Differential and Physical Geometry

Jeffrey M. Lee

Contents

0.1 Preface

`Note to reviewer: This book is in an incomplete state. It is being written in a highly nonlinear fashion. I intend to rewrite much, even in the first chapter. It seems possible that there may need to be two volumes since I do not want to exceed 450 pages per volume.`

`There will be several more comments such as this directed to the reviewer/reviewers which will indicate the authors plans for the text. I am interested in getting feedback before another rewrite.`

`The last section to be rewritten will be the preface since only at that point will I know exactly what I have been able to include.`

Classical differential geometry is that approach to geometry that takes full advantage of the introduction of numerical coordinates into a geometric space. This use of coordinates in geometry was the essential insight of Rene Descartes"s which allowed the invention of analytic geometry and paved the way for modern differential geometry. A differential geometric space is firstly a topological space on which are defined a sufficiently nice family of coordinate systems. These spaces are called differentiable manifolds and examples are abundant. Depending on what type of geometry is to be studied, extra structure is assumed which may take the form of a distinguished group of symmetries, or the presence of a distinguished tensor such as a metric tensor or symplectic form.

Despite the necessary presence of coordinates, modern differential geometers have learned to present much of the subject without direct reference to any specific (or even generic) coordinate system. This is called the "invariant" or "coordinate free" formulation of differential geometry. Of course, as just pointed out, the existence of (smooth) locally defined coordinate systems is part of what is assumed at the beginning[1]. The only way to really see exactly what this all means is by diving in and learning the subject.

The relationship between geometry and the physical world is fundamental on many levels. Geometry (especially differential geometry) clarifies, codifies and then generalizes ideas arising from our intuitions about certain aspects of the our world. Some of these aspects are those that we think of as forming the spatiotemporal background against which our lives are played out while others are derived from our experience of solid objects and their surfaces. Of course these two are related; the earth is both an object and a "lived in space"; thus the "geo" in the word geometry. Differential geometry is an appropriate mathematical setting for the study of what we initially conceive of as continuous physical phenomenon such as fluids, gases and electromagnetic fields. The fields live on differentiable manifolds but also can be seen as taking values in special manifolds called fiber bundles. Manifolds have dimension. The surface of the earth is two dimensional, while the configuration space of a mechanical system is a manifold which may easily have a very high dimension (dimension=degrees of freedom). Stretching the imagination further we can conceive of each possible field configuration as being a single abstract point in an infinite dimensional manifold.

Geometry has its philosophical aspects. Differential geometry makes explicit, and abstracts from, our intuitions about what the world itself is as revealed by, as the German philosopher Kant would have it, the action of understanding on the manifold presentations of the senses. Kant held that space and time were "categories of the understanding". Kant thought that the truths of Euclidean geometry where a prior truths; what he called a priori synthetic truths. This often taken to imply that Kant believed that non-Euclidean geometries were impossible (at least as far as physical space is concerned) but perhaps this doesn't give enough credit to Kant. Another interpretation of Kant's arguments simply concludes that human intuition about the geometry of space must necessarily be Euclidean; a statement that is still questionable but certainly not the same as the claim that the real geometry of space must be of logical necessity Euclidean in nature. Since the advent of Einstein's theory of gravitation it is assumed that space may in fact be non-Euclidean. Despite this it does seem to be a fact that our innate conceptualization of space is Euclidean. Contrary to Kant's assertions, this seems to be the result of evolution and the approximately Euclidean nature of space at human scales. Einstein's theory of gravitation is an extension of special relativity which can be seen as unifying space and time. From that point of view, the theory is more a theory about spacetime rather than space. For philosophers and scientist alike, *space* and *time* are words that refer to basic aspects of the world and/or, depending on one philosophical persuasion, basic aspects of the human experience of the world. An implicit understanding of space and time is part and parcel of what if means to be conscious. As the continental philosophers might describe it, space and time are what stand between

[1]Equivalently, one can start by assume the existence of a subsheaf of the sheaf of continuous functions on the underlying topological space. This sheaf must be locally equivalent to the sheaf of smooth functions on a Euclidean space.

us and our goals. Heidegger maintained that there was in human beings an understanding of a deeper kind of time that he called "temporality[2]" that was lost to the explicitly self conscious scientific form of understanding. I believe that while this might be true in some sense, Heidegger and many of his modern admirers underestimate the extent to which mathematics has succeeded in capturing and making explicit our implicit existential understanding of spatiality and temporality. Again, evolutionary theory suggests a different analysis of human temporality. In any case, geometry as a branch of mathematics can be seen as related to humankind's effort to come to an exact and explicit understanding of these ubiquitous aspects of experienced reality. In this form geometry is abstracted and idealized. Heidegger would say that much is lost once the abstraction is effected, even so, what is gained is immense. Geometry is the first exact science.

At root, what is the actual nature of reality which makes us experience it in this largely geometric way? As hinted at above, this is the topic of much discussion among philosophers and scientists. That the abstractions of geometry are not mere abstractions is proved by the almost unbelievable usefulness of geometric thinking in physics. Differential geometry is especially useful for classical physics-including and especially Einstein's theory of gravitation (general relativity). On the other hand, there is also the presumably more fundamental quantum nature of physical reality. One has to face the possibility that the quantum nature of the physical world is not a drama played out on a preexisting stage of a classically conceived space (or spacetime) but rather it may be the case that, like temperature, space and time are emergent "macroscopic properties" of nature. In fact, it is popular among physicists to look for so called "background free" theories where the most basic entities that make up (a model of) reality conspire to create what we perceive as space and time. On the mathematical side, this idea is connected with the emerging fields of discrete differential geometry and noncommutative geometry or quantum geometry[3]. Whatever the outcome of this drive toward background free theory, geometry in a broad sense can be expected to remain important for physics. After all, whatever nature is in her own secret nature, it is evident that her choreography adds up to something highly geometric when viewed at a large scale. If physics wants to replace geometry as we know it today by something else, then it is left to explain why geometry emerges in the macroscopic world. Thus even the most ambitious background free theory must be shown to contain the seeds of geometry and the mechanism by which macroscopic geometry emerges must be explained. It would be a mistake to underestimate this task.

The physicist is interested in geometry because s/he wants to understand the way the physical world is in "actuality". But there is also a discovered "logical world" of pure geometry that is in some sense a part of reality too.

[2]Strangely, Heidegger did not see the need to form a similar concept of primordial spatiality.

[3]Quantum geometry is often taken to by synonymous with noncommutative geometry but there are also such things as Ashtekar's "quantum Riemannian geometry" which is doesn't quite fit so neatly into the field of noncommutative geometry as it is usually conceived.

This is the reality which Roger Penrose calls the Platonic world[4]. Thus the mathematician is interested in the way worlds *could be in principal*- they are interested in what might be called the "possible geometric worlds". Since the inspiration for what we find interesting has its roots in our experience, even the abstract geometries that we study retain a certain physicality. From this point of view, the intuition that guides the pure geometer is fruitfully enhanced by an explicit familiarity with geometry as it occurs in modern physical theory. The spaces studied in pure differential often have their own physics as it were. For example, what is the relationship between the geometry of the geodesic flow on a Riemannian manifold and the spectrum of the Laplace operator which is always naturally defined. Depending on the manifold in question, this may not be related to real physics but it is certainly "physics-like". One could roughly say that the spectral geometry of the Laplace operator on a manifold is the quantum mechanics of the geodesic flow. We may also just as well consider any of the other basic equations associated with the Laplace operator such as the heat equation. There seems to be a kind of generalized "physical thinking" that can profitably be applied when studying certain branches of mathematics. I think this is true for much of what is of interest in differential geometry.

Knowledge of differential geometry is common among physicists thanks to the success of Einstein's highly geometric theory of gravitation and also because of the discovery of the differential geometric underpinnings of modern gauge theory[5] and string theory. It is interesting to note that the gauge field concept was introduced into physics within just a few years of the time that the notion of a connection on a fiber bundle (of which a gauge field is a special case) was making its appearance in mathematics. Perhaps the most exciting, as well as challenging, piece of mathematical physics to come along in a while is the "string theory" mentioned above. String theory is, at least in part, a highly differential geometric theory. It is too early to tell if string theory can be brought out of its current mathematically inchoate state and whether or not it will turn out to provide an accurate model of the physical world.

The usefulness of differential geometric ideas for physics is also apparent in the conceptual payoff enjoyed when classical mechanics is reformulated in the language of differential geometry. Mathematically, we are led to the relatively new subjects of symplectic geometry and Poisson geometry which form a major topic for this book.

To be clear, this book is not a physics book but is a mathematics book which take inspiration from, and uses examples from physics. Although there is a great deal of physics that might be included in a book of this sort, the usual constraints of time and space make it possible to include only a small part. A similar statement holds concerning the differential geometry covered in the text. Differential geometry is a huge field and even if we had restricted our attention to just Riemannian geometry only a small fragment of what could be addressed

[4]Penrose seems to take this Platonic world rather literally giving it a great deal of ontological weight as it were.

[5]The notion of a connection on a fiber bundle and the notion of a gauge field are essentially identical concepts discovered independently by mathematicians and physicists.

at this level could possibly be included.

In choosing what to include in this book, I was guided by personal interest and, more importantly, by the limitations of my own understanding. There will most likely be mistakes in the text, some serious. For this I apologize.

0.1.1 Notation

Differential geometry is one of the subjects where notation is a continual problem. Notation that is highly precise from the vantage point of set theory and logic tends to be fairly opaque to the underlying geometry. On the other hand, notation that is true to intuition is difficult to make precise. Notation that is uncluttered and handy for calculations tends to suffer from ambiguities when looked at closely. It is perhaps worth pointing out that the kind of ambiguities we are talking about include some that are of the same sort as are accepted by every calculus student without much thought. For instance, we find (x, y, z) being used to refer variously to "indeterminates", "a triple of numbers", or a triple of functions of some variable as for example when we write

$$\vec{x}(t) = (x(t), y(t), z(t)).$$

Also, we often write $y = f(x)$ and then, even write $y = y(x)$ and $y'(x)$ or dy/dx instead of $f'(x)$ or df/dx. Polar coordinates are generally thought of as living on the $xy-$plane even though it could be argued that $r\theta-$space is really a (subset of a) different copy of $\mathbb{R}^2$. The ambiguities of this type of notation are as apparent. This does not mean that this notation is bad. In fact, it can be quite useful to use slightly ambiguous notation. Human beings are generally very good at handling ambiguity. In fact, if a self conscious desire to avoid logical inconsistency in notation is given priority over everything else we quickly begin to become immobilized. The reader should be warned that while we will develop fairly pedantic notation, perhaps too pedantic at times, we shall also not hesitate to resort to abbreviation and notational shortcuts as the need arises. This will be done with increasing frequency in later chapters. We set out a few of our most frequently used notational conventions

1. Throughout the book the symbol combination "$:=$" means "is equal to by definition".

2. The reader is reminded that for two sets A and B the **Cartesian product** $A \times B$ is the set of pairs $A \times B := \{(a, b) : a \in A, b \in B\}$. More generally, for a family of sets $\{A_i\}_{i \in I}$ we may form the product set $\prod_{i \in I} A_i := \{(a_1, a_2, ...) : a_i \in A_i\}$. In particular, $\mathbb{R}^n := \mathbb{R} \times \cdots \times \mathbb{R}$ is the product of n copies of the set real numbers $\mathbb{R}$.

3. Whenever we represent a linear transformation by a matrix, then the matrix acts on column vectors from the left. This means that in this context elements of $\mathbb{R}^n$ are thought of as column vectors. It is sometimes convenient to represent elements of the dual space $(\mathbb{R}^n)^*$ as row vectors so

that if $\alpha \in (\mathbb{R}^n)^*$ is represented by $(a_1, ..., a_n)$ and $v \in \mathbb{R}^n$ is represented by $\left(v^1, ..., v^n\right)^t$ then

$$\alpha(v) = (a_1.....a_n)\begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix}.$$

Since we do not want to have to write the transpose symbol in every instance and for various other reasons we will usually use upper indices (superscripts) to label the entries of elements of $\mathbb{R}^n$ and lower indices (subscripts) to label the entries of elements of $(\mathbb{R}^n)^*$. Thus $\left(v^1, ..., v^n\right)$ invites one to think of a column vector (*even when the transpose symbol is not present*) while $(a_1, ..., a_n)$ is a row vector. On the other hand, a list of elements of a vector space such as a what might form a basis will be labeled using subscripts while superscripts label lists elements of the dual space. For example, let V be a vector space with a basis $(\mathbf{f}_1, ..., \mathbf{f}_n)$ and dual basis $(\mathbf{f}^1, ..., \mathbf{f}^n)$ for V^*. Then we can write an arbitrary element $v \in \mathrm{V}$ variously by

$$\begin{aligned} v &= v^1\mathbf{f}_1 + \cdots + v^n\mathbf{f}_n = \sum v^i\mathbf{f}_i \\ &= (\mathbf{f}_1, ..., \mathbf{f}_n)\begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix} \end{aligned}$$

while an element $\alpha \in \mathrm{V}^*$ would usually be written in one of the following ways

$$\begin{aligned} \alpha &= a_1\mathbf{f}^1 + \cdots + a_n\mathbf{f}^n = \sum \alpha_i\mathbf{f}^i \\ &= (a_1, ..., a_n)\begin{pmatrix} \mathbf{f}^1 \\ \vdots \\ \mathbf{f}^n \end{pmatrix}. \end{aligned}$$

4. We also sometimes use the convention that when an index is repeated once up and once down, then a summation is implied. This is the **Einstein summation convention**. For example, $\alpha(v) = a_iv^i$ means $\alpha(v) = \sum_i a_iv^i$ while $s_l^{kr} = a_l^{ijk}b_{ij}^r$ means $s_l^{kr} = \sum_{1\le i,j\le n} a_l^{ijk}b_{ij}^r$.

5. Another useful convention that we will use often is that when we have a list of objects $(o_1, ..., o_N)$ then $(o_1, ..., \widehat{o_i}, ..., o_N)$ will mean the same list with the i-th object omitted. Similarly, $a_1 + \cdots + \widehat{a_i} + \cdots + a_N$ is a sum of elements in a list with a_i omitted.

6. Finally, in some situations a linear function A of a variable, say h, is written as Ah or $A \cdot h$ instead of $A(h)$. This notation is particularly useful

when we have a family of linear maps depending on a parameter. For example, the derivative of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ at a point $x \in \mathbb{R}^n$ is a linear map $Df(x) : \mathbb{R}^n \to \mathbb{R}^m$ and as such we may apply it to a vector $h \in \mathbb{R}^n$. But to write $Df(x)(h)$ is a bit confusing in some contexts such as when the function f needs to be given more descriptive but bulky notation. For this reason we write $Df(x) \cdot h$ or $Df|_x h$ to clarify the different roles of the variables x and h. As another example, if $x \to A(x)$ is an $m \times n$ matrix valued function we might write $A_x h$ for the matrix multiplication of $A(x)$ and $h \in \mathbb{R}^n$.

A chart containing more extensive information about notational conventions is included in appendix I.

Part I

Part I

Chapter 1

Background

I never could make out what those damn dots meant.

Lord Randolph Churchill

Differential geometry is largely based on the foundation of the theory of differentiable manifolds. The theory of differentiable manifolds and the associated maps is a natural extension of multivariable calculus. Multivariable calculus is said to be done on (or in) an n-dimensional Euclidean space $\mathbb{R}^n$ (also called, perhaps more appropriately, a Cartesian space since no metric geometry is implied at this point). We hope that the great majority of readers will be perfectly comfortable with standard advanced undergraduate level multivariable calculus. A reader who felt the need for a review could do no better than to study the classic book "Calculus on Manifolds" by Michael Spivak. This book does advanced calculus[1] in a manner most congruent the spirit of modern differential geometry. It also has the virtue of being short. On the other hand, a sort of review will also be given later in this chapter that simultaneously generalizes from $\mathbb{R}^n$ to infinite dimensional vector spaces (Banach spaces). For now we will just introduce few terms and notational conventions which could conceivably be unfamiliar to readers coming from fields outside of mathematics proper.

The elements of $\mathbb{R}^n$ are thought of as points and also as directions. The directional derivative, in a direction $h \in \mathbb{R}^n$, of a function $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ at some point $p \in U$, is defined to be the limit

$$D_h f(p) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}(f(p + \varepsilon h) - f(p)) \tag{1.1}$$

whenever this limit exists. The partial derivatives are defined by letting the direction h be one of the standard basis vectors which point in the coordinate directions. The have the various notations $D_{e_i} f(p) := \frac{\partial f}{\partial x^i}(p) := \partial_i f(p)$. If the limit above exists for all h and the map $h \mapsto D_h f(p)$ is linear and continuous then it is a good candidate for what we shall call *the derivative* of f at p. *Thus*

[1] Despite the title, most of Spivak's book is about calculus rather than manifolds.

the derivative of f at a point p is to be a linear map $Df_p : \mathbb{R}^n \to \mathbb{R}^m$. A map f from an open set $U \subset \mathbb{R}^n$ into another Euclidean space $\mathbb{R}^m$ is given by m functions (say $f_1, ..., f_m$) of n variables (say $u^1, ..., u^n$). The linear map Df_p, should it exist, is represented by the matrix of first partials:

$$\begin{bmatrix} \frac{\partial \sigma_1}{\partial u^1}(p) & \cdots & \frac{\partial \sigma_1}{\partial u^n}(p) \\ \vdots & & \vdots \\ \frac{\partial \sigma_m}{\partial u^1}(p) & \cdots & \frac{\partial \sigma_m}{\partial u^n}(p) \end{bmatrix}.$$

(Here we have given rough and ready definitions that will refined in the next section on calculus on general Banach spaces). The usual definition of a derivative as a linear map requires something more than just the existence of these directional derivatives. We save the official definition for our outline of calculus on Banach spaces below but mention that this fuller and stronger kind of differentiability is guaranteed if all the partial derivatives exist and are continuous on an open set which we may as well take to be the domain of the function.

Along with our notation $\mathbb{R}^n$ for the vector space of $n-$tuples of real numbers $\{(x^1, ..., x^n) : x^i \in \mathbb{R}\}$ we also have $\mathbb{C}^n$, the complex vector space consisting of $n-$tuples of complex numbers generically denoted $(z^1, ..., z^n)$ where $z^i = x^i + \sqrt{-1}y^i$. Also, let $\mathbb{Z}$ denote the integers and $\mathbb{Z}^+$ the nonnegative integers. We will eventually have occasion to use the following notations which are commonly found in books on partial differential equations. For $\alpha = (\alpha_1, ..., \alpha_n)$ and $\beta = (\beta_1, ..., \beta_n)$, $n-$tuples of non-negative integers let

$$|\alpha| := \alpha_1 + ... + \alpha_k$$
$$\alpha! := \alpha_1! \cdots \alpha_k!$$

and

$$\binom{\alpha}{\beta} := \frac{\alpha!}{\beta!(\alpha - \beta)!}.$$

Now if $x = (x^1, ..., x^n)$ and $\alpha = (\alpha_1, ..., \alpha_n)$ we set $x^\alpha = \left(x^1\right)^{\alpha_1} \cdots (x^n)^{\alpha_n}$, $z^\alpha = \left(z^1\right)^{\alpha_1} \cdots (z^n)^{\alpha_n}$, $\partial x^\alpha = \partial x^{\alpha_1} \cdots \partial x^{\alpha_n}$ and $\partial^\alpha f = \partial^{\alpha_1 + ... + \alpha_n} f$. Thus if $|\alpha| = k$, the notation for a k-th order partial derivative is

$$\frac{\partial^\alpha f}{\partial x^\alpha} = \frac{\partial^{\alpha_1 + ... + \alpha_n} f}{\partial x^{\alpha_1} \cdots \partial x^{\alpha_n}} \text{ or also}$$
$$D^\alpha f = \left(\frac{\partial^{\alpha_1}}{\partial x^{\alpha_1}}\right) \cdots \left(\frac{\partial^{\alpha_n}}{\partial x^{\alpha_n}}\right) f.$$

Most readers will be familiar with the following sets of functions:

1. $C^k(U)$: If U is an open set in $\mathbb{R}^n$ then $f \in C^k(U)$ if and only if all partial derivatives of order $\leq k$ exist and are continuous on U. We sometimes just say that f is C^k. Similarly, if E is $\mathbb{R}^m$ for some m or, more generally, some normed space, then $C^k(U, \mathsf{E})$ denotes the set of all maps $f : U \to \mathsf{E}$

which are C^k. For the meaning of this when E is an infinite dimensional normed space the reader must wait for an explanation until later in this chapter. On the other hand, if $\mathsf{E} = \mathbb{R}^m$ then f is composed of m real valued functions of n variables and then we say that f is in $C^k(U, \mathbb{R}^m)$ if each of the component functions is in $C^k(U, \mathbb{R}^m)$

2. $C^\infty(U)$ is the set of f such that $f \in C^k(U)$ for all $k \geq 0$. These functions are said to be infinitely differentiable or "smooth".

3. $C^k_c(U)$: The **support** of a function f with domain U is a closed set denoted $\operatorname{supp}(f)$ and defined to be the closure of the set $\{x \in U : f(x) \neq 0\}$. It is the smallest closed set with the property that f is zero on its compliment. This definition of support applies also to maps taking values in a general vector space. The set of all functions f in $C^k(U)$ which have compact support in U is denoted $C^k_c(U)$. Similarly, if E is $\mathbb{R}^m$ or some normed space then $C^k_c(U, \mathsf{E})$ denotes the set of all maps f from the set $C^k(U, \mathsf{E})$ that happen to have compact support.

4. $C^\omega(U)$: The set $C^\omega(U)$ denotes the set of all functions f defined on U which have an absolutely convergent power series (radius of convergence > 0) of the form
$$f(x) = \sum_{0 \leq |\alpha| < \infty} r_\alpha (x-a)^\alpha$$
for all $a \in U$. The functions of $C^\omega(U)$ are said to be **real analytic** on U. We remind the reader that if $f \in C^\omega(U)$ then $f \in C^\infty(U)$ and the power series above is also the Taylor expansion of f at a :
$$f(x) = \sum_{0 \leq |\alpha| < \infty} \frac{1}{\alpha!} D^\alpha f(a)(x-a)^\alpha$$
(The notation will be further clarified below).

5. Let $n \leq m$. If $f : U \subset \mathbb{R}^n \to V \subset \mathbb{R}^m$ is a C^k map such that the matrix of partial derivatives is of rank n then is an C^k **immersion**. A C^k map $\phi : U \subset \mathbb{R}^n \to V \subset \mathbb{R}^n$ which is a homeomorphism and such that ϕ^{-1} is also a C^k map is called a **diffeomorphism**. By convention, if $k = 0$ then a diffeomorphism is really just a homeomorphism between open set of Euclidean space.

Later we generalize differential calculus to infinite dimensional spaces and some definitions given above will be repeated sometimes in a different formulation.

1.1 Naive Functional Calculus.

We have recalled the basic definitions of the directional derivative of a map such as $f : \mathbb{R}^n \to \mathbb{R}^m$. This is a good starting point for making the generalizations to come but let us think about a bit more about our "directions" h and "points" p. In both cases these refer to $n-$tuples in $\mathbb{R}^n$. The values taken by the function are also tuples ($m-$tuples in this instance). From one point of view a $n-$tuple is just a function whose domain is the finite set $\{1, 2, ..., n\}$. For instance, the n-tuple $h = (h^1, ..., h^n)$ is just the function $i \mapsto h^i$ which may as well have been written $i \mapsto h(i)$. This suggests that we generalize to functions whose domain is an infinite set. A sequence of real numbers is just such an example but so is any real (or complex) valued function. This brings us to the notion of a function space. An example of a function space is $C([0, 1])$, the space of continuous functions on the unit interval $[0, 1]$. So, whereas an element of $\mathbb{R}^3$, say $(1, \pi, 0)$ has 3 components or entries, an element of $C([0, 1])$, say $(t \mapsto \sin(2\pi t))$ has a continuum of "entries". For example, the $1/2$ entry of the latter element is $\sin(2\pi(1/2)) = 0$. So one approach to generalizing the usual setting of calculus might be to consider replacing the space of $n-$tuples $\mathbb{R}^n$ by a space of functions. Now we are interested in differentiating functions whose arguments are themselves functions. This type of function is sometimes called a functional. We shall sometimes follow the tradition of writing $F[f]$ instead of $F(f)$. Some books even write $F[f(x)]$. Notice that this is *not* a composition of functions. A simple example of a functional on $C([0, 1])$ is

$$F[f] = \int_0^1 f^2(x)dx.$$

We may then easily define a formal notion of directional derivative:

$$(D_h F)[f] = \lim_{\epsilon \to 0} \frac{1}{\epsilon}(F[f + \epsilon h] - F[f])$$

where h is some function which is the "direction vector". This also allows us to define the differential δF which is a linear map on the functions space given at f by $\delta F|_f h = (D_h F)[f]$. We use a δ instead of a d to avoid confusion between dx^i and δx^i which comes about when x^i is simultaneously used to denote a number and also a function of , say, t.

It will become apparent that choosing the right function space for a particular problem is highly nontrivial and in each case the function space must be given an appropriate topology. In the following few paragraphs our discussion will be informal and we shall be rather cavalier with regard to the issues just mentioned. After this informal presentation we will develop a more systematic approach (Calculus on Banach spaces).

The following is another typical example of a functional defined on the space $C^1([0, 1])$ of continuously differentiable functions defined on the interval $[0, 1]$:

$$S[c] := \int_0^1 \sqrt{1 + (dc/dt)^2}dt$$

The reader may recognize this example as the arc length functional. The derivative *at* the function c in the *direction of* a function $h \in C^1([0, 1])$ would be given by

$$\delta S|_c(h) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\left(S[c + \varepsilon h] - S[c]\right).$$

It is well known that if $\delta S|_c(h) = 0$ for every h then c is a linear function; $c(t) = at + b$. The condition $\delta S|_c(h) = 0 = 0$ (for all h) is often simply written as $\delta S = 0$. We shall have a bit more to say about this notation shortly. For examples like this one, the analogy with multi-variable calculus is summarized as

$$\begin{aligned} \text{The index or argument becomes continuous:} &\quad i \rightsquigarrow t \\ d\text{-tuples become functions:} &\quad x^i \rightsquigarrow c(t) \\ \text{Functions of a vector variable become functionals of functions:} &\quad f(\vec{x}) \rightsquigarrow S[c] \end{aligned}$$

Here we move from $d-$tuples (which are really functions with finite domain) to functions with a continuous domain. The function f of x becomes a functional S of functions c.

We now exhibit a common example from the mechanics which comes from considering a bead sliding along a wire. We are supposed to be given a so called "Lagrangian function" $L : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ which will be the basic ingredient in building an associated functional. A typical example is of the form $L(x, v) = \frac{1}{2}mv^2 - V(x)$. Define the action functional S by using L as follows: For a given function $t \longmapsto q(t)$ defined on $[a, b]$ let

$$S[q] := \int_a^b L(q(t), \dot{q}(t))dt.$$

We have used x and v to denote variables of L but since we are eventually to plug in $q(t), \dot{q}(t)$ we could also follow the common tradition of denoting these variables by q and $\dot{q}$ but then it must be remembered that we are using these symbols in two ways. In this context, one sometimes sees something like following expression for the so-called variation

$$\delta S = \int \frac{\delta S}{\delta q(t)} \delta q(t) dt \tag{1.2}$$

Depending on one's training and temperament, the meaning of the notation may be a bit hard to pin down. First, what is the meaning of δq as opposed to, say, the differential dq? Second, what is the mysterious $\frac{\delta S}{\delta q(t)}$? A good start might be to go back and settle on what we mean by the differential in ordinary multivariable calculus. For a differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ we take df to just mean the map

$$df : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$$

given by $df(p, h) = f'(p)h$. We may also fix p and write $df|_p$ or $df(p)$ for the linear map $h \mapsto df(p, h)$. With this convention we note that $dx^i\big|_p(h) = h^i$ where $h = (h^1, ..., h^d)$. Thus applying both sides of the equation

$$df|_p = \sum \frac{\partial f}{\partial x^i}(p)\, dx^i\big|_p \tag{1.3}$$

to some vector h we get

$$f'(p)h = \sum \frac{\partial f}{\partial x^i}(p)h^i. \tag{1.4}$$

In other words, $df|_p = D_h f(p) = \nabla f \cdot h = f'(p)$. Too many notations for the same concept. Equation 1.3 is clearly very similar to $\delta S = \int \frac{\delta S}{\delta q(t)} \delta q(t) dt$ and so we expect that δS is a linear map and that $t \mapsto \frac{\delta S}{\delta q(t)}$ is to δS as $\frac{\partial f}{\partial x^i}$ is to df:

$$\begin{aligned} df &\rightsquigarrow \delta S \\ \frac{\partial f}{\partial x^i} &\rightsquigarrow \frac{\delta S}{\delta q(t)}. \end{aligned}$$

Roughly, $\frac{\delta S}{\delta q(t)}$ is taken to be whatever function (or distribution) makes the equation 1.2 true. We often see the following type of calculation

$$\begin{aligned} \delta S &= \delta \int L dt \\ &= \int \left(\frac{\partial L}{\partial q}\delta q + \frac{\partial L}{\partial \dot{q}}\delta \dot{q}\right) dt \\ &= \int \left\{\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}}\right\} \delta q dt \end{aligned} \tag{1.5}$$

from which we are to conclude that

$$\frac{\delta S}{\delta q(t)} = \frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}}$$

Actually, there is a subtle point here in that we must restrict δS to variations for which the integration by parts is justified. We can make much better sense of things if we have some notion of derivative for functionals defined on some function space. There is also the problem of choosing an appropriate function space. On the one hand, we want to be able to take (ordinary) derivatives of these functions since they may appear in the very definition of S. On the other hand, we must make sense out of limits so we must pick a space of functions with a tractable and appropriate topology. We will see below that it is very desirable to end up with what is called a Banach space. Often one is forced to deal with more general topological vector spaces. Let us ignore all of these worries for a bit longer and proceed formally. If δS is somehow the variation due to a variation $h(t)$ of $q(t)$ then it depends on both the starting position in function space (namely, the function $q(.)$) and also the direction in function space that we move (which is the function $h(.)$). Thus we interpret $\delta q = h$ as some appropriate function and then interpret δS as short hand for

$$\begin{aligned} \delta S|_{q(.)} h &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}(S[q + \varepsilon h] - S[q]) \\ &= \int \left(\frac{\partial L}{\partial q} h + \frac{\partial L}{\partial \dot{q}} \dot{h}\right) dt \end{aligned}$$

Note: If we had been less conventional and more cautious about notation we would have used c for the function which we have been denoting by $q : t \mapsto q(t)$. Then we could just write $\delta S|_c$ instead of $\delta S|_{q(.)}$. The point is that the notation $\delta S|_q$ might leave one thinking that $q \in \mathbb{R}$ (which it is under one interpretation!) but then $\delta S|_q$ would make no sense. It is arguably better to avoid letting q refer both to a number and to a function even though this is quite common. At any rate, from here we restrict attention to "directions" h for which $h(a) = h(b) = 0$ and use integration by parts to obtain

$$\delta S|_{q(.)}\, h = \int \{\frac{\partial L}{\partial x^i}(q(t), \dot{q}(t)) - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}}(q(t), \dot{q}(t))\} h^i(t)dt.$$

So it seems that the function $E(t) := \frac{\partial L}{\partial q}(q(t), \dot{q}(t)) - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i}(q(t), \dot{q}(t))$ is the right candidate for the $\frac{\delta S}{\delta q(t)}$. However, once again, we must restrict to h which vanish at the boundary of the interval of integration. On the other hand, this family is large enough to force the desired conclusion. Despite this restriction the function $E(t)$ is clearly important. For instance, if $\delta S|_{q(.)} = 0$ (or even $\delta S|_{q(.)}\, h = 0$ for all functions that vanish at the end points) then we may conclude easily that $E(t) \equiv 0$. This gives an equation (or system of equations) known as the Euler-Lagrange equation for the function $q(t)$ corresponding to the action functional S :

$$\frac{\partial L}{\partial q}(q(t), \dot{q}(t)) - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}}(q(t), \dot{q}(t)) = 0$$

Exercise 1.1 *Replace $S[c] = \int L(c(t), \dot{c}(t))dt$ by the similar function of several variables $S(c_1, ...c_N) = \sum L(c_i, \triangle c_i)$. Here $\triangle c_i := c_i - c_{i-1}$ (taking $c_0 = c_N$) and L is a differentiable map $\mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$. What assumptions on $c = (c_1, ...c_N)$ and $h = (h_1, ...h_N)$ justify the following calculation?*

$$\begin{aligned}
dS|_{(c_1,...c_N)}\, h &= \sum \frac{\partial L}{\partial c_i} h^i + \frac{\partial L}{\partial \triangle c_i} \triangle h^i \\
&= \sum \frac{\partial L}{\partial c_i} h^i + \sum \frac{\partial L}{\partial \triangle c_i} h^i - \sum \frac{\partial L}{\partial \triangle c_i} h^{i-1} \\
&= \sum \frac{\partial L}{\partial c_i} h^i + \sum \frac{\partial L}{\partial \triangle c_i} h^i - \sum \frac{\partial L}{\partial \triangle c_{i+1}} h^i \\
&= \sum \frac{\partial L}{\partial c_i} h^i - \sum (\frac{\partial L}{\partial \triangle c_{i+1}} - \frac{\partial L}{\partial \triangle c_i}) h^i \\
&= \sum \{\frac{\partial L}{\partial c_i} h^i - (\triangle \frac{\partial L}{\partial \triangle c_i})\} h^i \\
&= \sum \frac{\partial S}{\partial c_i} h^i.
\end{aligned}$$

The upshot of our discussion is that the δ notation is just an alternative notation to refer to the differential or derivative. Note that q^i might refer to a coordinate or to a function $t \mapsto q^i(t)$ and so dq^i is the usual differential and maps $\mathbb{R}^d$ to $\mathbb{R}$ whereas $\delta x^i(t)$ is either taken as a variation function $h^i(t)$ as above or as the map $h \mapsto \delta q^i(t)(h) = h^i(t)$. In the first interpretation $\delta S = \int \frac{\delta S}{\delta q^i(t)} \delta q^i(t) dt$ is an abbreviation for $\delta S(h) = \int \frac{\delta S}{\delta q^i(t)} h^i(t) dt$ and in the second interpretation it is the map $\int \frac{\delta S}{\delta q^i(t)} \delta q^i(t) dt : h \mapsto \int \frac{\delta S}{\delta q^i(t)} (\delta q^i(t)(h)) dt = \int \frac{\delta S}{\delta q^i(t)} h^i(t) dt$. The various formulas make sense in either case and both interpretations are ultimately equivalent. This much the same as taking the dx^i in $df = \frac{\partial f}{\partial x^i} dx^i$ to be components of an arbitrary vector $(dx^1, ..., dx^d)$ or we may take the more modern view that dx^i is a linear map given by $dx^i : h \mapsto h^i$. If this seems strange recall that x^i itself is also interpreted both as a number and as a coordinate function.

Example 1.1 *Let $F[c] := \int_{[0,1]} c^2(t) dt$ as above and let $c(t) = t^3$ and $h(t) = \sin(t^4)$. Then*

$$\begin{aligned}
\delta F|_c (h) = D_h F(c) &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left(F[c + \varepsilon h] - F[c] \right) \\
&= \left. \frac{d}{d\varepsilon} \right|_{\varepsilon=0} F[c + \varepsilon h] \\
&= \left. \frac{d}{d\varepsilon} \right|_{\varepsilon=0} \int_{[0,1]} (c(t)) + \varepsilon h(t))^2 dt \\
&= 2 \int_{[0,1]} c(t) h(t) dt = 2 \int_0^1 t^3 \sin(\pi t^4) dx \\
&= \frac{1}{\pi}
\end{aligned}$$

Note well that h and c are functions but here they are, more importantly, "points" in a function space! What we are differentiating is F. Again, $F[c]$ is *not* a composition of functions; the function c itself is the dependent variable here.

Exercise 1.2 *Notice that for a smooth function $s : \mathbb{R} \to \mathbb{R}$ we may write*

$$\frac{\partial s}{\partial x^i}(x_0) = \lim_{h \to 0} \frac{s(x_0 + h e_i) - s(x_0)}{h}$$
$$\text{where } e_i = (0, ..., 1, ...0)$$

Consider the following similar statement which occurs in the physics literature quite often.

$$\frac{\delta S}{\delta c(t)} = \lim_{h \to 0} \frac{S[c + h\delta_t] - S[c]}{h} \quad ?$$

Here δ_t is the Dirac delta function (distribution) with the defining property $\int \delta_t \phi = \phi(t)$ for all continuous ϕ. To what extent is this rigorous? Try a formal calculation using this limit to determine $\frac{\delta S}{\delta c(t)}$ in the case that

$$S(c) := \int_0^1 c^3(t) dt.$$

1.2 Calculus on Normed Spaces

A basic object in differential geometry is the differentiable manifold. The theory of differentiable manifolds and the associated maps and other constructions, is really just an extension of the calculus in to a setting where, for topological reasons, we may need to use several coordinates systems. Once the coordinate systems are in place, many endeavors reduce to the kind of calculation already familiar from calculus of several variables. Indeed, on a certain level, a differentiable manifold is just a "place to do calculus". As we shall see, extra structures such as metrics, connections, symplectic forms and so on, need to be imposed on a differentiable manifold if certain parts of the calculus of several variables are to be extended to this more general setting. In this section we review the basics of differential calculus and exhibit calculus in a more general, infinite dimensional setting. Namely, we introduce calculus in Banach spaces. Thus we arrive at a version of the differential calculus that is more inclusive than what is usually taught in a standard advanced calculus course. Our reasons for doing this are as follows:

1. Exhibiting this generalization gives us a review of some notions from calculus from a coordinate free viewpoint.

2. The Banach space calculus often provides a more rigorous version of the naive functional calculus and can be useful for variational aspects of differential geometry such as the theory of harmonic maps and Morse theory.

3. The generalization is straightforward. Most proofs are formally identical to their finite dimensional versions.

The set of all maps of a certain type between finite dimensional differentiable manifolds (still to be officially defined below) is a set which often turns out to be an *infinite dimensional* differentiable manifold. We use the extension of calculus to Banach spaces to facilitate a definition of infinite dimensional differentiable manifold. This is a space that locally looks like a Banach space (defined below).

The discussion of functional calculus was informal and driven by analogies. It was an informal differential calculus for functionals whose arguments are functions living in some function space. This way of thinking of things is not favored by mathematicians not only because of the evident lack of rigor and problematic notation but also because it does not seem to focus on the essential features of the situation. The essential features are evident in the expression 1.1 which defines the derivative as it is usually given in an advanced calculus course. The additions and subtractions tell us that the vector space structure of the domain and range spaces are important while the limit tells us that topology must play a role. Following this lead we arrive at the preliminary conclusion that *calculus is done on an open set in a topological vector space.* Even this will have to be further generalized to the statement that differential calculus is done on differentiable manifolds (possibly infinite dimensional) but for now we stay with the present insight.

Before we give the formal definitions let us be up front about which aspects of the Euclidean spaces are most important for calculus:

1. $\mathbb{R}^d$ has a topology given by the distance function

$$d(x,y) = \|x-y\| = \sqrt{\sum_{i=i}^{d}(x^i - y^i)^2}$$

   and an important feature here is the fact that this distance function arises because of the existence of a norm; in this case the norm of an element $v = (v^1, ..., v^d)$ is given by $\|v\| = \sqrt{\sum_{i=i}^{d}(v^i)^2}$.

2. Addition and subtraction are continuous maps $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ where $\mathbb{R}^d \times \mathbb{R}^d$ is given the product topology which is the smallest topology so that the projection maps $pr_1 : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ and $pr_2 : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ given by $(x,y) \mapsto x$ and $(x,y) \mapsto y$ are continuous. Scalar multiplication is also a continuous map $\mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$.

3. The topology given by the norm in (1) makes $\mathbb{R}^d$ a complete metric space. Thus all Cauchy sequences converge.

4. If $l(v) = 0$ for all (continuous) linear maps $l : \mathbb{R}^d \to \mathbb{R}$ then $v = 0$.

Number (1) leads to the notion of a normed vector space and number (2) to the more general notion of a "topological vector space". A normed space will automatically gives a topological vector space. The importance of number (3) is less obvious but turns out to be an important ingredient for generalized versions of the implicit function theorem. Number (4) seems almost a tautology since the projection maps $(v^1, ..., v^d) \mapsto v^i$ are continuous linear functionals. In general this is not a triviality and is essentially the Hahn-Banach theorem. Also, the requirement that the linear functional is continuous is redundant in $\mathbb{R}^d$ since all linear functionals on $\mathbb{R}^d$ are continuous[2].

Even though some aspects of calculus can be generalized without problems for fairly general spaces, the most general case that we shall consider is the case of *complete* normed vector spaces. These spaces are the so called **Banach spaces**.

So it seems that we can do calculus on some infinite dimensional spaces but there are several subtle points that arise. As we have suggested, there must be a topology on the space with respect to which addition and scalar multiplication are continuous. This is the meaning of "topological vector space". Also, in order for the derivative to be unique the topology must be Hausdorff. But there are more things to worry about. What we usually think of as differentiability of a function at a point x_0 entails the existence of a linear map $Df(x_0) = Df|_{x_0}$ that appropriately approximates f near x_0. If the topology is given by a norm

[2] A related point is that all norms on a finite dimensional vector space are equivalent in a very strong sense. In particular, they always give the same topology.

$\|.\|$ then a function f is differentiable at x_0 if there is a *continuous* linear map $Df|_{x_0}$ such that

$$\lim_{\|h\|\to 0} \frac{\left\| f(x_0+h) - f(x_0) - Df|_{x_0} \cdot h \right\|}{\|h\|} = 0.$$

We are also interested in having a version of the implicit mapping theorem. It turns out that most familiar facts from multivariable calculus, including the implicit mapping theorem go through if we replace $\mathbb{R}^n$ by a Banach space. In some cases we will state definitions for more general space normed spaces.

Remark 1.1 *There are at least two issues that remain even if we restrict ourselves to Banach spaces. First, the existence of* smooth *cut-off functions and* smooth *partitions of unity (to be defined below) are not guaranteed. The existence of smooth cut-off functions and smooth partitions of unity for infinite dimensional spaces is a case-by-case issue while in the finite dimensional case their existence is guaranteed. Second, there is the fact that a subspace of a Banach space is not a Banach space unless it is also a closed subspace. This will make a slight difference in some definitions and proofs*

For simplicity and definiteness all normed spaces and Banach spaces in this section will be real Banach spaces. Given two normed spaces V_1 and V_2 with norms $\|.\|_1$ and $\|.\|_2$ we can form a normed space from the Cartesian product $\mathsf{V}_1 \times \mathsf{V}_2$ by using the norm $\|(v,u)\| := \max\{\|v\|_1, \|u\|_2\}$. The vector space structure on $\mathsf{V}_1 \times \mathsf{V}_2$ is that of the (outer) direct sum and we could also write $\mathsf{V}_1 \oplus \mathsf{V}_2$. Recall that two norms on V, say $\|.\|'$and $\|.\|''$ are equivalent if there exist positive constants c and C such that

$$c\,\|x\|' \leq \|x\|'' \leq C\,\|x\|'$$

for all $x \in \mathsf{V}$. There are many equivalent norms for $\mathsf{V}_1 \times \mathsf{V}_2$ including

$$\|(v,u)\|' := \sqrt{\|v\|_1^2 + \|u\|_2^2}$$

and also

$$\|(v,u)\|'' := \|v\|_1 + \|u\|_2\,.$$

If V_1 and V_2 are Banach spaces then so is $\mathsf{V}_1 \times \mathsf{V}_2$ with either of the above norms. Also, if W_1 and W_2 are closed subspaces of V such that $\mathsf{W}_1 \cap \mathsf{W}_2 = \{0\}$ and such that every $v \in \mathsf{V}$ can be written uniquely in the form $v = w_1 + w_2$ where $w_1 \in \mathsf{W}_1$ and $\mathsf{w}_2 \in \mathsf{W}_2$ then we write $\mathsf{V} = \mathsf{W}_1 \oplus \mathsf{W}_2$. This is the inner direct sum. In this case there is the natural continuous linear isomorphism $\mathsf{W}_1 \times \mathsf{W}_2 \cong \mathsf{W}_1 \oplus \mathsf{W}_2$ given by

$$(w_1, w_2) \longleftrightarrow w_1 + w_2.$$

When it is convenient, we can identify $\mathsf{W}_1 \oplus \mathsf{W}_2$ with $\mathsf{W}_1 \times \mathsf{W}_2$ and in this case we hedge our bets, as it were, and write $w_1 + w_2$ for either (w_1, w_2)

or $w_1 + w_2$ letting the context determine the precise meaning if it matters. Under the representation (w_1, w_2) we need to specify what norm we are using and as mentioned above, there is more than one natural choice. We take $\|(w_1, w_2)\| := \max\{\|w_1\|, \|w_2\|\}$ but if the spaces are Hilbert spaces we might also use $\|(w_1, w_2)\|_2 := \sqrt{\|w_1\|^2 + \|w_2\|^2}$.

Let E be a Banach space and $\mathsf{W} \subset \mathsf{E}$ a *closed* subspace. We say that W is **complemented** if there is a *closed* subspace W' such that $\mathsf{E} = \mathsf{W} \oplus \mathsf{W}'$. We also say that W is a **split subspace** of E.

Definition 1.1 *Let E and F be normed spaces. A linear map $A : \mathsf{E} \longrightarrow \mathsf{F}$ is said to be bounded if*

$$\|A(v)\| \leq C \|v\|$$

for all $v \in \mathsf{E}$. For convenience, we have used the same notation for the norms in both spaces. If A is surjective and if $\|A(v)\| = \|v\|$ for all $v \in \mathsf{E}$ we call A an ***isometry****.*

It is a standard fact that a linear map between normed spaces is bounded if and only if it is continuous.

Definition 1.2 (Notation) *We will denote the set of all continuous (bounded) linear maps from a normed space E to a normed space F by $L(\mathsf{E}, \mathsf{F})$. The set of all continuous linear isomorphisms from E onto F will be denoted by $G\mathrm{L}(\mathsf{E}, \mathsf{F})$. In case, $\mathsf{E} = \mathsf{F}$ the corresponding spaces will be denoted by $\mathfrak{gl}(\mathsf{E})$ and $G\mathrm{L}(\mathsf{E})$. Here $G\mathrm{L}(\mathsf{E})$ is a group under composition and is called the general linear group .*

Definition 1.3 *Let V_i, $i = 1, ..., k$ and W be normed spaces. A map $\mu : \mathsf{V}_1 \times \cdots \times \mathsf{V}_k \to \mathsf{W}$ is called multilinear (k-multilinear) if for each i, $1 \leq i \leq k$ and each fixed $(w_1, ..., \widehat{w_i}, ..., w_k) \in \mathsf{V}_1 \times \cdots \times \widehat{\mathsf{V}_1} \times \cdots \times \mathsf{V}_k$ we have that the map*

$$v \mapsto \mu(w_1, ..., \underset{i-th\ slot}{v}, ..., w_k),$$

obtained by fixing all but the i-th variable, is a bounded linear map. In other words, we require that μ be $\mathbb{R}$- linear in each slot separately. A multilinear map $\mu : \mathsf{V}_1 \times \cdots \times \mathsf{V}_k \to \mathsf{W}$ is said to be ***bounded*** *if and only if there is a constant C such that*

$$\|\mu(v_1, v_2, ..., v_k)\|_{\mathsf{W}} \leq C \|v_1\|_{\mathsf{E}_1} \|v_2\|_{\mathsf{E}_2} \cdots \|v_k\|_{\mathsf{E}_k}$$

for all $(v_1, ..., v_k) \in \mathsf{E}_1 \times \cdots \times \mathsf{E}_k$.

Now $\mathsf{V}_1 \times \cdots \times \mathsf{V}_k$ is a normed space in several equivalent ways just in the same way that we defined before for the case $k = 2$. The topology is the product topology. Now one can show with out much trouble that a multilinear map $\mu : \mathsf{V}_1 \times \cdots \times \mathsf{V}_k \to \mathsf{W}$ is bounded if and only if it is continuous.

Notation 1.1 *The set of all bounded multilinear maps* $\mathsf{E}_1 \times \cdots \times \mathsf{E}_k \to \mathsf{W}$ *will be denoted by* $L(\mathsf{E}_1, ..., \mathsf{E}_k; \mathsf{W})$. *If* $\mathsf{E}_1 = \cdots = \mathsf{E}_k = \mathsf{E}$ *then we write* $L^k(\mathsf{E}; \mathsf{W})$ *instead of* $L(\mathsf{E}, ..., \mathsf{E}; \mathsf{W})$

Definition 1.4 *Let* E *be a Hilbert space with inner product denoted by* $\langle .,. \rangle$. *Then* $\mathsf{O}(\mathsf{E})$ *denotes the group of linear isometries from* E *onto itself. That is, the bijective linear maps* $\Phi : \mathsf{E} \to \mathsf{E}$ *such that* $\langle \Phi v, \Phi w \rangle = \langle v, w \rangle$ *for all* $v, w \in \mathsf{E}$. *The group* $\mathsf{O}(\mathsf{E})$ *is called the* ***orthogonal group*** *(or sometimes the* ***Hilbert group*** *in the infinite dimensional case).*

Notation 1.2 *For linear maps* $T : \mathsf{V} \to \mathsf{W}$ *we sometimes write* $T \cdot v$ *instead of* $T(v)$ *depending on the notational needs of the moment. In fact, a particularly useful notational device is the following: Suppose we have map* $A : X \to L(\mathsf{V}; \mathsf{W})$. *Then* $A(x)(\ v)$ *makes sense but we may find ourselves in a situation where* $A|_x\ v$ *is even more clear. This latter notation suggests a family of linear maps* $\{\ A|_x\}$ *parameterized by* $x \in X$.

Definition 1.5 *A (bounded) multilinear map* $\mu : \mathsf{V} \times \cdots \times \mathsf{V} \to \mathsf{W}$ *is called* ***symmetric*** *if for any* $v_1, v_2, ..., v_k \in \mathsf{V}$ *we have that*

$$\mu(v_{\sigma 1}, v_{\sigma 2}, ..., v_{\sigma k}) = \mu(v_1, v_2, ..., v_k)$$

for all permutations σ *on the letters* $\{1, 2, ..., k\}$. *Similarly,* μ *is called* ***skew-symmetric*** *or* ***alternating if***

$$\mu(v_{\sigma 1}, v_{\sigma 2}, ..., v_{\sigma k}) = sgn(\sigma)\mu(v_1, v_2, ..., v_k)$$

for all permutations σ. *The set of all bounded symmetric (resp. skew-symmetric) multilinear maps* $\mathsf{V} \times \cdots \times \mathsf{V} \to \mathsf{W}$ *is denoted* $L^k_{sym}(\mathsf{V}; \mathsf{W})$ *(resp.* $L^k_{skew}(\mathsf{V}; \mathsf{W})$ *or* $L^k_{alt}(\mathsf{V}; \mathsf{W})$*).*

Now if the W is complete, that is, if W is a Banach space then the space $L(\mathsf{V}, \mathsf{W})$ is a Banach space in its own right with norm given by

$$\|A\| = \sup_{v \in \mathsf{V}, v \neq 0} \frac{\|A(v)\|_{\mathsf{W}}}{\|v\|_{\mathsf{V}}} = \sup\{\|A(v)\|_{\mathsf{W}} : \|v\|_{\mathsf{V}} = 1\}.$$

Also, if W is a Banach space then the spaces $L(\mathsf{E}_1, ..., \mathsf{E}_k; \mathsf{W})$ are also Banach spaces normed by

$$\|\mu\| := \sup\{\|\mu(v_1, v_2, ..., v_k)\|_{\mathsf{W}} : \|v_i\|_{\mathsf{E}_i} = 1 \text{ for } i = 1, .., k\}$$

Proposition 1.1 *A k-multilinear map* $\mu \in L(\mathsf{E}_1, ..., \mathsf{E}_k; \mathsf{W})$ *is continuous if and only if it is bounded.*

We shall need to have several normed spaces handy for examples.

Example 1.2 *Consider a bounded open subset Ω of $\mathbb{R}^n$. Let $L^p_k(\Omega)$ denote the Banach space obtained by taking the completion of the set $C^k(\Omega)$ of k-times continuously differentiable real valued functions on Ω with the norm given by*

$$\|f\|_{k,p} := \left(\int_\Omega \sum_{|\alpha| \le k} \left| \frac{\partial^\alpha}{\partial x^\alpha} f(x) \right|^p \right)^{1/p}$$

Note that in particular $L^p_0(\Omega) = L^p(\Omega)$ is the usual L^p-space from real analysis.

Exercise 1.3 *Show that the map $C^k(\Omega) \to C^{k-1}(\Omega)$ given by $f \mapsto \frac{\partial f}{\partial x^i}$ is bounded if we use the norms $\|.\|_{k,p}$ and $\|.\|_{k-1,p}$ defined above. Show that we may extend this to a bounded map $L^p_k(\Omega) \to L^p_{k-1}(\Omega)$.*

There is a natural linear bijection $L(\mathsf{V}, L(\mathsf{V}, \mathsf{W})) \cong L^2(\mathsf{V}, \mathsf{W})$ given by $T \mapsto \iota$ T where

$$(\iota T)(v_1)(v_2) = T(v_1, v_2)$$

and we identify the two spaces and write T instead of $\iota\ T$. We also have $L(\mathsf{V}, L(\mathsf{V}, L(\mathsf{V}, \mathsf{W})) \cong L^3(\mathsf{V}; \mathsf{W})$ and in general $L(\mathsf{V}, L(\mathsf{V}, L(\mathsf{V}, ..., L(\mathsf{V}, \mathsf{W})) \cong L^k(\mathsf{V}; \mathsf{W})$ etc. It is also not hard to show that the isomorphism above is continuous and in fact, norm preserving so in fact ι is an isometry.

Definition 1.6 *Let V and W be normed spaces with norms $\|.\|_\mathsf{V}$ and $\|.\|_\mathsf{W}$. Let $U \subset \mathsf{V}$ be open and $x_0 \in U$. Let $f : U \to \mathsf{W}$ be a function. If there is an element $\delta f(x_0, h) \in \mathsf{W}$ such that defined for $h \in \mathsf{V}$ such that*

$$\lim_{t \to 0} \left\| \frac{f(x_0 + th) - f(x_0)}{t} - \delta f(x_0, h) \right\|_\mathsf{W} = 0$$

*then we say that $\delta f(x_0, h)$ is the **Gateaux directional derivative of** f **at** x_0 **in the direction** h. If there is a map $h \mapsto \delta f(x_0, h)$ defined for all $h \in \mathsf{V}$ such that the above limit holds for all h then this map need not be continuous or linear but if it is both then we say that f is **Gateaux differentiable at** x_0. In this case the linear map $h \mapsto \delta f(x_0, h)$ is written $\delta f|_{x_0}$ and is called the **Gateaux differential** of, or also the **first variation** of f at x_0.*

This is a rather general notion of derivative and is rather close to our informal definition in the case that V is some space of functions. When we presented the informal functional calculus we were, among other things, less than forthcoming about the topology on the function space.

We can generalize the notion of first variation: If $t \mapsto f(x_0 + th)$ is k-times differentiable at $t = 0$ then $\delta^k f\big|_{x_0}(h) := \frac{d}{dt}\big|_{t=0} f(x_0 + th)$ is called the k-th variation of f at x_0 in the direction h. In case the $\delta^k f\big|_{x_0}(h)$ exists for all h then $\delta^k f\big|_{x_0}$ is the k-th variation of f at x_0.

Clearly if a function f has a local maximum or minimum at x_0 and if it has a Gateaux derivative in a direction h then $\delta f(x_0, h) = 0$. This necessary condition for an extremum is certainly not a sufficient condition.

As we mentioned previously, the usual results from multivariable calculus generalize most easily and completely in the case where the domain and codomain spaces are Banach spaces but the definition of differentiability make sense for general normed spaces.

Definition 1.7 *A function $f : \mathsf{V} \supset U \to \mathsf{W}$ between normed spaces and defined on an open set $U \subset \mathsf{V}$ is said to be **differentiable at** $p \in U$ if and only if there is a bounded linear map $A_p \in L(\mathsf{V}, \mathsf{W})$ such that*

$$\lim_{\|h\| \to 0} \frac{\|f(p+h) - f(p) - A_p \cdot h\|}{\|h\|} = 0$$

*In anticipation of the following proposition we write $A_p = Df(p)$. We will also use the notation $Df|_p$ or sometimes $f'(p)$. The linear map $Df(p)$ is called the **derivative** of f at p. We often write $Df|_p \cdot h$ instead of $Df(p)(h)$. The dot in the notation just indicates a linear dependence and is not a literal "dot product".*

Exercise 1.4 *Show that the map $F : L^2(\Omega) \to L^1(\Omega)$ given by $F(f) = f^2$ is differentiable at any $f_0 \in L^2(\Omega)$. (Notice that the domain is a function space and we f is the independent variable while the function (functional) is not upper-case F).*

If A_p exists for a given function f then it is unique (see appendix G.0.10). As might be expected, if f is differentiable at x_0 then it also has a Gateaux derivative at x_0 and the two derivatives coincide: $Df|_{x_0} = \delta f|_{x_0}$.

If we are interested in differentiating "in one direction at a time" then we may use the natural notion of directional derivative. A map has a directional derivative $D_h f$ at p in the direction h if the following limit exists:

$$(D_h f)(p) := \lim_{\varepsilon \to 0} \frac{f(p + \varepsilon h) - f(p)}{\varepsilon}$$

In other words, $D_h f(p) = \frac{d}{dt}\big|_{t=0} f(p+th)$. But a function may have a directional derivative in every direction (at some fixed p), that is, for every $h \in \mathsf{E}$ and yet still not be differentiable at p in the sense of definition 1.7.

Notation 1.3 *The directional derivative is written as $(D_h f)(p)$ and, in case f is actually differentiable at p, this is equal to $Df|_p h = Df(p) \cdot h$ (the proof is easy). Look closely; $D_h f$ should not be confused with $Df|_h$.*

From now on we will restrict ourselves to the setting of Banach spaces unless otherwise indicated.

If it happens that a function f is differentiable for all p throughout some open set U then we say that f is differentiable on U. We then have a map $Df : U \subset \mathsf{V} \to L(\mathsf{V}, \mathsf{W})$ given by $p \mapsto Df(p)$. If this map is differentiable at some $p \in \mathsf{V}$ then its derivative at p is denoted $DDf(p) = D^2 f(p)$ or $D^2 f\big|_p$ and is an element of $L(\mathsf{V}, L(\mathsf{V}, \mathsf{W})) \cong L^2(\mathsf{V}; \mathsf{W})$. Similarly, we may inductively define $D^k f \in L^k(\mathsf{V}; \mathsf{W})$ whenever f is sufficiently nice that the process can continue.

Definition 1.8 *We say that a map $f : U \subset \mathsf{V} \to \mathsf{W}$ is C^r-differentiable on U if $D^r f|_p \in L^r(\mathsf{V}, \mathsf{W})$ exists for all $p \in U$ and if $D^r f$ is continuous as map $U \to L^r(\mathsf{V}, \mathsf{W})$. If f is C^r-differentiable on U for all $r > 0$ then we say that f is C^∞ or **smooth** (on U).*

Exercise 1.5 *Show directly that a bounded multilinear map is C^∞.*

Definition 1.9 *A bijection f between open sets $U_\alpha \subset \mathsf{V}$ and $U_\beta \subset \mathsf{W}$ is called a C^r-**diffeomorphism** if and only if f and f^{-1} are both C^r-differentiable (on U_α and U_β respectively). If $r = \infty$ then we simply call f a diffeomorphism.*

*Let U be open in V. A map $f : U \to \mathsf{W}$ is called a **local C^r diffeomorphism** if and only if for every $p \in U$ there is an open set $U_p \subset U$ with $p \in U_p$ such that $f|_{U_p} : U_p \to f(U_p)$ is a C^r-diffeomorphism.*

We will sometimes think of the derivative of a curve [3] $c : I \subset \mathbb{R} \to \mathsf{E}$ at $t_0 \in I$, written $\dot{c}(t_0)$, as a velocity vector and so we are identifying $\dot{c}(t_0) \in L(\mathbb{R}, \mathsf{E})$ with $Dc|_{t_0} \cdot 1 \in \mathsf{E}$. Here the number 1 is playing the role of the unit vector in $\mathbb{R}$.

Let $f : U \subset \mathsf{E} \to \mathsf{F}$ be a map and suppose that we have a splitting $\mathsf{E} = \mathsf{E}_1 \times \mathsf{E}_2$. Let (x, y) denote a generic element of $\mathsf{E}_1 \times \mathsf{E}_2$. Now for every $(a, b) \in U \subset \mathsf{E}_1 \times \mathsf{E}_2$ the partial maps $f_{a,} : y \mapsto f(a, y)$ and $f_{,b} : x \mapsto f(x, b)$ are defined in some neighborhood of b (resp. a). We define the partial derivatives, when they exist, by $D_2 f(a, b) = Df_{a,}(b)$ and $D_1 f(a, b) = Df_{,b}(a)$. These are, of course, linear maps.

$$D_1 f(a, b) : \mathsf{E}_1 \to \mathsf{F}$$
$$D_2 f(a, b) : \mathsf{E}_2 \to \mathsf{F}$$

The partial derivative can exist even in cases where f might not be differentiable in the sense we have defined. The point is that f might be differentiable only in certain directions. On the other hand, if f has continuous partial derivatives $D_i f(x, y) : \mathsf{E}_i \to \mathsf{F}$ near $(x, y) \in \mathsf{E}_1 \times \mathsf{E}_2$ then $Df(x, y)$ exists and is continuous. In this case, we have for $\mathrm{v} = (\mathrm{v}_1, \mathrm{v}_2)$,

$$\begin{aligned} &Df(x, y) \cdot (\mathrm{v}_1, \mathrm{v}_2) \\ &= D_1 f(x, y) \cdot \mathrm{v}_1 + D_2 f(x, y) \cdot \mathrm{v}_2. \end{aligned}$$

The reader will surely not be confused if we also write $\partial_1 f$ or $\partial_x f$ instead of $D_1 f$ (and similarly $\partial_2 f$ or $\partial_y f$ instead of $D_2 f$).

1.2.1 Chain Rule, Product rule and Taylor's Theorem

Theorem 1.1 (Chain Rule) *Let U_1 and U_2 be open subsets of Banach spaces E_1 and E_2 respectively. Suppose we have continuous maps composing as*

$$U_1 \xrightarrow{f} U_2 \xrightarrow{g} \mathsf{E}_3$$

[3] We will often use the letter I to denote a generic (usually open) interval in the real line.

where E_3 *is a third Banach space. If* f *is differentiable at* p *and* g *is differentiable at* $f(p)$ *then the composition is differentiable at* p *and* $D(g \circ f) = Dg(f(p)) \circ Dg(p)$. *In other words, if* $v \in \mathsf{E}_1$ *then*

$$D(g \circ f)|_p \cdot v = Dg|_{f(p)} \cdot (Df|_p \cdot v).$$

Furthermore, if $f \in C^r(U_1)$ *and* $g \in C^r(U_2)$ *then* $g \circ f \in C^r(U_1)$.

We will often use the following lemma without explicit mention when calculating:

Lemma 1.1 *Let* $f : U \subset \mathsf{V} \to \mathsf{W}$ *be twice differentiable at* $x_0 \in U \subset \mathsf{V}$; *then the map* $D_v f : x \mapsto Df(x) \cdot v$ *is differentiable at* x_0 *and its derivative at* x_0 *is given by*

$$D(D_v f)|_{x_0} \cdot h = D^2 f(x_0)(h, v).$$

Theorem 1.2 *If* $f : U \subset \mathsf{V} \to \mathsf{W}$ *is twice differentiable on* U *such that* $D^2 f$ *is continuous, i.e. if* $f \in C^2(U)$ *then* $D^2 f$ *is symmetric:*

$$D^2 f(p)(w, v) = D^2 f(p)(v, w).$$

More generally, if $D^k f$ *exists and is continuous then* $D^k f\ (p) \in L^k_{sym}(\mathsf{V}; \mathsf{W})$.

Theorem 1.3 *Let* $\varrho \in L(\mathsf{F}_1, \mathsf{F}_2; \mathsf{W})$ *be a bilinear map and let* $f_1 : U \subset \mathsf{E} \to \mathsf{F}_1$ *and* $f_2 : U \subset \mathsf{E} \to \mathsf{F}_2$ *be differentiable (resp.* $C^r, r \geq 1$*) maps. Then the composition* $\varrho(f_1, f_2)$ *is differentiable (resp.* $C^r, r \geq 1$*) on* U *where* $\varrho(f_1, f_2) : x \mapsto \varrho(f_1(x), f_2(x))$. *Furthermore,*

$$D\varrho|_x (f_1, f_2) \cdot v = \varrho(Df_1|_x \cdot v, f_2(x)) + \varrho(f_1(x), Df_2|_x \cdot v).$$

In particular, if F *is an algebra with differentiable product* $\star$ *and* $f_1 : U \subset \mathsf{E} \to \mathsf{F}$ *and* $f_2 : U \subset \mathsf{E} \to \mathsf{F}$ *then* $f_1 \star f_2$ *is defined as a function and*

$$D(f_1 \star f_2) \cdot v = (Df_1 \cdot v) \star (f_2) + (Df_1 \cdot v) \star (Df_2 \cdot v).$$

It will be useful to define an integral for maps from an interval $[a, b]$ into a Banach space V. First we define the integral for step functions. A function f on an interval $[a, b]$ is a **step function** if there is a partition $a = t_0 < t_1 < \cdots < t_k = b$ such that f is constant, with value say f_i, on each subinterval $[t_i, t_{i+1})$. The set of step functions so defined is a vector space. We define the integral of a step function f over $[a, b]$ by

$$\int_{[a,b]} f := \sum_{i=0}^{k-1} f(t_i)\Delta t_i$$

where $\Delta t_i := t_{i+1} - t_i$. One checks that the definition is independent of the partition chosen. Now the set of all step functions from $[a, b]$ into V is a linear subspace of the Banach space $\mathcal{B}(a, b, \mathsf{V})$ of all bounded functions of $[a, b]$ into V

and the integral is a linear map on this space. Recall that the norm on $\mathcal{B}(a,b,\mathsf{V})$ is $\sup_{a\leq t<b}\{f(t)\}$. If we denote the closure of the space of step functions in this Banach space by $\bar{\mathcal{S}}(a,b,\mathsf{V})$ then we can extend the definition of the integral to $\bar{\mathcal{S}}(a,b,\mathsf{V})$ by continuity since on step functions we have

$$\left|\int_{[a,b]} f\right| \leq (b-a)\left\|f\right\|_\infty .$$

In the limit, this bound persists. This integral is called the **Cauchy-Bochner** integral and is a bounded linear map $\bar{\mathcal{S}}(a,b,\mathsf{V}) \to \mathsf{V}$. It is important to notice that $\bar{\mathcal{S}}(a,b,\mathsf{V})$ contains the continuous functions $C([a,b],\mathsf{V})$ because such may be uniformly approximated by elements of $\mathcal{S}(a,b,\mathsf{V})$ and so we can integrate these functions using the Cauchy-Bochner integral.

Lemma 1.2 *If $\ell : \mathsf{V} \to \mathsf{W}$ is a bounded linear map of Banach spaces then for any $f \in \bar{\mathcal{S}}(a,b,\mathsf{V})$ we have*

$$\int_{[a,b]} \ell \circ f = \ell\left(\int_{[a,b]} f\right)$$

Proof. This is obvious for step functions. The general result follows by taking a limit for step functions converging in $\bar{\mathcal{S}}(a,b,\mathsf{V})$ to f. ∎

Let U be an open subset of $\mathbb{R}^n$ and let $f : U \to \mathbb{R}^m$ be a map that is differentiable at $a = (a^1,...,a^n) \in \mathbb{R}^n$. The map f is given by m functions $f^i : U \to \mathbb{R}$, $1 \leq i \leq m$. Now with respect to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, the derivative is given by an $n \times m$ matrix called the **Jacobian** matrix:

$$J_a(f) := \begin{pmatrix} \frac{\partial f^1}{\partial x^1}(a) & \frac{\partial f^1}{\partial x^2}(a) & \cdots & \frac{\partial f^1}{\partial x^n}(a) \\ \frac{\partial f^2}{\partial x^1}(a) & & & \frac{\partial f^2}{\partial x^n}(a) \\ \vdots & & \ddots & \\ \frac{\partial f^m}{\partial x^1}(a) & & & \frac{\partial f^m}{\partial x^n}(a) \end{pmatrix}.$$

The rank of this matrix is called the rank of f at a. If $n = m$ then the Jacobian is a square matrix and $\det(J_a(f))$ is called the **Jacobian determinant** at a. If f is differentiable near a then it follows from the inverse mapping theorem proved below that if $\det(J_a(f)) \neq 0$ then there is some open set containing a on which f has a differentiable inverse. The Jacobian of this inverse at $f(x)$ is the inverse of the Jacobian of f at x.

Notation 1.4 *The Jacobian matrix is tedious to write down. Of course we have the abbreviation $J_a(f)$ but let us also use the suggestive notation*

$$\frac{\partial(f^1,..,f^m)}{\partial(x^1,..,x^n)} := \begin{pmatrix} \frac{\partial f^1}{\partial x^1} & \frac{\partial f^1}{\partial x^2} & \cdots & \frac{\partial f^1}{\partial x^n} \\ \frac{\partial f^2}{\partial x^1} & & & \\ \vdots & & \ddots & \\ \frac{\partial f^m}{\partial x^1} & & & \frac{\partial f^m}{\partial x^n} \end{pmatrix}.$$

In case $m = n$, we do not use this notation to mean the determinant of this Jacobian matrix as is done by many authors. Instead we will use the (admittedly mysterious) notation

$$\frac{df^1 \wedge \cdots \wedge df^n}{dx^1 \wedge \cdots \wedge dx^n}(a) = \det(J_a(f))$$

The motivation for this notation will be made clear in the sequel.

The following is the **mean value theorem**:

Theorem 1.4 *Let V and W be Banach spaces. Let $c : [a, b] \to \mathsf{V}$ be a C^1-map with image contained in an open set $U \subset \mathsf{V}$. Also, let $f : U \to \mathsf{W}$ be a C^1 map. Then*

$$f(c(b)) - f(c(a)) = \int_0^1 Df(c(t)) \cdot c'(t)dt.$$

If $c(t) = (1 - t)x + ty$ then

$$f(y) - f(x) = \int_0^1 Df(c(t))dt \cdot (y - x).$$

Notice that $\int_0^1 Df(c(t))dt \in L(\mathsf{V}, \mathsf{W})$.

Corollary 1.1 *Let U be a convex open set in a Banach space V and $f : U \to \mathsf{W}$ a C^1 map into another Banach space W. Then for any $x, y \in U$ we have*

$$\|f(y) - f(x)\| \leq C_{x,y} \|y - x\|$$

where $C_{x,y}$ is the supremum over all values taken by f along the line segment that is the image of the path $t \mapsto (1 - t)x + ty$.

Recall that for a fixed x, higher derivatives $D^p f\big|_x$ are symmetric multilinear maps. For the following let $(y)^k$ denote $\underbrace{(y, y, ..., y)}_{k\text{-times}}$. With this notation we have the following version of Taylor's theorem.

Theorem 1.5 (Taylor's theorem) *Given Banach spaces V and W, a C^r function $f : U \to \mathsf{W}$ and a line segment $t \mapsto (1 - t)x + ty$ contained in U, we have that $t \mapsto D^p f(x + ty) \cdot (y)^p$ is defined and continuous for $1 \leq p \leq k$ and*

$$f(x + y) = f(x) + \frac{1}{1!}\, Df\big|_x \cdot y + \frac{1}{2!}\, D^2 f\big|_x \cdot (y)^2 + \cdots + \frac{1}{(k-1)!}\, D^{k-1} f\big|_x \cdot (y)^{k-1}$$
$$+ \int_0^1 \frac{(1 - t)^{k-1}}{(k - 1)!} D^k f(x + ty) \cdot (y)^k dt$$

The proof is by induction and follows the usual proof closely. The point is that we still have an integration by parts formula coming from the product rule and we still have the fundamental theorem of calculus.

1.2.2 Local theory of maps

Inverse Mapping Theorem

The main reason for restricting our calculus to Banach spaces is that the inverse mapping theorem holds for Banach spaces and there is no simple and general inverse mapping theory on more general topological vector spaces. The so called hard inverse mapping theorems such as that of Nash and Moser require special estimates and are constructed to apply only in a very controlled situation. Recently, Michor and Kriegl et. al. have promoted an approach that defines smoothness in terms of mappings of $\mathbb{R}$ into a very general class of topological vector spaces which makes a lot of the formal parts of calculus valid under their modified definition of smoothness. However, the general (and easy) inverse and implicit mapping theorems still remain limited as before to Banach spaces and more general cases have to be handled case by case.

Definition 1.10 *Let* E *and* F *be Banach spaces. A map will be called a* C^r ***diffeomorphism near*** p *if there is some open set* $U \subset \mathrm{dom}(f)$ *containing* p *such that* $f|_U : U \to f(U)$ *is a* C^r *diffeomorphism onto an open set* $f(U)$*. The set of all maps that are diffeomorphisms near* p *will be denoted* $\mathrm{Diff}^r_p(\mathsf{E},\mathsf{F})$*. If* f *is a* C^r *diffeomorphism near* p *for all* $p \in U = \mathrm{dom}(f)$ *then we say that* f *is a* ***local*** C^r ***diffeomorphism****.*

Lemma 1.3 *The space* $G\mathrm{L}(\mathsf{E},\mathsf{F})$ *of continuous linear isomorphisms is an open subset of the Banach space* $L(\mathsf{E},\mathsf{F})$*. In particular, if* $\|\mathrm{id} - A\| < 1$ *for some* $A \in G\mathrm{L}(\mathsf{E})$ *then* $A^{-1} = \lim_{N\to\infty} \sum_{n=0}^{N} (\mathrm{id} - A)^n$*.*

Lemma 1.4 *The map* $\mathcal{I} : G\mathrm{L}(\mathsf{E},\mathsf{F}) \to G\mathrm{L}(\mathsf{E},\mathsf{F})$ *given by taking inverses is a* C^∞ *map and the derivative of* $\mathcal{I} : g \mapsto g^{-1}$ *at some* $g_0 \in G\mathrm{L}(\mathsf{E},\mathsf{F})$ *is the linear map given by the formula:* $D\mathcal{I}|_{g_0} : A \mapsto -g_0^{-1} A g_0^{-1}$*.*

Theorem 1.6 (Inverse Mapping Theorem) *Let* E *and* F *be Banach spaces and* $f : U \to \mathsf{F}$ *be a* C^r *mapping defined an open set* $U \subset \mathsf{E}$*. Suppose that* $x_0 \in U$ *and that* $f'(x_0) = Df|_x : \mathsf{E} \to \mathsf{F}$ *is a continuous linear isomorphism. Then there exists an open set* $V \subset U$ *with* $x_0 \in V$ *such that* $f : V \to f(V) \subset \mathsf{F}$ *is a* C^r-*diffeomorphism. Furthermore the derivative of* f^{-1} *at* y *is given by* $Df^{-1}|_y = (Df|_{f^{-1}(y)})^{-1}$*.*

Corollary 1.2 *Let* $U \subset \mathsf{E}$ *be an open set and* $0 \in U$*. Suppose that* $f : U \to \mathsf{F}$ *is differentiable with* $Df(p) : \mathsf{E} \to \mathsf{F}$ *a (bounded) linear isomorphism for each* $p \in U$*. Then* f *is a local diffeomorphism.*

Theorem 1.7 (Implicit Function Theorem I) *Let* $\mathsf{E}_1, \mathsf{E}_2$ *and* F *be Banach spaces and* $U \times V \subset \mathsf{E}_1 \times \mathsf{E}_2$ *open. Let* $f : U \times V \to \mathsf{F}$ *be a* C^r *mapping such that* $f(x_0, y_0) = 0$*. If* $D_2 f_{(x_0,y_0)} : \mathsf{E}_2 \to \mathsf{F}$ *is a continuous linear isomorphism then there exists a (possibly smaller) open set* $U_0 \subset U$ *with* $x_0 \in U_0$ *and unique mapping* $g : U_0 \to V$ *with* $g(x_0) = y_0$ *and such that*

$$f(x, g(x)) = 0$$

for all $x \in U_0$.

Theorem 1.8 (Implicit Function Theorem II) *Let $\mathsf{E}_1, \mathsf{E}_2$ and F be Banach spaces and $U \times V \subset \mathsf{E}_1 \times \mathsf{E}_2$ open. Let $f : U \times V \to \mathsf{F}$ be a C^r mapping such that $f(x_0, y_0) = w_0$. If $D_2 f(x_0, y_0) : \mathsf{E}_2 \to \mathsf{F}$ is a continuous linear isomorphism then there exists (possibly smaller) open sets $U_0 \subset U$ and $W_0 \subset \mathsf{F}$ with $x_0 \in U_0$ and $w_0 \in W_0$ together with a unique mapping $g : U_0 \times W_0 \to V$ such that*

$$f(x, g(x, w)) = w$$

for all $x \in U_0$. Here unique means that any other such function h defined on a neighborhood $U_0' \times W_0'$ will equal g on some neighborhood of (x_0, w_0).

In the case of a map $f : U \to V$ between open subsets of Euclidean spaces (say $\mathbb{R}^n$ and $\mathbb{R}^m$) we have the notion of **rank at** $p \in U$ which is just the rank of the linear map $D_p f : \mathbb{R}^n \to \mathbb{R}^m$.

Notation 1.5 *A standard fact from linear algebra is that for a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ that is injective with rank r there exist linear isomorphisms $C_1 : \mathbb{R}^m \to \mathbb{R}^m$ and $C_2 : \mathbb{R}^n \to \mathbb{R}^n$ such that $C_1 \circ A \circ C_2^{-1}$ is just a projection followed by an injection:*

$$\mathbb{R}^n = \mathbb{R}^r \times \mathbb{R}^{n-r} \to \mathbb{R}^r \times 0 \to \mathbb{R}^r \times \mathbb{R}^{m-r} = \mathbb{R}^m$$

We have obvious special cases when $r = n$ or $r = m$. This fact has a local version that applies to C^∞ nonlinear maps. In order to facilitate the presentation of the following theorems we will introduce the following terminology:

Linear case.

Definition 1.11 *We say that a continuous linear map $A_1 : \mathsf{E}_1 \to \mathsf{F}_1$ is* ***equivalent*** *to a map $A_2 : \mathsf{E}_2 \to \mathsf{F}_2$ if there are continuous linear isomorphisms $\alpha : \mathsf{E}_1 \to \mathsf{E}_2$ and $\beta : \mathsf{F}_1 \to \mathsf{F}_2$ such that $A_2 = \beta \circ A_1 \circ \alpha^{-1}$.*

Definition 1.12 *Let $A : \mathsf{E} \to \mathsf{F}$ be an injective continuous linear map. We say that A is a* ***splitting injection*** *if there are Banach spaces F_1 and F_2 with $\mathsf{F} \cong \mathsf{F}_1 \times \mathsf{F}_2$ and if A is equivalent to the linear injection $\mathsf{E} \to \mathsf{F}_1 \times \mathsf{F}_2$ defined by $x \mapsto (x, 0)$.*

Lemma 1.5 *If $A : \mathsf{E} \to \mathsf{F}$ is a splitting injection as above then there exists a linear isomorphism $\delta : \mathsf{F} \to \mathsf{E} \times \mathsf{F}_2$ such that $\delta \circ A : \mathsf{E} \to \mathsf{E} \times \mathsf{F}_2$ is the injection $x \mapsto (x, 0)$.*

Proof. By definition there are isomorphisms $\alpha : \mathsf{E} \to \mathsf{F}_1$ and $\beta : \mathsf{F} \to \mathsf{F}_1 \times \mathsf{F}_2$ such that $\beta \circ A \circ \alpha^{-1}$ is the injection $\mathsf{F}_1 \to \mathsf{F}_1 \times \mathsf{F}_2$. Since α is an isomorphism we may compose as follows

$$\begin{aligned}
&(\alpha^{-1} \times \mathrm{id}_{\mathsf{E}}) \circ \beta \circ A \circ \alpha^{-1} \circ \alpha \\
&= (\alpha^{-1} \times \mathrm{id}_{\mathsf{E}}) \circ \beta \circ A \\
&= \delta \circ A
\end{aligned}$$

to get a map that is easily seen to have the correct form. ■

If A is a splitting injection as above it easy to see that there are closed subspaces F_1 and F_2 of F such that $\mathsf{F} = \mathsf{F}_1 \oplus \mathsf{F}_2$ and such that A maps E isomorphically onto F_1 .

Definition 1.13 *Let $A : \mathsf{E} \to \mathsf{F}$ be an surjective continuous linear map. We say that A is a* ***splitting surjection*** *if there are Banach spaces E_1 and E_2 with $\mathsf{E} \cong \mathsf{E}_1 \times \mathsf{E}_2$ and if A is equivalent to the projection $pr_1 : (x, y) \mapsto x$.*

Lemma 1.6 *If $A : \mathsf{E} \to \mathsf{F}$ is a splitting surjection then there is a linear isomorphism $\delta : \mathsf{F} \times \mathsf{E}_2 \to \mathsf{E}$ such that $A \circ \delta : \mathsf{F} \times \mathsf{E}_2 \to \mathsf{F}$ is the projection $(x, y) \mapsto x$.*

Proof. By definition there exist isomorphisms $\alpha : \mathsf{E} \to \mathsf{E}_1 \times \mathsf{E}_2$ and $\beta : \mathsf{F} \to \mathsf{E}_1$ such that $\beta \circ A \circ \alpha^{-1}$ is the projection $pr_1 : \mathsf{E}_1 \times \mathsf{E}_2 \to \mathsf{E}_1$. We form another map by composition of isomorphisms;

$$\begin{aligned} &\beta^{-1} \circ \beta \circ A \circ \alpha^{-1} \circ (\beta, \mathrm{id}_{\mathsf{E}_2}) \\ &= A \circ \alpha^{-1} \circ (\beta, \mathrm{id}_{\mathsf{E}_2}) := A \circ \delta \end{aligned}$$

and check that this does the job. ■

If A is a splitting surjection as above it easy to see that there are closed subspaces E_1 and E_2 of E such that $\mathsf{E} = \mathsf{E}_1 \oplus \mathsf{E}_2$ and such that A maps E onto E_1 as a projection $x + y \mapsto x$.

Local (nonlinear) case.

For the nonlinear case it is convenient to keep track of points. More precisely, we will be working in the category of local pointed maps. The objects of this category are pairs (U, p) where $U \subset \mathsf{E}$ a Banach space and $p \in \mathsf{E}$. The morphisms are maps of the form $f : (U, p) \to (V, q)$ where $f(p) = q$. As before, we shall let $f :: (\mathsf{E}, p) {\to} (\mathsf{F}, q)$ refer to such a map in case we do not wish to specify the opens sets U and V.

Let X, Y be topological spaces. When we write $f :: \mathsf{X} \to \mathsf{Y}$ we imply only that f is defined on some open set in X. We shall employ a similar use of the symbol "::" when talking about continuous maps between (open subsets of) topological spaces in general. If we wish to indicate that f is defined near $p \in \mathsf{X}$ and that $f(p) = q$ we will used the pointed category notation together with the symbol "::":

$$f :: (\mathsf{X}, p) \to (\mathsf{Y}, q)$$

We will refer to such a map as a **pointed local map**. Of course, every map $f : U \to V$ determines a pointed local map $f :: (\mathsf{X}, p) \to (\mathsf{Y}, f(p))$ for every $p \in U$. Notice that we use the same symbol f for the pointed map. This is a convenient abuse of notation and allows us to apply some of the present terminology to maps without explicitly mentioning pointed maps. Local maps may be composed with the understanding that the domain of the composite map may become smaller: If $f :: (\mathsf{X}, p) \to (\mathsf{Y}, q)$ and $g :: (\mathsf{Y}, q) \to (\mathsf{G}, z)$ then $g \circ f ::$

$(\mathsf{X},p) \to (\mathsf{G},z)$ and the domain of $g \circ f$ will be a non-empty open set containing p. Also, we will say that two such maps $f_1 : (\mathsf{X},p)\to(\mathsf{Y},q)$ and $f_2 : (\mathsf{X},p)\to(\mathsf{Y},q)$ are **equal near** p if there is an open set O with $p \in O \subset \mathrm{dom}(f_1) \cap \mathrm{dom}(f_2)$ such that the restrictions of these maps to O are equal:

$$f_1|_O = f_2|_O$$

in this case will simply write "$f_1 = f_2$ (near p)". We also say that f_1 and f_2 have the same **germ** at p.

Definition 1.14 *Let $f_1 ::\to (\mathsf{F}_1, q_1)$ be a pointed local map and $f_2 :: (\mathsf{E}_2, p_2) \to (\mathsf{F}_2, q_2)$ another such local map. We say that f_1 and f_2 are* ***(locally) equivalent*** *if there exist local diffeomorphisms $\alpha :: \mathsf{E}_1 \to \mathsf{E}_2$ and $\beta :: \mathsf{F}_1 \to \mathsf{F}_2$ such that $f_1 = \alpha \circ f_2 \circ \beta^{-1}$ (near p_1) or equivalently if $f_2 = \beta^{-1} \circ f_1 \circ \alpha$ (near p_2).*

$$\begin{array}{ccc} (\mathsf{E}_1,p_1) & \stackrel{\alpha}{\to} & (\mathsf{E}_2,p_2) \\ \downarrow & & \downarrow \\ (\mathsf{F}_1,q_1) & \stackrel{\beta}{\to} & (\mathsf{F}_2,q_2) \end{array}$$

Definition 1.15 *Let $f :: (\mathsf{E},p)\to(\mathsf{F},q)$ be a pointed local map. We say that f is a* ***locally splitting injection*** *(at p) or* ***local immersion*** *(at p) if there are Banach spaces F_1 and F_2 with $\mathsf{F} \cong \mathsf{F}_1 \times \mathsf{F}_2$ and if f is locally equivalent to the injection $\mathrm{inj}_1 :: (\mathsf{F}_1, 0) \to (\mathsf{F}_1 \times \mathsf{F}_2, 0)$.*

Lemma 1.7 *If f is a locally splitting injection at p as above there is an open set U_1 containing p and local diffeomorphism $\varphi : U_1 \subset \mathsf{F} \to U_2 \subset \mathsf{E} \times \mathsf{F}_2$ and such that $\varphi \circ f(\mathsf{x}) = (x, \mathsf{0})$ for all $x \in U_1$.*

Proof. This is done using the same idea as in the proof of lemma 1.5.

$$\begin{array}{cccc} & (\mathsf{E},p) & \to & (\mathsf{F},q) \\ \alpha & \updownarrow & & \updownarrow \beta \\ & (\mathsf{F}_1,0) & \to & (\mathsf{F}_1 \times \mathsf{F}_2, \quad (0,0)) \\ \alpha & \updownarrow & \mathrm{inj}_1 & \updownarrow \alpha^{-1} \times \mathrm{id} \\ & (\mathsf{E},p) & \to & (\mathsf{E} \times \mathsf{F}_2, \quad (p,0)) \end{array}$$

■

Definition 1.16 *Let $f :: (\mathsf{E},p) \to (\mathsf{F},q)$ be a pointed local map. We say that f is a* ***locally splitting surjection*** *or* ***local submersion*** *(at p) if there are Banach spaces E_1 and E_2 with $\mathsf{E} \cong \mathsf{E}_1 \times \mathsf{E}_2$ and if f is locally equivalent (at p) to the projection $pr_1 : \mathsf{E}_1 \times \mathsf{E}_2 \to \mathsf{E}_1$.*

Lemma 1.8 *If f is a locally splitting surjection as above there are open sets $U_1 \times U_2 \subset \mathsf{F} \times \mathsf{E}_2$ and $V \subset \mathsf{F}$ together with a local diffeomorphism $\varphi : U_1 \times U_2 \subset \mathsf{F} \times \mathsf{E}_2 \to V \subset \mathsf{E}$ such that $f \circ \varphi(u, v) = u$ for all $(u, v) \in U_1 \times U_2$.*

Proof. *This is the local (nonlinear) version of lemma 1.6 and is proved just as easily. Examine the diagram for guidance if you get lost:*

$$\begin{array}{ccc}
(\mathsf{E}, p) & \rightarrow & (\mathsf{F}, q) \\
\updownarrow & & \updownarrow \\
(\mathsf{E}_1 \times \mathsf{E}_2, (0,0)) & \rightarrow & (\mathsf{E}_1, 0) \\
\updownarrow & pr_1 & \updownarrow \\
(\mathsf{F} \times \mathsf{E}_2, (q,0)) & \rightarrow & (\mathsf{F}, q)
\end{array}$$

■

Theorem 1.9 (local immersion) *Let* $f :: (\mathsf{E}, p) \to (\mathsf{F}, q)$ *be a local map. If* $Df|_p : (\mathsf{E}, p) \to (\mathsf{F}, q)$ *is a splitting injection then* $f :: (\mathsf{E}, p) \to (\mathsf{F}, q)$ *is a local immersion at* p.

Theorem 1.10 (local immersion; finite dimensional case) *Let* $f :: \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^k)$ *be a map of constant rank* n *in some neighborhood of* $0 \in \mathbb{R}^n$. *Then there is* $g_1 :: \mathbb{R}^n \to \mathbb{R}^n$ *with* $g_1(0) = 0$, *and a* $g_2 :: \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^k$ *such that* $g_2 \circ f \circ g_1^{-1}$ *is just given by* $x \mapsto (x, 0) \in \mathbb{R}^n \times \mathbb{R}^k$ *.*

We have a similar but complementary theorem that we state in a slightly more informal manner.

Theorem 1.11 (local submersion) *Let* $f :: (\mathsf{E}, p) \to (\mathsf{F}, q)$ *be a local map. If* $Df|_p : (\mathsf{E}, p) \to (\mathsf{F}, q)$ *is a splitting surjection then* $f :: (\mathsf{E}, p) \to (\mathsf{F}, q)$ *is a local submersion.*

Theorem 1.12 (local submersion -finite dimensional case) *Let* $f :: (\mathbb{R}^n \times \mathbb{R}^k, 0) \to (\mathbb{R}^n, 0)$ *be a local map with constant rank* n *near* 0. *Then there are diffeomorphisms* $g_1 :: (\mathbb{R}^n \times \mathbb{R}^k, 0) \to (\mathbb{R}^n \times \mathbb{R}^k, 0)$ *and* $g_2 :: (\mathbb{R}^n, 0) \to (\mathbb{R}^n, 0)$ *such that near* 0 *the map*

$$g_2 \circ f \circ g_1^{-1} :: (\mathbb{R}^n \times \mathbb{R}^k, 0) \to (\mathbb{R}^n, 0)$$

is just the projection $(x, y) \mapsto x$.

If the reader thinks about what is meant by local immersion and local submersion they will realize that in each case the derivative map $Df(p)$ has full rank. That is, the rank of the Jacobian matrix in either case is a big as the dimensions of the spaces involved allow. Now rank is only semicontinuous and this is what makes full rank extend from points out onto neighborhoods so to speak. On the other hand, we can get more general maps into the picture if we explicitly assume that the rank is locally constant. We will state the following theorem only for the finite dimensional case.

Theorem 1.13 (The Rank Theorem) *Let* $f : (\mathbb{R}^n, p) \to (\mathbb{R}^m, q)$ *be a local map such that* Df *has constant rank* r *in an open set containing* p. *Then there are local diffeomorphisms* $g_1 : (\mathbb{R}^n, p) \to (\mathbb{R}^n, 0)$ *and* $g_2 : (\mathbb{R}^m, q) \to (\mathbb{R}^m, 0)$ *such that* $g_2 \circ f \circ g_1^{-1}$ *is a local diffeomorphism near* 0 *with the form*

$$(x^1, ...x^n) \mapsto (x^1, ...x^r, 0, ..., 0).$$

1.2.3 The Tangent Bundle of an Open Subset of a Banach Space

Later on we will define the notion of a tangent space and tangent bundle for a differentiable manifold that locally looks like a Banach space. Here we give a definition that applies to the case of an open set U in a Banach space.

Definition 1.17 *Let E be a Banach space and $U \subset \mathsf{E}$ an open subset. A **tangent vector** at $x \in U$ is a pair (x, v) where $v \in \mathsf{E}$. The **tangent space** at $x \in U$ is defined to be $T_xU := T_x\mathsf{E} := \{x\} \times \mathsf{E}$ and the tangent bundle TU over U is the union of the tangent spaces and so is just $TU = U \times \mathsf{E}$. Similarly the cotangent bundle over U is defined to be $T^*U = U \times \mathsf{E}^*$ where E^* is the dual space of bounded linear functionals on E.*

We give this definition in anticipation of our study of the tangent space at a point of a differentiable manifold. In this case however, it is often not necessary to distinguish between T_xU and E since we can often tell from context that an element $v \in \mathsf{E}$ is to be interpreted as based at some point $x \in U$. For instance a vector field in this setting is just a map $X : U \to \mathsf{E}$ but where $X(x)$ should be thought of as based at x.

Definition 1.18 *If $f : U \to \mathsf{F}$ is a C^r map into a Banach space F then the **tangent map** $Tf : TU \to T\mathsf{F}$ is defined by*

$$Tf \cdot (x, v) = (f(x), Df(x) \cdot v).$$

The map takes the tangent space $T_xU = T_x\mathsf{E}$ linearly into the tangent space $T_{f(x)}\mathsf{F}$ for each $x \in U$. The projection onto the first factor is written $\tau_U : TU = U \times \mathsf{E} \to U$ and given by $\tau_U(x, v) = x$. We also have a projection $\pi_U : T^*U = U \times \mathsf{E}^* \to U$ defined similarly.

If $f : U \to V$ is a diffeomorphism of open sets U and V in E and F respectively then Tf is a diffeomorphism that is linear on the fibers and such that we have a commutative diagram:

$$\begin{array}{cccc} TU = & U \times \mathsf{E} & \stackrel{Tf}{\to} & V \times \mathsf{F} = TV \\ pr_1 & \downarrow & & \downarrow pr_1 \\ & U & \to & V \\ & & f & \end{array}$$

The pair is an example of what is called a **local bundle map**. In this context we will denote the projection map $TU = U \times \mathsf{E} \to U$ by τ_U.

The chain rule looks much better if we use the tangent map:

Theorem 1.14 *Let U_1 and U_2 be open subsets of Banach spaces E_1 and E_2 respectively. Suppose we have differentiable (resp. $C^r, r \geq 1$) maps composing as*

$$U_1 \xrightarrow{f} U_2 \xrightarrow{g} \mathsf{E}_3$$

where E_3 *is a third Banach space. Then the composition is* $g \circ f$ *differentiable (resp.* $C^r, r \geq 1$*) and* $T(g \circ f) = Tg \circ Tf$

$$\begin{array}{ccccc} TU_1 & \stackrel{Tf}{\to} & TU_2 & \stackrel{Tg}{\to} & T\mathsf{E}_3 \\ \tau_{U_1} \downarrow & & \tau_{U_2} \downarrow & & \downarrow \tau_{E_3} \\ U_1 & \stackrel{f}{\to} & U_2 & \stackrel{g}{\to} & \mathsf{E}_3 \end{array}$$

Notation 1.6 *(and convention) There are three ways to express the "differential/derivative" of a differentiable map* $f : U \subset \mathsf{E} \to \mathsf{F}$*. These are depicted in figure* **??**.

1. *The first is just* $Df : \mathsf{E} \to \mathsf{F}$ *or more precisely* $Df|_x : \mathsf{E} \to \mathsf{F}$ *for any point* $x \in U$.

2. *This one is new for us. It is common but not completely standard :*

$$dF : TU \to \mathsf{F}$$

This is just the map $(x, v) \to Df|_x v$*. We will use this notation also in the setting of maps from manifolds into vector spaces where there is a canonical trivialization of the tangent bundle of the target manifold (all of these terms will be defined). The most overused symbol for various "differentials" is d.*

3. *Lastly, we have the tangent map* $Tf : TU \to T\mathsf{F}$ *which we defined above. This is the one that generalizes to manifolds without problems.*

In the local setting that we are studying now these three all contain essentially the same information so the choice to use one over the other is merely aesthetic.

It should be noted that some authors use *df* to mean any of the above maps and their counterparts in the general manifold setting. This leads to less confusion than one might think since one always has context on one's side.

1.2.4 Another look at the Functional Derivative

In this section we use the calculus we have developed in the setting of Banach spaces to provide one possible rigorous definition of the functional derivative of a differentiable functional $F : \mathsf{F} \supset U \to \mathbb{R}$. Actually for much of what follows, it would suffice to use the weaker notion of derivative at $g \in U \subset \mathsf{F}$ which requires only that all the directional derivatives exist and that the linear map $h \mapsto D_h F(p)$ is bounded.

One meaning for the phrase "functional derivative" is just the ordinary derivative of a function on a Banach space of functions (Or more generally, the Gateaux derivative on some a more general normed space). There is another meaning, a meaning more in line with the informal functional calculus.

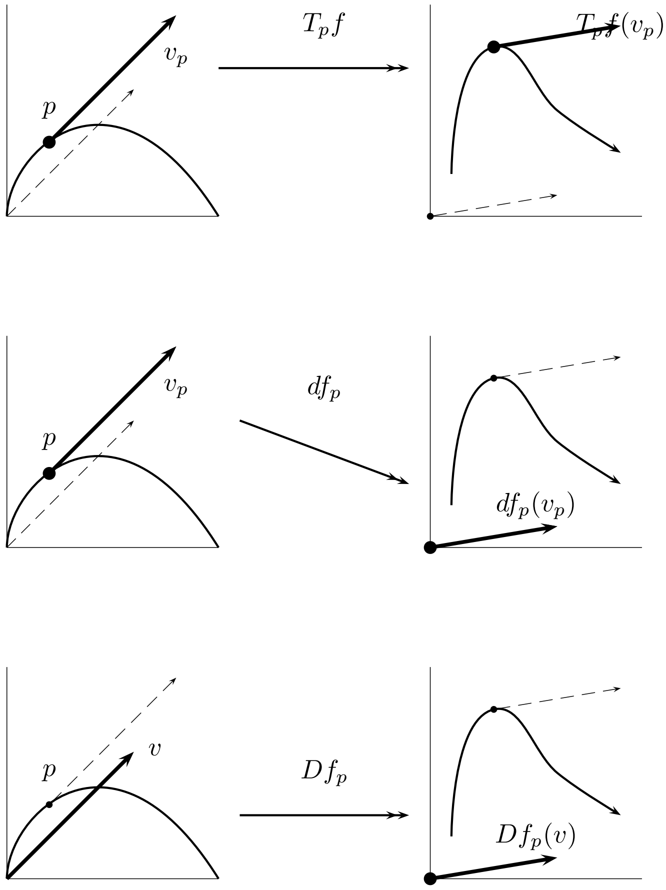


Figure 1.1: Hello

Roughly, we take the functional derivative of a functional, say F, to be is whatever object $\frac{\delta F}{\delta \varphi(t)}$ it is that occurs in expressions like $\delta F = \int \frac{\delta F}{\delta \varphi(t)} \delta h(t)\, dt$. Thus, from this point of view $\frac{\delta F}{\delta \varphi(t)}$ is a sort of integral kernel for the derivative. Thus $\frac{\delta F}{\delta \varphi(t)}$ refers to a function $\frac{\delta F}{\delta \varphi} : t \mapsto \frac{\delta F}{\delta \varphi(t)}$ whose value at t would have normally been denoted as $\frac{\delta F}{\delta \varphi}(t)$. Of course, we could just take $\frac{\delta F}{\delta \varphi}$ in a distributional sense but then it is not clear what we have gained. It is also not clear to the author what should be the appropriate ultimate definition of the functional derivative. Notice that we have taken the derivative of F at some φ. Presumably, this is the φ in the denominator of $\frac{\delta F}{\delta \varphi}$. One approach to making sense out of $\frac{\delta F}{\delta \varphi(t)}$ that works for a fairly large number of cases is that found in [A,B,R]. We now quickly explain this approach. The reader should be aware that this is not the only approach, nor the most general. Also, $\frac{\delta F}{\delta \varphi}$ is mainly a mere notation that must be faced in some of the literature and is not really a deep concept. For a given function F defined on a normed vector space, the very meaning of $\frac{\delta F}{\delta \varphi}$ depends on a choice of pairing which is usually given by some integration such as $\langle f, g \rangle := \int f(x) g(x) dx$. We generalize a bit:

Definition 1.19 *A continuous bilinear mapping $\langle ., . \rangle : \mathsf{E} \times \mathsf{F} \to \mathbb{R}$ is said to be $\mathsf{E}-$**nondegenerate** if $\langle v, w \rangle = 0$ for all w always implies that $v = 0$. Similarly, $\langle ., . \rangle$ is $\mathsf{F}-$**nondegenerate** if $\langle v, w \rangle = 0$ for all v implies that $w = 0$. If the map $v \mapsto \langle v, . \rangle \in \mathsf{F}^*$ is an isomorphism we say that the bilinear map is $\mathsf{E}-$**strongly nondegenerate**. Similarly for $\mathsf{F}-$**strongly nondegenerate**. If $\langle ., . \rangle : \mathsf{E} \times \mathsf{F} \to \mathbb{R}$ is both $\mathsf{E}-$nondegenerate and $\mathsf{F}-$nondegenerate we say that $\langle ., . \rangle$ is a nondegenerate pairing*

We sometimes refer to a bilinear map such as in the last definition as a **pairing** of E and F, especially if it is nondegenerate (both sides).

Definition 1.20 *Let $\langle ., . \rangle : \mathsf{E} \times \mathsf{F} \to \mathbb{R}$ be a fixed $\mathsf{E}-$nondegenerate pairing. If $F : U \subset \mathsf{F} \to \mathbb{R}$ is differentiable at $f \in \mathsf{F}$ then the unique element $\frac{\delta F}{\delta f} \in \mathsf{E}$ such that $\langle \frac{\delta F}{\delta f}, . \rangle = DF|_f$ is called the functional derivative of F.*

In the special (but common) case that both E and F are spaces of functions defined on a common measure space and if the pairing is given by integration $\langle h, g \rangle := \int h(x) g(x) \mu(dx)$ with respect to some measure μ, then we sometimes write $\frac{\delta F}{\delta f(x)}$ instead of $\frac{\delta F}{\delta f}(x)$. The only reason for this is to harmonize with existing notation used in the naive versions of the functional derivative found in the physics literature.

Example 1.3 *Let D be a compact domain in $\mathbb{R}^n$ and let E and F both be the Banach space $C^0(D)$. We use the pairing $\langle f, g \rangle = \int_D f(x) g(x) dx$. For our example of a function (functional) on F we take F defined by*

$$F(f) := \int_D (f(x))^k dx$$

Then is easy to see that for any $f \in C^0(D)$ we have $DF(f) \cdot h = \int_D k(f(x))^{k-1} h(x) dx$ and so $\frac{\delta F}{\delta f(x)} = k(f(x))^{k-1}$.

Since the pairing might not be strong it is possible that the functional derivative does not exist. However, it is often the case that there is a dense subspace F_0 such that

$$DF(f) \cdot h = \langle \frac{\delta F}{\delta f}, h \rangle \text{ for all } h \in \mathsf{F}_0$$

for some $\frac{\delta F}{\delta f} \in \mathsf{E}$. In this case we should really specify the fact that $\frac{\delta F}{\delta f}$ only does the job on the small space. It is not standard but one could say that "$\frac{\delta F}{\delta f}$ is the functional derivative of F (rel F_0)". The following is obvious:

Criterion 1.1 *Suppose that $\frac{\delta F}{\delta f}$ is the functional derivative of F (rel F_0). Since F_0 is dense we see that $DF(f) \equiv 0$ if and only if $\frac{\delta F}{\delta f} = 0$. If F has a relative maximum or minimum at f then $\frac{\delta F}{\delta f} = 0$.*

$\frac{\delta F}{\delta f} = 0$ is usually a differential equation and is a general Euler-Lagrange equation.

Example 1.4 *Let E and F both be $C^2(D)$ and F_0 be $C_c^2(D)$ and use the pairing $\langle f, g \rangle = \int_D f(x)g(x)dx$. Now consider the functional*

$$F(f) := \frac{1}{2} \int_D \nabla f \cdot \nabla f \, dx$$

We shall discover the functional derivative by calculation of an arbitrary directional derivative

$$\begin{aligned}
&\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} F(f + \varepsilon h) \\
&= \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \frac{1}{2} \int_D \nabla(f(x) + \varepsilon h(x)) \cdot \nabla(f(x) + \varepsilon h(x)) dx \\
&= \int_D \nabla f(x) \cdot \nabla h(x) dx \\
&= -\int_D \nabla^2 f(x) h(x) dx \text{ (boundary terms disappear)} \\
&= \langle -\nabla^2 f, h \rangle \text{ (for } h \in C_c^2(D))
\end{aligned}$$

Thus $-\nabla^2 f = \frac{\delta F}{\delta f}$ rel $C_c^2(D)$.

As a final example of some hieroglyphics from the informal functional calculus consider the following version of Taylor's formula for a real valued functional F defined on functions whose domain is, for example, an interval:

$$\begin{aligned}
F(f + \delta f) = F(f) &+ \int \frac{\delta F}{\delta f(x)} \delta f(x) dx + \iint \frac{\delta^2 F}{\delta f(x) \delta f(y)} \delta f(x) \delta f(y) \\
&+ \iiint \frac{\delta^3 F}{\delta f(x) \delta f(y) \delta f(z)} \delta f(x) \delta f(y) \delta f(z) dx dy dz + \ldots.
\end{aligned}$$

Clearly, the expression implies the existence of integral kernels for higher derivatives and the expression is likely to be valid only if δf satisfies some condition (compact support) which would nullify boundary terms in any integration by parts. This seems likely to be unnecessary restriction in a generic situation and the superiority of the Banach space version of Taylor's formula is evident.

1.2.5 Extrema

A real valued function f on a topological space X is continuous at $x_0 \in X$ if for every $\epsilon > 0$ there is a open set U_ϵ containing x_0 such that

$$U_\epsilon \subset \{x : f(x) > f(x_0) - \epsilon\} \cap \{x : f(x) < f(x_0) + \epsilon\}.$$

Baire introduced the notion of semicontinuity by essentially using only one of the intersected sets above. Namely, f is lower semicontinuous at x_0 if for every $\epsilon > 0$ there is a open set U_ϵ containing x_0 such that

$$U_\epsilon \subset \{x : f(x) > f(x_0) - \epsilon\}.$$

Of course there is the symmetrical notion of upper semicontinuity. Lower semicontinuity is appropriately introduced in connection with the search for x such that $f(x)$ is a (local) minimum. Since replacing f by $-f$ interchanges upper and lower semicontinuity and also maxima and minima it we be sufficient to limit the discussion to minima. If the topological space is Hausdorff then we can have a simple and criterion for the existence of a minimum:

Theorem 1.15 *Let M be Hausdorff and $f : M \to \mathbb{R} \cup \{+\infty\}$ a lower semicontinuous function such that there exists a number $c < \infty$ such that $M_{f,c} := \{x \in M : f(x) \leq c\}$ is nonempty and sequentially compact then there exists a minimizing $x_0 \in M$:*

$$f(x_0) = \inf_{x \in M} f(x).$$

This theorem is a sort of generalization of the theorem that a continuous function on a compact set achieve a minimum (and maximum). We need to include semicontinuous functions since even some of the most elementary examples for geometric minimization problems naturally involve functionals that are only semicontinuous for very natural topologies on the set M.

Now if M is a convex set in some vector space then if $f : M \to \mathbb{R}$ is strictly convex on M (meaning that $0 < t < 1 \implies f(tx_1 + (1-t)x_2) < tf(x_1) + (1-t)f(x_2)$) then a simple argument shows that f can have at most one minimizer x_0 in M. We are thus led to a wish list in our pursuit of a minimum for a function. We hope to simultaneously have

1. M convex,
2. $f : M \to \mathbb{R}$ strictly convex,
3. f lower semicontinuous,

4. there exists $c < \infty$ so that $M_{f,c}$ is nonempty and sequentially compact.

Just as in elementary calculus, a differentiable function f has an extrema at a point x_0 only if x_0 is a boundary point of the domain of f or, if x_0 is an interior point, $Df(x_0) = 0$. In case $Df(x_0) = 0$ for some interior point x_0, we may not jump to the conclusion that f achieves an extrema at x_0. As expected, there is a second derivative test that may help in this case:

Theorem 1.16 *Let $f : U \subset \mathsf{E} \to \mathsf{F}$ be twice differentiable at $x_0 \in U$ (assume U is open). If f has a local extrema at x_0 then $D^2 f(x_0)(v, v) \geq 0$ for all $v \in \mathsf{E}$. If $D^2 f(x_0)$ is a nondegenerate bilinear form then if $D^2 f(x_0)(v, v) > 0$ (resp. < 0) then f achieves a minimum (resp. maximum) at x_0.*

In practice we may have a choice of several possible sets M, topologies for M or the vector space contains M and so on. But then we must strike a balance since and topology with more open sets has more lower semicontinuous functions while a less open sets means more sequentially compact sets.

1.2.6 Lagrange Multipliers and Ljusternik's Theorem

`Note: This section is under construction. It is still uncertain if this material will be included at all.`

The next example show how to use Lagrange multipliers to handle constraints.

Example 1.5 *Let E and F and F_0 be as in the previous example. We define two functionals*

$$\mathcal{F}[f] := \int_D \nabla f \cdot \nabla f dx$$
$$\mathcal{C}[f] = \int_D f^2 dx$$

We want a necessary condition on f such that f extremizes $\mathcal{D}$ subject to the constraint $\mathcal{C}[f] = 1$. The method of Lagrange multipliers applies here and so we have the equation $D\mathcal{F}|_f = \lambda \ D\mathcal{C}|_f$ which means that

$$\langle \frac{\delta \mathcal{F}}{\delta f}, h \rangle = \lambda \langle \frac{\delta \mathcal{C}}{\delta f}, h \rangle \text{ for all } h \in C_c^2(D)$$
$$\text{or}$$
$$\frac{\delta \mathcal{F}}{\delta f} = \lambda \frac{\delta \mathcal{C}}{\delta f}$$

After determining the functional derivatives we obtain

$$-\nabla^2 f = \lambda f$$

This is not a very strong result since it is only a necessary condition and only hints at the rich spectral theory for the operator ∇^2.

Theorem 1.17 *Let* E *and* F *be Banach spaces and* $U \subset \mathsf{E}$ *open with a differentiable map* $f : U \to \mathsf{F}$*. If for* $x_0 \in U$ *with* $y_0 = f(x_0)$ *we have that* $Df|_{x_0}$ *is onto and* $\ker Df|_{x_0}$ *is complemented in* E *then the set* $x_0 + \ker Df|_{x_0}$ *is tangent to the level set* $f^{-1}(y_0)$ *in the following sense: There exists a neighborhood* $U' \subset U$ *of* x_0 *and a homeomorphism* $\phi : U' \to V$ *where* V *is another neighborhood of* x_0 *and where* $\phi(x_0 + h) = x_0 + h + \varepsilon(h)$ *for some continuous function* ε *with the property that*

$$\lim_{h\to 0} \frac{\|\varepsilon(h)\|}{\|h\|} = 0.$$

Proof. $Df|_{x_0}$ is surjective. Let $K := \ker Df|_{x_0}$ and let L be the complement of K in E. This means that there are projections $p : \mathsf{E} \to K$ and $q : \mathsf{E} \to L$

$$\begin{aligned} p^2 &= p \text{ and } q^2 = q \\ p + q &= id \end{aligned}$$

Let $r > 0$ be chosen small enough that $x_0 + B_r(0) + B_r(0) \subset U$. Define a map

$$\psi : K \cap B_r(0) \times L \cap B_r(0) \to \mathsf{F}$$

by $\psi(h_1, h_2) := f(x_0 + h_1 + h_2)$ for $h_1 \in K \cap B_r(0)$ and $h_2 \in L \cap B_r(0)$. We have $\psi(0,0) = f(x_0) = y_0$ and also one may verify that ψ is C^1 with $\partial_1\psi = Df(x_0)|\, K = 0$ and $\partial_2\psi = Df(x_0)|\, L$. Thus $\partial_2\psi : L \to \mathsf{F}$ is a continuous isomorphism (use the open mapping theorem) and so we have a continuous linear inverse $(\partial_2\psi)^{-1} : \mathsf{F} \to L$. We may now apply the implicit function theorem to the equation $\psi(h_1, h_2) = y_0$ to conclude that there is a locally unique function $\varepsilon : K \cap B_\delta(0) \to L$ for small $\delta > 0$ (less than r) such that

$$\begin{aligned} \psi(h, \varepsilon(h)) &= y_0 \text{ for all } h \in K \cap B_\delta(0) \\ \varepsilon(0) &= 0 \\ D\varepsilon(0) &= -\left(\partial_2\psi\right)^{-1} \circ \partial_1\psi|_{(0,0)} \end{aligned}$$

But since $\partial_1\psi = Df(x_0)|\, K = 0$ this last expression means that $D\varepsilon(0) = 0$ and so

$$\lim_{h\to 0} \frac{\|\varepsilon(h)\|}{\|h\|} = 0$$

Clearly the map $\phi : (x_0 + K \cap B_\delta(0)) \to \mathsf{F}$ defined by $\phi(x_0 + h) := x_0 + h + \varepsilon(h)$ is continuous and also since by construction $y_0 = \psi(h, \varepsilon(h)) = \phi(x_0 + h + \varepsilon(h))$ we have that ϕ has its image in $f^{-1}(y_0)$. Let the same symbol ϕ denote the map $\phi : (x_0 + K \cap B_\delta(0)) \to f^{-1}(y_0)$ which only differs in its codomain. Now h and $\varepsilon(h)$ are in complementary subspaces and so ϕ must be injective. Thus its restriction to the set $V := \{x_0 + h + \varepsilon(h) : h \in K \cap B_\delta(0)$ is invertible and in fact we have $\phi^{-1}(x_0 + h + \varepsilon(h)) = x_0 + h$. That V is open follows from the way we have used the implicit function theorem. Now recall the projection p. Since the range of p is K and its kernel is L we have that $\phi^{-1}(x_0 + h + \varepsilon(h)) = x_0 + p(h + \varepsilon(h))$

and we see that ϕ^{-1} is continuous on V. Thus ϕ (suitably restricted) is a homeomorphism of $U' := x_0 + K \cap B_\delta(0)$ onto $V \subset f^{-1}(y_0)$. We leave it to the reader to provide the easy verification that ϕ has the properties claimed by statement of the theorem. ■

1.3 Problem Set

1. Find the matrix that represents the derivative the map $f : \mathbb{R}^n \to \mathbb{R}^m$ given by

   a) $f(x) = Ax$ for an $m \times n$ matrix A.

   b) $f(x) = x^t A x$ for an $n \times n$ matrix A (here $m = 1$)

   c) $f(x) = x^1 x^2 \cdots x^n$ (here $m = 1$)

2. Find the derivative of the map $F : L^2([0,1]) \to L^2([0,1])$ given by

$$F[f](x) = \int_0^1 k(x,y)\,[f(y)]^2\,dy$$

   where $k(x,y)$ is a bounded continuous function on $[0,1] \times [0,1]$.

3. Let $L : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ be C^∞ and define

$$S[c] = \int_0^1 L(c(t), c'(t), t)dt$$

   which is defined on the Banach space B of all C^1 curves $c : [0,1] \to R^n$ with $c(0) = 0$ and $c(1) = 0$ and with the norm $\|c\| = \sup_{t\in[0,1]}\{|c(t)| + |c'(t)|\}$. Find a function $g_c : [0,1] \to R^n$ such that

$$DS|_c \cdot b = \int_0^1 \langle g_c(t), b(t)\rangle dt$$

   or in other words,

$$DS|_c \cdot b = \int_0^1 \sum_{i=1}^n g_c^i(t) b^i(t) dt.$$

4. In the last problem, if we hadn't insisted that $c(0) = 0$ and $c(1) = 0$, but rather that $c(0) = x_0$ and $c(1) = x_1$, then the space wouldn't even have been a vector space let alone a Banach space. But this fixed endpoint family of curves is exactly what is usually considered for functionals of this type. Anyway, convince yourself that this is not a serious problem by using the notion of an affine space (like a vector space but no origin and only differences are defined.). Is the tangent space of the this space of fixed endpoint curves a Banach space?

Hint: If we choose a fixed curve c_0 which is the point in the Banach space at which we wish to take the derivative then we can write $\mathcal{B}_{\vec{x}_0\vec{x}_1} = \mathcal{B} + c_0$ where

$$\begin{aligned} \mathcal{B}_{\vec{x}_0\vec{x}_1} &= \{c : c(0) = \vec{x}_0 \text{ and } c(1) = \vec{x}_1\} \\ \mathcal{B} &= \{c : c(0) = 0 \text{ and } c(1) = 0\} \end{aligned}$$

Then we have $T_{c_0}\mathcal{B}_{\vec{x}_0\vec{x}_1} \cong \mathcal{B}$. Thus we should consider $\left.DS\right|_{c_0} : \mathcal{B} \to \mathcal{B}$.

5. Let $\mathrm{Fl}_t(.)$ be defined by $\mathrm{Fl}_t(x) = (t+1)x$ for $t \in (-1/2, 1/2)$ and $x \in \mathbb{R}^n$. Assume that the map is jointly C^1 in both variable. Find the derivative of
$$f(t) = \int_{D(t)} (tx)^2 \, dx$$
at $t = 0$, where $D(t) := \mathrm{Fl}_t(D)$ the image of the disk $D = \{|x| \leq 1\}$.

   Hint: Find the Jacobian $J_t := \det[DFl_t(x)]$ and then convert the integral above to one just over $D(0) = D$.

6. Let $\mathrm{GL}(n, \mathbb{R})$ be the nonsingular $n \times n$ matrices and show that $\mathrm{GL}(n, \mathbb{R})$ is an open subset of the vector space of all matrices $\mathbb{M}_{n\times n}(\mathbb{R})$ and then find the derivative of the determinant map: $\det : \mathrm{GL}(n, \mathbb{R}) \to \mathbb{R}$ (for each A this should end up being a linear map $\left.D\det\right|_A : \mathbb{M}_{n\times n}(\mathbb{R}) \to \mathbb{R}$).

   What is $\frac{\partial}{\partial x_{ij}} \det X$ where $X = (x_{ij})$?

7. Let $A : U \subset \mathsf{E} \to L(\mathsf{F}, \mathsf{F})$ be a C^r map and define $F : U \times \mathsf{F} \to \mathsf{F}$ by $F(u, f) := A(u)f$. Show that F is also C^r.

   Hint: Leibniz rule theorem.

8. Show that if F is any closed subset of $\mathbb{R}^n$ there is a C^∞-function f whose zero set $\{x : f(x) = 0\}$ is exactly F.

9. Let U be an open set in $\mathbb{R}^n$. For $f \in C^k(U)$ and $S \subset U$ a compact set, let $\|f\|_k^S := \sum_{|\alpha|\leq k} \sup_{x\in S} \left|\frac{\partial^{|\alpha|} f}{\partial x^\alpha}(x)\right|$. a) Show that (1) $\|rf\|_k^S = |r| \, \|f\|_k^S$ for any $r \in \mathbb{R}$, (2) $\|f_1 + f_2\|_k^S \leq \|f_1\|_k^S + \|f_2\|_k^S$ for any $f_1, f_2 \in C^k(U)$, (3) $\|fg\|_k^S \leq \|f\|_k^S \, \|g\|_k^S$ for $f, g \in C^k(U)$.
   b) Let $\{K_i\}$ be a compact subsets of U such that $U = \bigcup_i K_i$. Show that $d(f, g) := \sum_{i=0}^\infty \frac{1}{2^i} \frac{\|f-g\|_k^{K_i}}{1+\|f-g\|_k^{K_i}}$ defines a complete metric space structure on $C^k(U)$.

10. Let E and F be real Banach spaces. A function $f : \mathsf{E} \to \mathsf{F}$ is said to be homogeneous of degree k if $f(rx) = rf(x)$ for all $r \in \mathbb{R}$ and $x \in \mathsf{E}$. Show that if f is homogeneous of degree k and is differentiable, then $Df(v) \cdot v = kf(v)$.

Chapter 2

Appetizers

The art of doing mathematics consists in finding that special case which contains all the germs of generality.

David Hilbert (1862-1943).

This chapter is about elementary examples and ideas that serve to reveal a bit of "spirit of geometry"; particularly, the role of groups and group actions in geometry. Although I assume the reader has some background in linear algebra and basic group theory, I have found it convenient to review some definitions and notations. In the first few paragraphs I will also make several references to things that are not studied in earnest until later in the book and the reader is advised to return to this chapter for a second reading after having penetrated further into the text. In order that the reader will not feel completely at loss we shall provide a short glossary with purely heuristic definitions:

1. *Differentiable manifold*: A topological space that supports coordinate systems which are so related to each other that we may give meaning to the differentiability of functions defined on the space. A notion central to all that we do. Examples include the Cartesian spaces $\mathbb{R}^n$, surfaces in $\mathbb{R}^3$, the projective plane, the torus, etc. Differentiable manifolds are studied systematically starting in chapter 3.

2. *Lie group*: A group that is also a differentiable manifold. The group multiplication is required to be differentiable (actually real analytic). Most Lie groups are isomorphic to a group of matrices such as the general linear group $\mathrm{GL}(n)$ or the group of orthogonal matrices. Our proper study of Lie groups begins with chapter 5.

3. *Tangent vectors and vector fields*: A tangent vector at a point on a manifold is a generalization of a vector tangent to a surface. It is the infinitesimal (first order) analogue of a curve in the surface through the point. Also to be thought of as the velocity of a curve through that point.

   A vector field is a (smooth) assignment of a tangent vector to each point on a manifold and might, for example, model the velocity field of a fluid flow. Just as a vector is the velocity of a curve through a point, so a vector field is the "velocity" of a smooth family of deformations of the manifold. Tangent vectors and vector fields are introduced in chapter 4 and studied in further depth in chapter 7.

4. *A differential $k-$ form:* A differential $1-$form at a point encodes infinitesimal information which can be thought of as the infinitesimal version of a function defined near the point. A differentiable 1-form is a (smooth) assignment of a $1-$form to each point of a differentiable manifold. Differentiable $1-$forms are introduced in chapter 4and the more general notion of a differentiable $k-$forms is studied in chapter 8. Differential $1-$forms are objects which may be integrated over parameterized curves, differential 2 forms may be integrated over parameterized surfaces. Generally, a differentiable $k-$form may be integrated over parameterized $k-$dimensional spaces.

5. *Lie algebra*: Every Lie group has an associated Lie algebra. A Lie algebra is a vector space together with a special kind of product called the bracket. For matrix groups the Lie algebras are all just subspaces of matrices with the bracket being the commutator $[A, B] = AB - BA$. For example, the Lie algebra of the group of the rotation group $SO(3)$ is the space of all skewsymmetric 3×3 matrices and is denoted $\mathfrak{so}(3)$. The Lie algebra of a Lie group is a sort of first order approximation to the group near the identity element. Because Lie groups have so much structure it turns out that the Lie algebra contains almost all information about the group itself.

6. An $\mathfrak{so}(n)$-valued $1-$form may be though of as a skewsymmetric matrix of $1-$forms and plays an important role in the local study of covariant derivatives. The gauge potentials in particle physics are also matrix valued $1-$forms.

7. *Metric*: A metric tensor is an extra piece of structure on a differentiable manifold that allows the measurement of lengths of vectors and then, by integration, the lengths of curves. In a given coordinate system, a metric is given by a symmetric matrix of functions.

8. *Riemannian manifold*: A differentiable manifold together with a metric. Central object for Riemannian geometry-See chapter 20.

9. *Isometry group*: The isometry group of a Riemannian manifold is the group of metric preserving (hence length and volume preserving) maps from the manifold to itself. A generic Riemannian manifold has a trivial isometry group.

"What is Geometry?". Such is the title of a brief expository article ([Chern1]) by Shiing-Shen Chern -one of the giants of 20th century geometry. In the article, Chern does not try to answer the question directly but rather identifies a few of the key ideas which, at one time or another in the history of geometry, seem to have been central to the very meaning of geometry. It is no surprise that the first thing Chern mentions is the axiomatic approach to geometry which was the method of Euclid. Euclid did geometry without the use of coordinates and proceeded in a largely logico-deductive manner supplemented by physical argument. This approach is roughly similar to geometry as taught in (U.S.) middle and high schools and so I will not go into the subject. Suffice it to say that the subject has a very different flavor as compared modern approaches such as differential geometry, algebraic geometry and group theoretic geometry[1]. Also, even though it is commonly thought that Euclid though of geometry as a purely abstract discipline, it seems that for Euclid geometry was the study of a physical reality. The idea that geometry had to conform to physical space persisted far after Euclid and this prejudice made the discovery of non-Euclidean geometries a difficult process.

Next in Chern's list of important ideas is Descarte's idea of introducing coordinates into geometry. As we shall see, the use of (or at least existence of) coordinates is one of the central ideas of differential geometry. Descarte's coordinate method put the algebra of real numbers to work in service of geometry. A major effect of this was to make the subject easier (thank you, Descarte!). Coordinates paved the way for the differential calculus and then the modern theory of differentiable manifolds which will be a major topic of the sequel.

It was Felix Klein's program (Klein's Erlangen Programm) that made groups and the ensuing notions of symmetry and invariance central to the very meaning of geometry. This approach was advanced also by Cayley and Cartan. Most of this would fit into what is now referred to as the geometry of homogeneous spaces. Homogenous spaces are not always studied from a purely geometric point of view and, in fact, they make an appearance in many branches of mathematics and physics. Homogeneous spaces also take a central position within the theory of Lie group representation theory and Harmonic analysis[2].

These days when one thinks of differential geometry it is usually Riemannian geometry. Riemann's approach to geometry differs from that of Klein's and at first it is hard to find common ground outside of the seemingly accidental fact that some homogeneous spaces carry a natural Riemannian structure. The basic object of Riemannian geometry is a Riemannian manifold. On a Riemannian manifold length and distance are defined using first the notion of lengths of tangent vectors and paths. A Riemannian manifold may well have a trivial symmetry group (isometry group). This would seem to put group theory in the back seat unless we happen to be studying highly symmetric spaces like "Riemannian homogeneous spaces" or "Riemannian symmetric spaces" such as the sphere or Euclidean space itself. Euclidean geometry is both a Klein geometry and a Riemannian geometry and so it is the basis of two different generalizations shown as (1) and (2) in figure 2. The notion of a *connection* is an important unifying notion for modern differential geometry. A connection is a device to measure constancy and change and allows us to take derivatives of vector fields and more

[1] In fact, the author did not enjoy high school geometry at all.
[2] Lie group representations are certainly important for geometry.

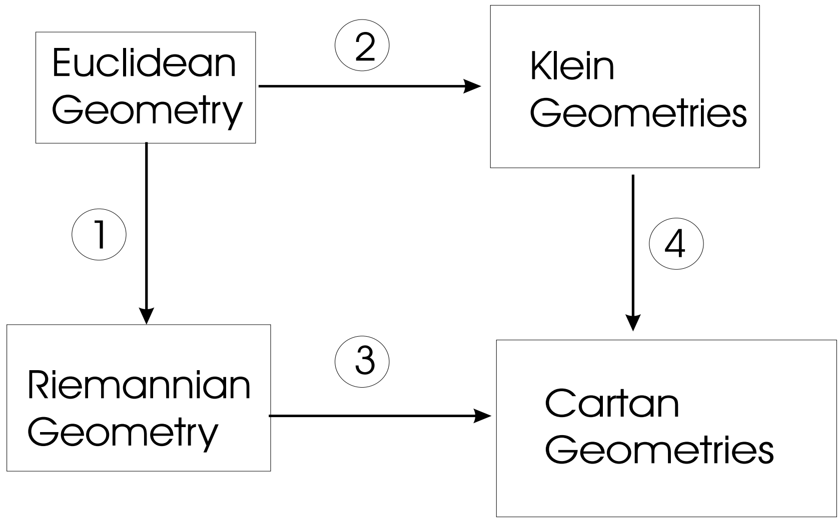
general fields. In Riemannian geometry, the central bit of extra structure is that of a metric tensor which allows us to measure lengths, areas, volumes and so on. In particular, every Riemannian manifold has a distance function and so it a metric space ("metric" in the sense of general topology). In Riemannian geometry the connection comes from the metric tensor and is called the Levi-Civita connection. Thus distance and length are at the root of how change is reckoned in Riemannian geometry. Locally, a connection on an $n-$dimensional Riemannian manifold is described by an $\mathfrak{so}(n)-$valued differential $1-$form (differential forms are studied in chapter 8).

We use the term Klein geometry instead of homogeneous space geometry when we take on the specifically geometric attitude characterized in part by emphasis on the so called Cartan connection which comes from the Maurer-Cartan form on the group itself. It is the generalization of this Lie algebra valued 1-form which gives rise to Cartan Geometry (this is the generalization (4) in the figure). Now Cartan geometry can also be seen as a generalization inspired directly from Riemannian geometry. Riemannian geometry is an example of a Cartan geometry for sure but the way it becomes such is only fully understood from the point of view of the generalization 4 in the figure. From this standpoint the relevant connection is the Cartan version of the Levi-Civita connection which takes values in the Lie algebra of $\mathrm{Euc}(n)$ rather than the Lie algebra of $SO(n)$. This approach is still unfamiliar to many professional differential geometers but as well shall see it is superior in many ways. For a deep understanding of the Cartan viewpoint on differential geometry the reader should look at R.W. Sharp's excellent text (addithere). The presentation of Cartan geometries found in this book it heavily indebted to Sharp's book.

Klein geometry will be pursued at various point throughout the book but especially in chapters 16, 22, and 25. Our goal in the present section will be to introduce a few of the most important groups and related homogeneous spaces. The most exemplary examples are exactly the Euclidean spaces (together with the group of Euclidean motions).

Most of the basic examples involve groups acting on spaces which topologically equivalent to finite dimensional vector spaces. In the calculus of several variables we think of $\mathbb{R}^3$ as being a model for 3-dimensional physical space. Of course, everyone realizes that $\mathbb{R}^3$ is not quite a faithful model of space for several reasons not the least of which is that, unlike physical space, $\mathbb{R}^3$ is a vector space and so, for instance, has a unique special element (point) called the zero element. Physical space doesn't seem to have such a special point (an origin). A formal mathematical notion more appropriate to the situation which removes this idea of an origin is that of an **affine space** defined below 2.1. As actors in physical space, we implicitly and sometimes explicitly impose coordinate systems onto physical space. Rectangular coordinates are the most familiar type of coordinates that physicists and engineers use. In the physical sciences, coordinates are really implicit in our measurement methods and in the instruments used.

We also *impose* coordinates onto certain objects or onto their surfaces. Particularly simple are flat surfaces like tables and so forth which intuitively are



2-dimensional analogues of the space in which we live. Experience with such objects and with 3-dimensional space itself leads to the idealizations known as the Euclidean plane, 3-dimensional Euclidean space. Our intuition about the Euclidean plane and Euclidean 3−space is extended to higher dimensions by analogy. Like 3-dimensional space, the Euclidean plane also has no preferred coordinates or implied vector space structure. Euclid's approach to geometry used neither (at least no explicit use). On the other hand, there does seem to be a special family of coordinates on the plane that makes the equations of geometry take on a simple form. These are the rectangular (orthonormal) coordinates mentioned above. In rectangular coordinates the distance between two points is given by the usual Pythagorean prescription involving the sum of squares. What set theoretic model should a mathematician exhibit as the best mathematical model of these Euclidean spaces of intuition? Well, Euclidean space may not be a vector space as such but since we have the notion of translation in space we do have a vector space "acting by translation". Paying attention to certain features of Euclidean space leads to the following notion:

Definition 2.1 *Let* $\mathbf{A}$ *be a set and* V *be a vector space over a field* $\mathbb{F}$. *We say that* $\mathbf{A}$ *is an* ***affine space*** *with* ***difference space*** V *if there is a map* $+ : V \times \mathbf{A} \to \mathbf{A}$ *written* $(v, p) \mapsto v + p$ *such that*
i) $(v + w) + p = v + (w + p)$ *for all* $v, w \in V$ *and* $p \in \mathbf{A}$
ii) $0 + p = p$ *for all* $p \in \mathbf{A}$.
iii) for each fixed $p \in \mathbf{A}$ *the map* $v \mapsto v + p$ *is a bijection.*

If we have an affine space with difference space V then there is a unique map called the **difference map** $- : \mathbf{A} \times \mathbf{A} \to V$ which is defined by

$$q - p = \text{the unique } v \in V \text{ such that } v + p = q.$$

Thus we can form difference of two points in an affine space but we cannot add two points in an affine space or at least if we can then that is an accidental feature and is not part of what it means to be an affine space. We need to know that such things really exist in some sense. What should we point at as an example of an affine space? In a ironic twist the set which is most ready-to-hand for this purpose is $\mathbb{R}^n$ itself. We just allow $\mathbb{R}^n$ to play the role of both the affine space and the difference space. We take advantage of the existence of a predefine vector addition $+$ to provide the translation map. Now $\mathbb{R}^n$ has many features that are not an essential part of its affine space character. When we let $\mathbb{R}^n$ play this role as an affine space, we must be able to turn a sort of blind eye to features which are not essential to the affine space structure such as the underlying vector spaces structure. Actually, it would be a bit more honest to admit that we will in fact use features of $\mathbb{R}^n$ which are accidental to the affine space structure. After all, the first thing we did was to use the vector space addition from $\mathbb{R}^n$ to provide the translation map. This is fine- we only need to keep track of what aspects of our constructions and which of our calculations retain significant for the properly affine aspects of $\mathbb{R}^n$. The introduction of group theory into the picture helps a great deal in sorting things out. More generally any vector space can be turned into an affine space simply by letting the translation operator $+$ required by the definition of an affine space to be the addition that is part of the vector space structure of V. This is just the same as what we did with $\mathbb{R}^n$ and again V itself is its own difference space and the set V is playing two roles. Similar statements apply it the field is $\mathbb{C}$. In fact, algebraic geometers refer to $\mathbb{C}^n$ as complex affine space.

Let $\mathbb{F}$ be one of the fields $\mathbb{R}$ or $\mathbb{C}$. Most of the time $\mathbb{R}$ is the field intended when no specification is made. It is a fairly common practice to introduce multiple notations for the same set. For example, when we think of the vector space $\mathbb{F}^n$ as an affine space we sometimes denote it by $\mathbf{A}^n(\mathbb{F})$ or just $\mathbf{A}^n$ if the underlying field is understood.

Definition 2.2 *If $\mathbf{A}_i$ is an affine space with difference space V_i for $i = 1, 2$ then we say that a map $F : \mathbf{A}_1 \to \mathbf{A}_2$ is an affine transformation if it has the form $x \mapsto F(x_0) + L(x - x_0)$ for some linear transformation $L : V_1 \to V_2$.*

It is easy to see that for a fixed point p in an affine space $\mathbf{A}$ we immediately obtain a bijection $V \to \mathbf{A}$ which is given by $v \mapsto p + v$. The inverse of this bijection is a map $c_p : \mathbf{A} \to V$ which we put to use presently. An affine space always has a globally defined coordinate system. To see this pick a basis for V. This gives a bijection $V \to \mathbb{R}^n$. Now choose a point $p \in \mathbf{A}$ and compose with the canonical bijection $c_p : \mathbf{A} \to V$ just mentioned. This a coordinates system. Any two such coordinates systems are related by an bijective affine transformation (an affine automorphism). These special coordinate systems can be seen as an alternate way of characterizing the affine structure on the set $\mathbf{A}$. In fact, singling out a family of specially related coordinate systems is an often used method saying what we mean when we say that a set has a structure of a given type and this idea will be encounter repeatedly.

If V is a finite dimensional vector space then it has a distinguished topology[3] which is transferred to the affine space using the canonical bijections. Thus we may consider open sets in A and also continuous functions on open sets. Now it is usually convenient to consider the "coordinate representative" of a function rather than the function itself. By this we mean that if $x : \mathbf{A} \to \mathbb{R}^n$ is a affine coordinate map as above, then we may replace a function $f : U \subset \mathbf{A} \to \mathbb{R}$ (or $\mathbb{C}$) by the composition $f \circ x$. The latter is often denoted by $f(x^1, ..., x^n)$. Now this coordinate representative may, or may not, turn out to be a differentiable function but if it is then all other coordinate representatives of f obtained by using other affine coordinate systems will also be differentiable. For this reason, we may say that f itself is (or is not) differentiable. We say that the family of affine coordinate systems provide the affine space with a differentiable structure and this means that every (finite dimensional) affine space is also an example of differentiable manifold (defined later). In fact, one may take the stance that an affine space is the local model for all differentiable manifolds.

Another structure that will seem a bit trivial in the case of an affine space but that generalizes in a nontrivial way is the idea of a tangent space. We know what it means for a vector to be tangent to a surface at a point on the surface but the abstract idea of a tangent space is also exemplified in the setting of affine spaces. If we have a curve $\gamma : (a, b) \subset \mathbb{R} \to \mathbf{A}$ then the affine space structure allows us to make sense of the difference quotient

$$\dot{\gamma}(t_0) := \lim_{h \to 0} \frac{\gamma(t_0 + h) - \gamma(t_0)}{h} \ , \ t_0 \in (a, b)$$

which defines an element of the difference space V (assuming the limit exists). Intuitively we think of this as the velocity of the curve *based at* $\gamma(t_0)$. We may want to explicitly indicate the base point. If $p \in \mathbf{A}$, then the **tangent space** at p is $\{p\} \times V$ and is often denoted by $T_p\mathbf{A}$. The set of all tangent spaces for points in an open set $U \subset \mathbf{A}$ is called the tangent bundle over U and is another concept that will generalize in a very interesting way. Thus we have not yet arrived at the usual (metric) Euclidean space or Euclidean geometry where distances and angle are among the prominent notions. On the other hand, an affine space supports a geometry called *affine geometry.* The reader who has encountered axiomatic affine geometry will remember that one is mainly concerned about lines and the their mutual points of incidence. In the current analytic approach a line is the set of image points of an affine map $a : \mathbb{R} \to \mathbf{A}$. We will not take up affine geometry proper. Rather we shall be mainly concerned with geometries that result from imposing more structure on the affine space. For example, an affine space is promoted to a metric Euclidean space once we introduce length and angle. Once we introduce length and angle into $\mathbb{R}^n$ it becomes the standard model of n-dimensional Euclidean space and might choose to employ the notation $\mathbf{E}^n$ instead of $\mathbb{R}^n$. The way in which length and angle is introduced into $\mathbb{R}^n$ (or an abstract finite dimensional vector space V) is via an inner product and this is most likely familiar ground for the reader.

[3] If V is infinite dimensional we shall usually assume that V is a topological vector space.

Nevertheless, we will take closer look at what is involved so that we might place Euclidean geometry into a more general context. In order to elucidate both affine geometry and Euclidean geometry and to explain the sense in which the word geometry is being used, it is necessary to introduce the notion of "group action".

2.1 Group Actions, Symmetry and Invariance

One way to bring out the difference between $\mathbb{R}^n$, $\mathbf{A}^n$ and $\mathbf{E}^n$, which are all the same considered as sets, is by noticing that each has its own natural set of coordinates and coordinate transformations. The natural family of coordinate systems for $\mathbf{A}^n$ are the affine coordinates and are related amongst themselves by affine transformations under which lines go to lines (planes to planes etc.). Declaring that the standard coordinates obtained by recalling that $\mathbf{A}^n = \mathbb{R}^n$ determines all other affine coordinate systems. For $\mathbf{E}^n$ the natural coordinates are the orthonormal rectangular coordinates which are related to each other by affine transformations whose linear parts are orthogonal transformations of the difference space. These are exactly the length preserving transformations or *isometries*. We will make this more explicit below but first let us make some comments about the role of symmetry in geometry.

We assume that the reader is familiar with the notion of a group. Most of the groups that play a role in differential geometry are matrix groups and are the prime examples of so called 'Lie groups' which we study in detail later in the book. For now we shall just introduce the notion of a topological group. All Lie groups are topological groups. Unless otherwise indicated finite groups G will be given the discreet topology where every singleton $\{g\} \subset G$ is both open and closed.

Definition 2.3 *Let G be a group. We say that G is a topological group if G is also a topological space and if the maps $\mu : G \times G \to G$ and $\mathrm{inv} : G \to G$, given by $(g_1, g_2) \to g_1 g_2$ and $g_1 \mapsto g^{-1}$ respectively, are continuous maps.*

If G is a countable or finite set we usually endow G with the discrete topology so that in particular, every point would be an open set. In this case we call G a discrete group.

Even more important for our present discussion is the notion of a group action. Recall that if M is a topological space then so is $G \times M$ with the product topology.

Definition 2.4 *Let G and M be as above. A left (resp. right)* ***group action*** *is a map $\alpha : G \times M \to M$ (resp. $\alpha : M \times G \to M$) such that for every $g \in G$ the* ***partial map*** *$\alpha_g(.) := \alpha(g, .)$ (resp. $\alpha_{,g}(.) := \alpha(., g)$) is continuous and such that the following hold:*

1) $\alpha(g_2, \alpha(g_1, x)) = \alpha(g_2 g_1, x)$ (resp. $\alpha(\alpha(x, g_1), g_2) = \alpha(x, g_1 g_2)$) for all $g_1, g_2 \in G$ and all $x \in M$.

2) $\alpha(e, x) = x$ (resp. $\alpha(x, e) = x$) for all $x \in M$.

It is traditional to write $g \cdot x$ or just gx in place of the more pedantic notation $\alpha(g, x)$. Using this notation we have $g_2 \cdot (g_1 \cdot x) = (g_2 g_1) \cdot x$ and $e \cdot x = x$.

We shall restrict our exposition for left actions only since the corresponding notions for a right action are easy to deduce. Furthermore, if we have a right action $x \to \alpha^R(x, g)$ then we can construct an essentially equivalent left action by $\alpha^L : x \to \alpha^L(g, x) := \alpha^R(x, g^{-1})$.

If we have an action $\alpha : G \times M \to M$ then for a fixed x, the set $G \cdot x := \{g \cdot x : g \in G\}$ is called the **orbit** of x. The set of orbits is denoted by M/G or sometimes $G|M$ if we wish to emphasize that the action is a left action. It is easy to see that two orbits $G \cdot x$ and $G \cdot y$ are either disjoint or identical and so define an equivalence relation. The natural projection onto set of orbits $p : M \to M/G$ is given by

$$x \mapsto G \cdot x.$$

If we give M/G the quotient topology then of course p is continuous.

Definition 2.5 *If G acts on M the we call M a G-**space**.*

Exercise 2.1 *Convince yourself that an affine space A with difference space V is a $V-$space where we consider V as an abelian group under addition.*

Definition 2.6 *Let G act on M_1 and M_2. A map $f : M_1 \to M_2$ is called a $G-$map if*

$$f(g \cdot x) = g \cdot f(x)$$

*for all $x \in M$ and all $g \in G$. If f is a bijection then we say that f is a G-automorphism and that M_1 and M_2 are isomorphic as G-spaces. More generally, if G_1 acts on M_1 and G_2 acts on M_2 then a **weak group action morphism** from M_1 to M_2 is a pair of maps (f, ϕ) such that*

(i) $f : M_1 \to M_2$,
(ii) $\phi : G_1 \to G_2$ is a group homomorphism and
*(iii) $f(g \cdot x) = \phi(g) \cdot f(x)$ for all $x \in M$ and all $g \in G$. If f is a bijection and ϕ is an isomorphism then the group actions on M_1 and M_2 are **equivalent**.*

We have refrained from referring to the equivalence in (iii) as a "weak" equivalence since if M_1 and M_2 are G_1 and G_2 spaces respectively which are equivalent in the sense of (iii) then $G_1 \cong G_2$ by definition. Thus if we then identify G_1 and G_2 and call the resulting abstract group G then we recover a $G-$space equivalence between M_1 and M_2.

One main class of examples are those where M is a vector space (say V) and each α_g is a linear isomorphism. In this case, the notion of a left group action is identical to what is called a **group representation** and for each

$g \in G$, the map $g \mapsto \alpha_g$ is a group homomorphism[4] with image in the space $GL(V)$ of linear isomorphism of G to itself. The most common groups that occur in geometry are matrix groups, or rather, groups of automorphisms of a fixed vector space. Thus we are first to consider subgroups of $GL(V)$. A typical way to single out subgroups of $GL(V)$ is provided by the introduction of extra structure onto V such as an orientation and/or a special bilinear form such as an real inner product, Hermitian inner product, or a symplectic from to name just a few. We shall introduce the needed structures and the corresponding groups as needed. As a start we ask the reader to recall that an automorphism of a finite dimensional vector space V is *orientation preserving* if its matrix representation with respect to some basis has positive determinant. (A more geometrically pleasing definition of orientation and of determinant etc. will be introduced later). The set of all orientation preserving automorphisms of V is denoted a subgroup of $GL(V)$ and is denoted $Gl^+(V)$ and referred to as the **proper general linear group**. We are also interested in the group of automorphisms that have determinant equal to 1 which gives us a "*special* linear group". (These will be seen to be volume preserving maps). Of course, once we pick a basis we get an identification of any such linear automorphism group with a group of matrices. For example, we identify $GL(\mathbb{R}^n)$ with the group $GL(n, \mathbb{R})$ of nonsingular $n \times n$ matrices. Also, $Gl^+(\mathbb{R}^n) \cong Gl^+(n, \mathbb{R}) = \{A \in GL(n, \mathbb{R}) : \det A = 1\}$. A fairly clear pattern of notation is emerging and we shall not be so explicit about the meaning of the notation in the sequel.

(oriented) Vector Space	General linear	Proper general linear	Special linear
General V	$GL(V)$	$Gl^+(V)$	$SL(V)$
$\mathbb{F}^n$	$GL(\mathbb{F}^n)$	$Gl^+(\mathbb{F}^n)$	$SL(\mathbb{F}^n)$
Matrix group	$GL(n, \mathbb{F})$	$Gl^+(n, \mathbb{F})$	$SL(n, \mathbb{F})$
Banach space E	$GL(\mathsf{E})$	?	?

If one has a group action then one has a some sort of geometry. From this vantage point, the geometry is whatever is 'preserved' by the group action.

Definition 2.7 *Let $\mathcal{F}(M)$ be the vector space of all complex valued functions on M. An action of G on M produces an action of G on $\mathcal{F}(M)$ as follows*

$$(g \cdot f)(x) := f(g^{-1}x)$$

*This action is called the **induced action**.*

This induced action preserves the vector space structure on $\mathcal{F}(M)$ and so we have produced an example of a group representation. If there is a function $f \in \mathcal{F}(M)$ such that $f(x) = f(g^{-1}x)$ for all $g \in G$ then f is called an **invariant**. In other words, f is an invariant if it remains fixed under the induced action on $\mathcal{F}(M)$. The existence of an invariant often signals the presence of an underlying geometric notion.

Example 2.1 *Let $G = O(n, \mathbb{R})$ be the **orthogonal group** (all invertible matrices Q such that $Q^{-1} = Q^t$), let $M = \mathbb{R}^n$ and let the action be given by matrix*

[4] A right group action does not give rise to a homomorphism but an "anti-homomorphism".

multiplication on column vectors in $\mathbb{R}^n$*;*

$$(Q, v) \mapsto Qv$$

Since the length of a vector v *as defined by* $\sqrt{v \cdot v}$ *is preserved by the action of* $\mathrm{O}(\mathbb{R}^n)$ *we pick out this notion of length as a geometric notion. The* ***special orthogonal group*** $S\mathrm{O}(n, \mathbb{R})$ *(or just* $S\mathrm{O}(n)$*) is the subgroup of* $\mathrm{O}(n, \mathbb{R})$ *consisting of elements with determinant* 1*. This is also called the rotation group (especially when* $n = 3$*).*

More abstractly, we have the following

Definition 2.8 *If* V *be a vector space (over a field* $\mathbb{F}$*) which is endowed with a distinguished nondegenerate symmetric bilinear form* b *(a real inner product if* $\mathbb{F} = \mathbb{R}$*), then we say that* V *has an orthogonal structure. The set of linear transformations* $L : V \to V$ *such that* $b(Lv, Lw) = b(v, w)$ *for all* $v, w \in V$ *is a group denoted* $\mathrm{O}(V, \langle, \rangle)$ *or simply* $\mathrm{O}(V)$*. The elements of* $\mathrm{O}(V)$ *are called* ***orthogonal transformations.***

A map A between complex vector spaces which is linear over $\mathbb{R}$ and satisfies $A(\bar{v}) = \overline{A(v)}$ is conjugate linear (or antilinear). Recall that for a complex vector space V, a map $h : V \times V \to \mathbb{C}$ that is linear in one variable and conjugate linear in the other is called a **sesquilinear form**. If, further, h is nondegenerate then it is called a **Hermitian form** (or simply a complex inner product).

Definition 2.9 *Let* V *be a complex vector space which is endowed distinguished nondegenerate Hermitian* h*, then we say that* V *has a* ***unitary structure****. The set of linear transformations* $L : V \to V$ *such that* $h(Lv, Lw) = h(v, w)$ *for all* $v, w \in V$ *is a group denoted* $U(V)$*. The elements of* $\mathrm{O}(V)$ *are called* ***unitary transformations.*** *The standard Hermitian form on* $\mathbb{C}^n$ *is* $(x, y) \mapsto \sum \bar{x}^i y^i$ *(or depending on taste* $\sum x^i \bar{y}^i$*). We have the obvious identification* $U(\mathbb{C}^n) = U(n, \mathbb{C})$*.*

It is common to assume that $\mathrm{O}(n)$ refers to $\mathrm{O}(n, \mathbb{R})$ while $U(n)$ refers to $U(n, \mathbb{C})$. It is important to notice that with the above definitions $U(\mathbb{C}^n)$ is not the same as $\mathrm{O}(\mathbb{C}^n, \sum x^i y^i)$ since $b(x, y) = \sum x^i y^i$is bilinear rather than sesquilinear.

Example 2.2 *Let* G *be the* ***special linear group*** $\mathrm{SL}(n, \mathbb{R}) = \{A \in \mathrm{GL}(n, \mathbb{R}) : \det A = 1\}$*. The length of vectors are not preserved but something else is preserved. If* $v^1, ..., v^n \in \mathbb{R}^n$ *then the determinant function* $\det(v^1, ..., v^n)$ *is preserved:*

$$\det(v^1, ..., v^n) = \det(Qv^1, ..., Qv^n)$$

Since $\det$ *is not really a function on* $\mathbb{R}^n$ *this invariant doesn't quite fit the definition of an invariant above. On the other hand,* $\det$ *is a function of several variables, i.e. a function on* $\mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$ *and so there is the obvious induced action on this space of function. Clearly we shall have to be flexible if* *we are to capture all the interesting invariance phenomenon. Notice that the* $S\mathrm{O}(n,\mathbb{R}) \subset \mathrm{SL}(n,\mathbb{R})$ *and so the action of* $S\mathrm{O}(n,\mathbb{R})$ *on* $\mathbb{R}^n$ *is also orientation preserving.*

Orthogonal Structure	Full orthogonal	Special Orthogonal
V + nondegen. sym. form b	$\mathrm{O}(V,b)$	$S\mathrm{O}(V,b)$
$\mathbb{F}^n, \sum x^i y^i$	$\mathrm{O}(\mathbb{F}^n)$	$S\mathrm{O}(\mathbb{F}^n)$
Matrix group	$\mathrm{O}(n,\mathbb{F})$	$S\mathrm{O}(n,\mathbb{F})$
Hilbert space E	$\mathrm{O}(\mathsf{E})$	?

2.2 Some Klein Geometries

Definition 2.10 *A group action is said to be* ***transitive*** *if there is only one orbit. A group action is said to be* ***effective*** *if* $g \cdot x = x$ *for all* $x \in M$ *implies that* $g = e$ *(the identity element). A group action is said to be* ***free*** *if* $g \cdot x = x$ *for some* x *implies that* $g = e$*.*

Klein's view is that a geometry is just a transitive G-space. Whatever properties of figures (or more general objects) that remain unchanged under the action of G are deemed to be geometric properties by definition.

Example 2.3 *We give an example of an* $\mathrm{SL}(n,\mathbb{R})$*-space as follows: Let* M *be the upper half complex plane* $\{z \in \mathbb{C} : \operatorname{Im} z > 0\}$*. Then the action of* $\mathrm{SL}(n,\mathbb{R})$ *on* M *is given by*

$$(A, z) \mapsto \frac{az+b}{cz+d}$$

where $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

Exercise 2.2 *Let* $A \in \mathrm{SL}(n,\mathbb{R})$ *as above and* $w = A \cdot z = \frac{az+b}{cz+d}$*. Show that if* $\operatorname{Im} z > 0$ *then* $\operatorname{Im} w > 0$ *so that the action of the last example is well defined.*

2.2.1 Affine Space

Now let us consider the notion of an affine space again. Careful attention to the definitions reveals that an affine space is a set on which a vector space acts (a vector space is also an abelian group under the vector addition $+$). We can capture more of what an affine space $\mathbf{A}^n$ is geometrically by enlarging this action to include the full group of nonsingular affine maps of $\mathbf{A}^n$ onto itself. This group is denoted $\mathrm{Aff}(\mathbf{A}^n)$ and is the set of transformations of $\mathbf{A}^n$ of the form $A : x \mapsto Lx_0 + L(x - x_0)$ for some $L \in \mathrm{GL}(n,\mathbb{R})$ and some $x_0 \in \mathbf{A}^n$. More abstractly, if $\mathbf{A}$ is an affine space with difference space V then we have the group of all transformations of the form $A : x \mapsto Lx_0 + L(x - x_0)$ for some $x_0 \in A$ and $L \in \mathrm{GL}(V)$. This is the **affine group** associated with $\mathbf{A}$ and is denoted by $\mathrm{Aff}(\mathbf{A})$. Using our previous terminology, A is an $\mathrm{Aff}(\mathbf{A})-$space. $\mathbf{A}^n$ $(= \mathbb{R}^n)$ is our standard model of an n-dimensional affine space. It is an $\mathrm{Aff}(\mathbf{A}^n)$-space.

Exercise 2.3 *Show that if* $\mathbf{A}$ *an is a* n*-dimensional affine space with difference space* V *then* Aff*(*$\mathbf{A}$*)*$\cong$ Aff$(\mathbf{A}^n)$ *and the* Aff*(*$\mathbf{A}$*)-space* $\mathbf{A}$ *is equivalent to the* Aff$(\mathbf{A}^n)-$*space* $\mathbf{A}^n$*.*

Because of the result of this last exercise it is sufficient mathematically to restrict the study of affine space to the concrete case of $\mathbf{A}^n$ the reason that we refer to $\mathbf{A}^n$ as the standard affine space. We can make things even nicer by introducing a little trick which allows us to present $Aff(\mathbf{A}^n)$ as a matrix group. Let $\mathbf{A}^n$ be identified with the set of column vectors in $\mathbb{R}^{n+1}$ of the form

$$\begin{bmatrix} 1 \\ x \end{bmatrix} \text{ where } x \in \mathbb{R}^n$$

The set of all matrices of the form

$$\begin{bmatrix} 1 & 0 \\ x_0 & L \end{bmatrix} \text{ where } Q \in \mathrm{GL}(n, \mathbb{R}) \text{ and } x_0 \in \mathbb{R}^n$$

is a group. Now

$$\begin{bmatrix} 1 & 0 \\ x_0 & L \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix} = \begin{bmatrix} 1 \\ Lx + x_0 \end{bmatrix}$$

Remember that affine transformations of $\mathbf{A}^n$ ($= \mathbb{R}^n$ as a set) are of the form $x \mapsto Lx + x_0$. In summary, when we identify $\mathbf{A}^n$ with the vectors of the form $\begin{bmatrix} 1 \\ x \end{bmatrix}$ then the group $\mathrm{Aff}(\mathbf{A}^n)$ is the set of $(n+1) \times (n+1)$ matrices of the form indicated and the action of $\mathrm{Aff}(\mathbf{A}^n)$ is given by matrix multiplication.

It is often more appropriate to consider the **proper affine group** $Aff^+(\mathbf{A}, V)$ obtained by adding in the condition that its elements have the form $A : x \mapsto Lx_0 + L(x - x_0)$ with $\det L > 0$.

We have now arrive at another meaning of affine space. Namely, the system $(\mathbf{A}, V, Aff^+(\mathbf{A}), \cdot, +)$. What is new is that we are taking the action "$\cdot$" of $Aff^+(\mathbf{A})$ on as part of the structure so that now A is an $Aff^+(\mathbf{A})-$space.

Exercise 2.4 *Show that the action of* V *on* $\mathbf{A}$ *is transitive and free. Show that the action of* $Aff^+(\mathbf{A})$ *on* $\mathbf{A}$ *is transitive and effective but not free.*

Affine geometry is the study of properties attached to figures in an affine space that remain, in some appropriate sense, unchanged under the action of the affine group (or the proper affine group Aff^+). For example, coincidence properties of lines and points are preserved by Aff.

2.2.2 Special Affine Geometry

We can introduce a bit more rigidity into affine space by changing the linear part of the group to $\mathrm{SL}(n)$.

Definition 2.11 *The group* $SAff(\mathbf{A})$ *of affine transformations of an affine space* $\mathbf{A}$ *that are of the form*

$$A : x \mapsto Lx_0 + L(x - x_0)$$

with $\det L = 1$ *will be called the special affine group.*

We will restrict ourselves to the standard model of $\mathbf{A}^n$. With this new group the homogeneous space structure is now different and the geometry has changed. For one thing volume now makes sense. The are many similarities between special affine geometry and Euclidean geometry. As we shall see in chapter 16, in dimension 3, there exist special affine versions of arc length and surface area and mean curvature.

The volume of a parallelepiped is preserved under this action.

2.2.3 Euclidean space

Suppose now that we have an affine space $(A, V, +)$ together with an inner product on the vector space V. Consider the group of affine transformations the form $A : x \mapsto Lx_0 + L(x - x_0)$ where now $L \in \mathrm{O}(V)$. This is called the Euclidean motion group of A. Then $\mathbf{A}$ becomes, in this context, a Euclidean space and we might choose a different symbol, say $\mathbf{E}$ for the space to encode this fact. For example, when $\mathbb{R}^n$ is replaced by $\mathbf{E}^n$ and the Euclidean motion group is denoted $\mathrm{Euc}(\mathbf{E}^n)$.

Remark: Now every tangent space $T_pA = \{p\} \times V$ inherits the inner product from V in the obvious and trivial way. This is our first example of a metric tensor. Having a metric tensor on the more general spaces that we study later on will all us to make sense of lengths of curves, angles between velocity curves, volumes of subsets and much more.

In the concrete case of $\mathbf{E}^n$ (secretly $\mathbb{R}^n$again) we may represent the Euclidean motion group $\mathrm{Euc}(\mathbf{E}^n)$ as a matrix group, denoted $\mathrm{Euc}(n)$, by using the trick we used before. If we want to represent the transformation $x \to Qx + b$ where $x, b \in \mathbb{R}^n$ and $Q \in \mathrm{O}(n, \mathbb{R})$, we can achieve this by letting column vectors

$$\begin{bmatrix} 1 \\ x \end{bmatrix} \in \mathbb{R}^{n+1}$$

represent elements x. Then we take the group of matrices of the form

$$\begin{bmatrix} 1 & 0 \\ b & Q \end{bmatrix}$$

to play the role of $\mathrm{Euc}(n)$. This works because

$$\begin{bmatrix} 1 & 0 \\ b & Q \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix} = \begin{bmatrix} 1 \\ b + Qx \end{bmatrix}.$$

The set of all coordinate systems related to the standard coordinates by an element of $\mathrm{Euc}(n)$ will be called **Euclidean coordinates** (or sometimes orthonormal coordinates). It is these coordinate which we should use to calculate distance.

Basic fact 1: The (square of the) distance between two points in $\mathbf{E}^n$ can be calculated in any of the equivalent coordinate systems by the usual formula. This means that if $P, Q \in \mathbf{E}^n$ and we have some Euclidean coordinates for these two points: $x(P) = (x^1(P), ..., x^n(P))$, $x(Q) = (x^1(Q), ..., x^n(Q))$ then vector difference in these coordinates is $\Delta x = x(P) - x(Q) = (\Delta x^1, ..., \Delta x^n)$. If $y = (y^1, ..., y^n)$ are any other coordinates related to x by $y = Tx$ for some $T \in \mathrm{Euc}(n)$ then we have $\sum(\Delta y^i)^2 = \sum(\Delta x^i)^2$ where $\Delta y = y(P) - y(Q)$ etc. The distance between two points in a Euclidean space is a simple example of a geometric invariant.

Let us now pause to consider again the abstract setting of $G-$spaces. Is there geometry here? We will not attempt to give an ultimate definition of geometry associated to a $G-$space since we don't want to restrict the direction of future generalizations. Instead we offer the following two (somewhat tentative) definitions:

Definition 2.12 *Let M be a $G-$space. A figure in M is a subset of M. Two figures S_1, S_2 in M are said to be **geometrically equivalent** or **congruent** if there exists an element $g \in G$ such that $g \cdot S_1 = S_2$.*

Definition 2.13 *Let M be a $G-$space and let $\mathcal{S}$ be some family of figures in M such that $g \cdot S \in \mathcal{S}$ whenever $.S \in \mathcal{S}$. A function $I : \mathcal{S} \to \mathbb{C}$ is called a (complex valued) geometric invariant if $I(g \cdot S) = I(S)$ for all $S \in \mathcal{S}$.*

Example 2.4 *Consider the action of $\mathrm{O}(n)$ on $\mathbf{E}^n$. Let $\mathcal{S}$ be the family of all subsets of $\mathbf{E}^n$ which are the images of C^1 of the form $c : [a, b] \to \mathbf{E}^n$ such that c' is never zero (regular curves). Let $S \in \mathcal{S}$ be such a curve. Taking some regular C^1map c such that S is the image of c we define the length*

$$L(S) = \int_a^b \|c'(t)\| \, dt$$

It is common knowledge that $L(S)$ is independent of the parameterizing map c and so L is a geometric invariant (this is a prototypical example)

Notice that two curves may have the same length without being congruent. The invariant of length by itself does not form a complete set of invariants for the family of regular curves $\mathcal{S}$. The definition we have in mind here is the following:

Definition 2.14 *Let M be a $G-$space and let $\mathcal{S}$ be some family of figures in M. Let $\mathcal{I} = \{I_\alpha\}_{\alpha \in A}$ be a set of geometric invariants for $\mathcal{S}$. We say that $\mathcal{I}$ is a complete set of invariants if for every $S_1, S_2 \in \mathcal{S}$ we have that S_1 is congruent to S_2 if and only if $I_\alpha(S_1) = I_\alpha(S_2)$ for all $\alpha \in A$.*

Example 2.5 *Model Euclidean space by $\mathbb{R}^3$ and consider the family $\mathcal{C}$ of all regular curves $c : [0, L] \to \mathbb{R}^3$ such that $\frac{dc}{dt}$ and $\frac{d^2c}{dt^2}$ never vanish. Each such curve has a parameterization by arc length which we may take advantage of for the purpose of defining a complete set of invariant for such curves. Now let $c : [0, L] \to \mathbb{R}^3$ a regular curve parameterized by arc length. Let*

$$\mathbf{T}(s) := \frac{\frac{dc}{dt}(s)}{\left|\frac{dc}{dt}(s)\right|}$$

define the unit tangent vector field along c. The curvature is an invariant defined on c that we may think of as a function of the arc length parameter. It is defined by $\kappa(s) := \left|\frac{dc}{dt}(s)\right|$. We define two more fields along c. First the normal field is defined by $\mathbf{N}(s) = \frac{dc}{dt}(s)$. Next, define the unit binormal vector field $\mathbf{B}$ by requiring that $\mathbf{T}, \mathbf{N}, \mathbf{B}$ is a positively oriented triple of orthonormal unit vectors. By positively oriented we mean that

$$\det[\mathbf{T}, \mathbf{N}, \mathbf{B}] = 1.$$

*We now show that $\frac{d\mathbf{N}}{ds}$ is parallel to $\mathbf{B}$. For this it suffices to show that $\frac{d\mathbf{N}}{ds}$ is normal to both $\mathbf{T}$ and $\mathbf{N}$. First, we have $\mathbf{N}(s) \cdot \mathbf{T}(s) = 0$. If this equation is differentiated we obtain $2\frac{d\mathbf{N}}{ds} \cdot \mathbf{T}(s) = 0$. On the other hand we also have $1 = \mathbf{N}(s) \cdot \mathbf{N}(s)$ which differentiates to give $2\frac{d\mathbf{N}}{ds} \cdot \mathbf{N}(s) = 0$. From this we see that there must be a function $\tau = \tau(s)$ such that $\frac{d\mathbf{N}}{ds} := \tau\mathbf{B}$. This is a function of arc length but should really be thought of as a function on the curve. This invariant is called the **torsion**. Now we have a matrix defined so that*

$$\frac{d}{ds}[\mathbf{T}, \mathbf{N}, \mathbf{B}] = [\mathbf{T}, \mathbf{N}, \mathbf{B}] \begin{bmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & \tau & 0 \end{bmatrix}$$

or in other words

$$\begin{aligned} \frac{d\mathbf{T}}{ds} &= & \kappa\mathbf{N} & \\ \frac{d\mathbf{N}}{ds} &= -\kappa\mathbf{T} & & \tau\mathbf{B} \\ \frac{d\mathbf{B}}{ds} &= & \tau\mathbf{N} & \end{aligned}$$

Since $F = [\mathbf{T}, \mathbf{N}, \mathbf{B}]$ is by definition an orthogonal matrix we have $F(s)F^t(s) = I$. It is also clear that there is some matrix function $A(s)$ such that $F' = F(s)A(s)$. Also, Differentiating we have $\frac{dF}{ds}(s)F^t(s) + F(s)\frac{dF}{ds}^t(s) = 0$ and so

$$\begin{aligned} FAF^t + FA^tF^t &= 0 \\ A + A^t &= 0 \end{aligned}$$

since F is invertible. Thus $A(s)$ is antisymmetric. But we already have established that $\frac{d\mathbf{T}}{ds} = \kappa\mathbf{N}$ and $\frac{d\mathbf{B}}{ds} = \tau\mathbf{N}$ and so the result follows. It can be shown that the functions κ and τ completely determine a sufficiently regular curve up to reparameterization and rigid motions of space. The three vectors form a vector field along the curve c. At each point $p = \mathbf{c}(s)$ along the curve c the provide *and oriented orthonormal basis (or frame) for vectors based at* p*. This basis is called the Frenet frame for the curve. Also,* $\kappa(s)$ *and* $\tau(s)$ *are called the (unsigned) curvature and torsion of the curve at* $\mathbf{c}(s)$*. While,* κ *is never negative by definition we may well have that* $\tau(s)$ *is negative. The curvature is, roughly speaking, the reciprocal of the radius of the circle which is tangent to* $\mathbf{c}$ *at* $\mathbf{c}(s)$ *and best approximates the curve at that point. On the other hand,* τ *measures the twisting of the plane spanned by* $\mathbf{T}$ *and* $\mathbf{N}$ *as we move along the curve. If* $\gamma : I \to \mathbb{R}^3$ *is an arbitrary speed curve then we define* $\kappa_\gamma(t) := \kappa \circ h^{-1}$ *where* $h : I' \to I$ *gives a unit speed reparameterization* $\mathbf{c} = \gamma \circ h : I' \to \mathbb{R}^n$*. Define the torsion function* τ_γ *for* γ *by* $\tau \circ h^{-1}$*. Similarly we have*

$$\mathbf{T}_\gamma(t) := \mathbf{T} \circ h^{-1}(t)$$
$$\mathbf{N}_\gamma(t) := \mathbf{N} \circ h^{-1}(t)$$
$$\mathbf{B}_\gamma(t) := B \circ h^{-1}(t)$$

Exercise 2.5 *If* $\mathbf{c} : I \to \mathbb{R}^3$ *is a unit speed reparameterization of* $\gamma : I \to \mathbb{R}^3$ *according to* $\gamma(t) = \mathbf{c} \circ h$ *then show that*

1. $\mathbf{T}_\gamma(t) = \gamma' / \|\gamma'\|$
2. $\mathbf{N}_\gamma(t) = \mathbf{B}_\gamma(t) \times \mathbf{T}_\gamma(t)$
3. $\mathbf{B}_\gamma(t) = \frac{\gamma' \times \gamma''}{\|\gamma' \times \gamma''\|}$
4. $\kappa_\gamma = \frac{\|\gamma' \times \gamma''\|}{\|\gamma''\|^3}$
5. $\tau_\gamma = \frac{(\gamma' \times \gamma'') \cdot \gamma'''}{\|\gamma' \times \gamma''\|^2}$

Exercise 2.6 *Show that* $\gamma'' = \frac{dv}{dt}\mathbf{T}_\gamma + v^2 \kappa_\gamma \mathbf{N}_\gamma$ *where* $v = \|\gamma'\|$.

Example 2.6 *For a curve confined to a plane we haven't got the opportunity to define* $\mathbf{B}$ *or* τ*. However, we can obtain a more refined notion of curvature. We now consider the special case of curves in* $\mathbb{R}^2$*. Here it is possible to define a signed curvature which will be positive when the curve is turning counterclockwise. Let* $J : \mathbb{R}^2 \to \mathbb{R}^2$ *be given by* $J(a, b) := (-b, a)$*. The signed curvature* $\kappa_\gamma^\pm$ *of* γ *is given by*

$$\kappa_\gamma^\pm(t) := \frac{\gamma''(t) \cdot J\gamma'(t)}{\|\gamma'(t)\|^3}.$$

If γ *is a parameterized curve in* $\mathbb{R}^2$ *then* $\kappa_\gamma \equiv 0$ *then* γ *(parameterizes) a straight line. If* $\kappa_\gamma \equiv k_0 > 0$ *(a constant) then* γ *parameterizes a portion of a circle of radius* $1/k_0$*. The unit tangent is* $\mathbf{T} = \frac{\gamma'}{\|\gamma'\|}$*. We shall redefine the normal* $\mathbf{N}$ *to a curve to be such that* $\mathbf{T}, \mathbf{N}$ *is consistent with the orientation given by the standard basis of* $\mathbb{R}^2$*.*

Exercise 2.7 *If* $\mathbf{c} : I \to \mathbb{R}^2$ *is a unit speed curve then*

1. $\frac{d\mathbf{T}}{ds}(s) = \kappa_{\mathbf{c}}(s)\mathbf{N}(s)$
2. $\mathbf{c}''(s) = \kappa_{\mathbf{c}}(s)\,(J\mathbf{T}(s))$

Example 2.7 *Consider the action of $Aff^+(2)$ on the affine plane $\mathbf{A}^2$. Let $\mathcal{S}$ be the family of all subsets of $\mathbf{A}^2$ which are zero sets of quadratic polynomials of the form*

$$ax^2 + bxy + cy^2 + dx + ey + f.$$

with the nondegeneracy condition $4ac - b^2 \neq 0$. If S is simultaneously the zero set of nondegenerate quadratics $p_1(x, y)$ and $p_2(x, y)$ then $p_1(x, y) = p_2(x, y)$. Furthermore, if $g \cdot S_1 = S_2$ where S_1 is the zero set of p_1 then S_2 is the zero set of the nondegenerate quadratic polynomial $p_2 := p_1 \circ g^{-1}$. A little thought shows that we may as well replace $\mathcal{S}$ by the set of set of nondegenerate quadratic polynomials and consider the induced action on this set: $(g, p) \mapsto p \circ g^{-1}$. We may now obtain an invariant: Let $S = p^{-1}(0)$ and let $I(S) =: I(p) =: sgn(4ac - b^2)$ where $p(x, y) = ax^2 + bxy + cy^2 + dx + ey + f$. In this case, $\mathcal{S}$ is divided into exactly two equivalence classes

Notice that in example 2.4 above the set of figures considered was obtained as the set of images for some convenient family of maps *into* the set M. This raises the possibility of a slight change of viewpoint: maybe we should be studying the maps themselves. Given a G-set M, we could consider a family of maps $\mathcal{C}$ from some fixed topological space T (or family of topological spaces like intervals in R say) into M which is invariant in the sense that if $c \in \mathcal{C}$ then the map $g \cdot c$ defined by $g \cdot c : t \mapsto g \cdot c(t)$ is also in $\mathcal{C}$. Then two elements c_1 and c_2 would be "congruent" if and only if $g \cdot c_1 = c_2$ for some $g \in G$. Now suppose that we find a function $I : \mathcal{C} \to \mathbb{C}$ which is an invariant in the sense that $I(g \cdot c) = I(c)$ for all $c \in \mathcal{C}$ and all $g \in G$. We do not necessarily obtain an invariant for the set of images of maps from $\mathcal{C}$. For example, if we consider the family of regular curves $c : [0, 1] \to \mathbf{E}^2$ and let $G = \mathrm{O}(2)$ with the action introduced earlier, then the *energy functional* defined by

$$E(c) := \int_0^1 \frac{1}{2} \left\| c'(t) \right\|^2 dt$$

is an invariant for the induced action on this set of curves but even if c_1 and c_2 have the same image it does not follow that $E(c_1) = E(c_2)$. Thus E is a geometric invariant of the curve but not of the set which is its image. In the elementary geometric theory of curves in Euclidean spaces one certainly wants to understand curves as sets. One starts out by studying the maps $c : I \to \mathbf{E}^n$ first. In order to get to the geometry of the subsets which are images of regular curves one must consider how quantities defined depend on the choice of parameterization. Alternatively, a standard parameterization (parameterization by arc length) is always possible and this allows one to provide geometric invariants of the image sets (see Appendix J) .

Similarly, example 2.14 above invites us to think about maps *from M* into some topological space T (like $\mathbb{R}$ for example). We should pick a family of maps

$\mathcal{F}$ such that if $f \in \mathcal{F}$ then $g \cdot f$ is also in $\mathcal{F}$ where $g \cdot f : x \mapsto f(g^{-1} \cdot x)$. Thus we end up with G acting on the set $\mathcal{F}$. This is an induced action. We have chosen to use g^{-1} in the definition so that we obtain a left action on $\mathcal{F}$ rather than a right action. In any case, we could then consider f_1 and f_2 to be congruent if $g \cdot f_1 = f_2$ for some $g \in G$.

There is much more to the study of figure in Euclidean space than we have indicated here. We prefer to postpone introduction of these concepts until after we have a good background in manifold theory and then introduce the more general topic of Riemannian geometry. Under graduate level differential geometry courses usually consist mostly of the study of curves and surfaces in $3-$dimensional Euclidean space and the reader who has been exposed to this will already have an idea of what I am talking about. A quick review of curves and surfaces is provided in appendix J. The study of Riemannian manifolds and submanifolds that we take up in chapters 20 and 21.

We shall continue to look at simple homogeneous spaces for inspiration but now that we are adding in the notion of time we might try thinking in a more dynamic way. Also, since the situation has become decidedly more physical it would pay to start considering the possibility that the question of what counts as geometric might be replaced by the question of what counts as physical. We must eventually also ask what other group theoretic principals (if any) are need to understand the idea of invariants of motion such as conservation laws.

2.2.4 Galilean Spacetime

Spacetime is the set of all events in a (for now 4 dimensional) space. At first it might seem that time should be included by simply taking the Cartesian product of space $\mathbf{E}^3$ with a copy of $\mathbb{R}$ that models time: Spacetime=Space$\times\mathbb{R}$. Of course, this leads directly to $\mathbb{R}^4$, or more appropriately to $\mathbf{E}^3 \times \mathbb{R}$. Topologically this is right but the way we have written it implies an inappropriate and unphysical decomposition of time and space. If we only look at (affine) self transformations of $\mathbf{E}^3 \times \mathbb{R}$ that preserve the decomposition then we are looking at what is sometimes called Aristotelean spacetime (inappropriately insulting Aristotel). The problem is that we would not be taking into account the relativity of motion. If two spaceships pass each other moving a constant relative velocity then who is to say who is moving and who is still (or if both are moving). The decomposition $\mathbf{A}^4 = \mathbb{R}^4 = \mathbf{E}^3 \times \mathbb{R}$ suggests that a body is at rest if and only if it has a career (worldline) of the form $p \times \mathbb{R}$; always stuck at p in other words. But relativity of constant motion implies that no such assertion can be truly objective relying as it does on the assumption that one coordinate system is absolutely "at rest". Coordinate transformations between two sets of coordinates on $4-$dimensional spacetime which are moving relative to one another at constant nonzero velocity should mix time and space in some way. There are many ways of doing this and we must make a choice. The first concept of spacetime that we can take seriously is Galilean spacetime. Here we lose absolute motion (thankfully) but retain an absolute notion of time. In Galilean spacetime, it is entirely appropriate to ask whether two events are simultaneous or not. Now the idea that simultaneity is a well define is very intuitive but we shall shortly introduce Minkowski spacetime as more physically realistic and then discover that simultaneity will become a merely relative concept! The appropriate group of coordinate changes for Galilean spacetime is the **Galilean group** and denoted by Gal. This is the group of transformations of $\mathbb{R}^4$ (thought of as an affine space) generated by the follow three types of transformations:

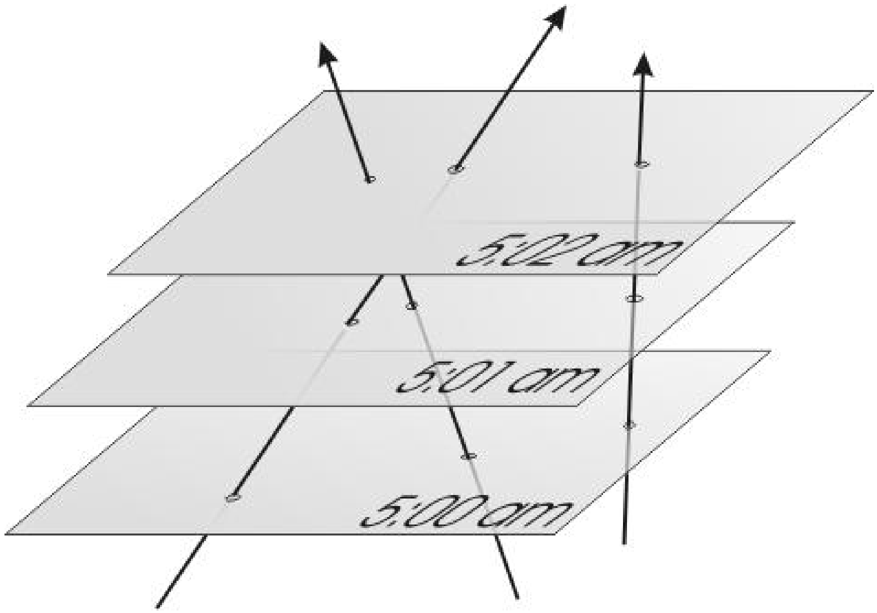
1. **Spatial rotations**: These are of the form
$$\begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & R & \\ 0 & & & \end{bmatrix} \begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix}$$
where $R \in \mathrm{O}(3)$,

2. **Translations of the origin**
$$\begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix} + \begin{bmatrix} t_0 \\ x_0 \\ y_0 \\ z_0 \end{bmatrix}$$
for some $(t_0, x_0, y_0, z_0) \in \mathbb{R}^4$.

3. **Uniform motions**. These are transformations of the form
$$\begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} t \\ x + v_1 t \\ y + v_2 t \\ z + v_3 t \end{bmatrix}$$
for some (velocity) $v = (v_1, v_2, v_3)$. The picture here is that there are two observers each making measurements with respect to their own rectangular spatial coordinate systems which are moving relative to each other with a constant velocity v. Each observer is able to access some universally available clock or synchronized system of clock which will give time measurements that are unambiguous except for choice of the "zero time" (and choice of units which we will assume fixed).

The set of all coordinates related to the standard coordinates of $\mathbb{R}^4$ by an element of Gal will be referred to as Galilean inertial coordinates. When the set $\mathbb{R}^4$ is provided with the action by the Galilean group we refer to it a Galilean spacetime. We will not give a special notation for the space. As a matter of notation it is common to denoted (t, x, y, z) by (x^0, x^1, x^2, x^3). The spatial part of an inertial coordinate system is sometimes denoted by $\mathbf{r} := (x, y, z) = (x^1, x^2, x^3)$.

Basic fact 2: The "spatial separation" between any two events E_1 and E_2 in Galilean spacetime is calculated as follows: Pick some Galilean inertial



coordinates $x = (x^0, x^1, x^2, x^3)$ and let $\Delta\mathbf{r} := \mathbf{r}(E_2) - \mathbf{r}(E_1)$ then the (square of the) spatial separation is calculated as $s = |\Delta\mathbf{r}| = \sum_{i=1}^{3}(x^i(E_2) - x^i(E_1))^2$. The result **definitely does** depend on the choice of Galilean inertial coordinates. Spacial separation is a relative concept in Galilean spacetime. On the other hand, the **temporal separation** $|\Delta t| = |t(E_2) - t(E_1)|$ **does not depend of the choice of coordinates**. Thus, it makes sense in this world to ask whether two events occurred at the same time or not.

2.2.5 Minkowski Spacetime

As we have seen, the vector space $\mathbb{R}^4$ may be provided with a special scalar product given by $\langle x, y\rangle := x^0y^0 - \sum_{i=1}^{3} x^iy^i$ called the Lorentz scalar product (in the setting of Geometry this is usual called a Lorentz metric). If one considers the physics that this space models then we should actual have $\langle x, y\rangle := c^2x^0y^0 - \sum_{i=1}^{3} x^iy^i$ where the constant c is the speed of light in whatever length and time units one is using. On the other hand, we can follow the standard trick of using units of length and time such that in these units the speed of light is equal to 1. This scalar product space is sometimes denoted by $\mathbb{R}^{1,3}$. More abstractly, a Lorentz vector space $V^{1,3}$ is a $4-$dimensional vector space with scalar product $\langle .,.\rangle$ which is isometric to $\mathbb{R}^{1,3}$. An orthonormal basis for a Lorentz space is by definition a basis (e_0, e_1, e_2, e_3) such that the matrix which represents the scalar product with respect to basis is

$$\eta = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Once we pick such a basis for a Lorentz space we get an isomorphism with $\mathbb{R}^{1,3}$. Physically and geometrically, the standard basis of $\mathbb{R}^{1,3}$ is just one among many orthonormal bases so if one is being pedantic, the abstract space $V^{1,3}$ would be more appropriate. The group associated with a Lorentz scalar product space $V^{1,3}$ is the Lorentz group $L = \mathrm{O}(V^{1,3})$ which is the group of linear isometries of $V^{1,3}$. Thus $g \in \mathrm{O}(V^{1,3}, \langle .,. \rangle)$ if and only if

$$\langle gv, gw \rangle = \langle v, w \rangle$$

for all $v, w \in V^{1,3}$.

Now the origin is a preferred point for a vector space but is not a physical reality and so we want to introduce the appropriate metric affine space.

Definition 2.15 ***Minkowski space*** *is the metric affine space M^{1+3} (unique up to isometry) with difference space given by a Lorentz scalar product space $V^{1,3}$ (Lorentz space).*

Minkowski space is sometimes referred to as Lorentzian affine space.

The group of coordinate transformations appropriate to M^{1+3} is described the group of affine transformations of M^{1+3} whose linear part is an element of $\mathrm{O}(V^{1,3}, \langle .,. \rangle)$. This is called the Poincaré group P. If we pick an origin $p \in M^{1+3}$ an orthonormal basis for $V^{1,3}$ then we may identify M^{1+3} with $\mathbb{R}^{1,3}$ (as an affine space[5]). Having made this arbitrary choice the Lorentz group is identified with the group of matrices $\mathrm{O}(1,3)$ which is defined as the set of all 4×4 matrices Λ such that

$$\Lambda^t \eta \Lambda = \eta.$$

and a general Poincaré transformation is of the form $x \mapsto \Lambda x + x_0$ for $x_0 \in \mathbb{R}^{1,3}$ and $\Lambda \in \mathrm{O}(1,3)$. We also have an alternative realization of M^{1+3} as the set of all column vectors of the form

$$\begin{bmatrix} 1 \\ x \end{bmatrix} \in \mathbb{R}^5$$

where $x \in \mathbb{R}^{1,3}$. Then the a Poincaré transformation is given by a matrix of the form the group of matrices of the form

$$\begin{bmatrix} 1 & 0 \\ b & Q \end{bmatrix} \text{ for } Q \in \mathrm{O}(1,3).$$

[5]The reader will be relieved to know that we shall eventually stop needling the reader with the pedantic distinction between $\mathbb{R}^n$ and $\mathbb{A}^n$, $\mathbb{E}^n$ and so on.

Basic fact 3. The spacetime interval between two events E_1 and E_2 in Minkowski spacetime is may be calculated in any (Lorentz) inertial coordinates by $\Delta\tau := -(\Delta x^0)^2 + \sum_{i=1}^{4}(\Delta x^i)^2$ where $\Delta x = x(E_2) - x(E_1)$. The result is independent of the choice of coordinates. Spacetime separation in M^4 is "absolute". On the other hand, in Minkowski spacetime spatial separation and temporal separation are both relative concepts and only make sense within a particular coordinate system. It turns out that real spacetime is best modeled by Minkowski spacetime (at least locally and in the absence of strong gravitational fields). This has some counter intuitive implications. For example, it does not make any sense to declare that some supernova exploded into existence at the precise time of my birth. There is simply no fact of the matter. It is similar to declaring that the star Alpha Centaury is "above" the sun. Now if one limits oneself to coordinate systems that have a small relative motion with respect to each other then we may speak of events occurring at the same time (approximately). If one is speaking in terms of precise time then even the uniform motion of a train relative to an observer on the ground destroys our ability to declare that two events happened at the same time. If the fellow on the train uses the best system of measurement (best inertial coordinate system) available to him and his sister on the ground does the same then it is possible that they may not agree as to whether or not two firecrackers, one on the train and one on the ground, exploded at the same time or not. It is also true that the question of whether the firecrackers exploded when they were 5 feet apart or not becomes relativized. The sense in which the spaces we have introduced here are "flat" will be made clear when we study curvature.

In the context of special relativity, there are several common notations used for vectors and points. For example, if (x^0, x^1, x^2, x^3) is a Lorentz inertial coordinate system then we may also write (t, x, y, z) or $(t, \mathbf{r})$ where $\mathbf{r}$ is called the spatial position vector. In special relativity it is best to avoid curvilinear coordinates because the simple from that the equations of physics take on when expressed in the rectangular inertial coordinates is ruined in a noninertial coordinate systems. This is implicit in all that we do in Minkowski space. Now while the set of points of M^4 has lost its vector space structure so that we no longer consider it legitimate to add points, we still have the ability to take the difference of two points and the result will be an element of the scalar product space $V^{1,3}$. If one takes the difference between p and q in M^4 we have an element $v \in V$ and if we want to consider this vector as a tangent vector based at p then we can write it as (p, v) or as v_p. To get the expression for the inner product of the tangent vectors $(p, v) = \overrightarrow{pq_1}$ and $(p, w) = \overrightarrow{pq_2}$ in coordinates (x^μ), let $v^\mu := x^\mu(q_1) - x^\mu(p)$ and $w^\mu := x^\mu(q_2) - x^\mu(p)$ and then calculate: $\langle v_p, w_p \rangle = \eta_{\mu\nu} v^\mu w^\nu$. Each tangent space is naturally an inner product. Of course, associated with each inertial coordinate system (x^μ) there is a set of $4-$vectors fields which are the coordinate vector fields denoted by $\partial_0, \partial_1, \partial_2$,and ∂_3. A contravariant vector or vector field $v = v^\mu \partial_\mu$ on M^4 has a twin in covariant form[6]. This is the covector (field) $v^\flat = v_\mu dx^\mu$ where $v_\mu := \eta_{\mu\nu} v^\nu$. Similarly

[6]This is a special case of an operation that one can use to get 1-forms from vector fields

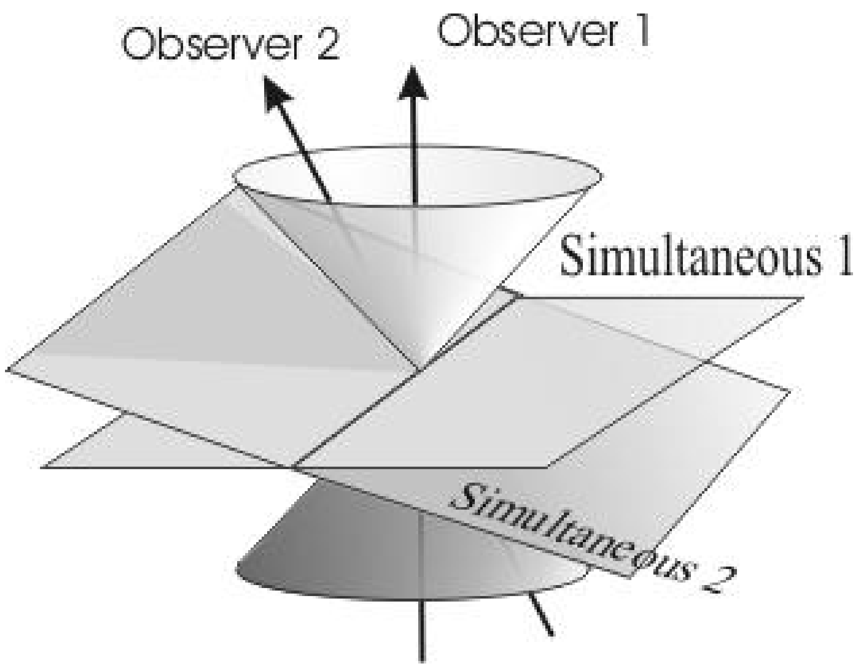
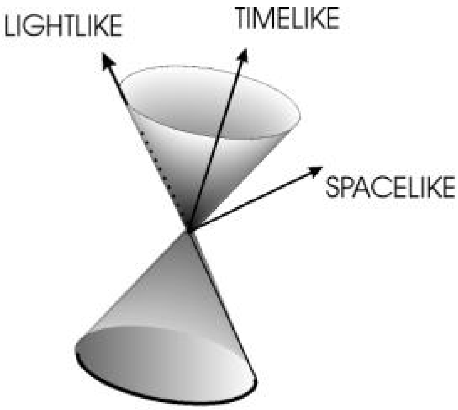


Figure 2.1: Relativity of Simultaneity



if $\alpha = \alpha_\mu dx^\mu$ is a covector field then there is an associated vector field given $\alpha^\# = \alpha^\mu \partial_\mu$ where $\alpha^\mu := \alpha_\nu \eta^{\mu\nu}$ and the matrix $(\eta^{\mu\nu})$ is the inverse of $(\eta_{\mu\nu})$ which in this context might seem silly since $(\eta_{\mu\nu})$ is its own inverse. The point is that this anticipates a more general situation and also maintains a consistent use of index position. The Lorentz inner product is defined for any pair of vectors based at the same point in spacetime and is usually called the Lorentz metric-a special case of a semi-Riemannian metric which is the topic of a later chapter. If v^μ and w^μ are the components of two vectors in the current Lorentz frame then $\langle v, w\rangle = \eta_{\mu\nu} v^\mu w^\nu$.

Definition 2.16 *A 4-vector v is called **space-like** if and only if $\langle v, v\rangle < 0$, **time-like** if and only if $\langle v, v\rangle > 0$ and **light-like** if and only if $\langle v, v\rangle = 0$. The set of all light-like vectors at a point in Minkowski space form a double cone in $\mathbb{R}^4$ referred to as the **light cone**.*

Remark 2.1 (Warning) *Sometimes the definition of the Lorentz metric given is opposite in sign from the one we use here. Both choices of sign are popular. One consequence of the other choice is that time-like vectors become those for which $\langle v, v\rangle < 0$.*

Definition 2.17 *A vector v based at a point in M^4 such that $\langle \partial_0, v\rangle > 0$ will be called **future pointing** and the set of all such forms the interior of the "future" light-cone.*

Definition 2.18 *A Lorentz transformation that sends future pointing timelike vector to future pointing timelike vectors is called an orthochronous Lorentz transformation.*

Now an important point is that there are two different ways that physics gives rise to a vector and covector field. The first is exemplified by the case

---

and visa versa on any semi-Riemannian manifold and we will explain this in due course.

of a single particle of mass m in a state of motion described in our coordinate system by a curve $\gamma : t \to (t, x(t), y(t), z(t))$ such that $\dot{\gamma}$ is a timelike vector for all parameter values t. The 3-velocity is a concept that is coordinate dependent and is given by $\mathbf{v} = (\frac{dx}{dt}(t), \frac{dy}{dt}(t), \frac{dz}{dt}(t))$. In this case, the associated 4−velocity u is a vector field along the curve γ an defined to be the unit vector (that is $\langle u, u \rangle = -1$) which is in the direction of $\dot{\gamma}$. The the contravariant 4−momentum is the 4-vector field along γ given by and $p = mu$. The other common situation is where matter is best modeled like a fluid or gas.

Now if γ is the curve which gives the career of a particle then the spacetime interval between two events in spacetime given by $\gamma(t)$ and $\gamma(t+\epsilon)$ is $\langle \gamma(t+\epsilon) - \gamma(t), \gamma(t+\epsilon) - \gamma(t) \rangle$ and should be a timelike vector. If ϵ is small enough then this should be approximately equal to the time elapsed on an ideal clock traveling with the particle. Think of zooming in on the particle and discovering that it is actually a spaceship containing a scientist and his equipment (including say a very accurate atomic clock).The actual time that will be observed by the scientist while traveling from significantly separated points along her career, say $\gamma(t_1)$ and $\gamma(t)$ with $t > t_1$ will be given by

$$\tau(t) = \int_{t_1}^{t} |\langle \dot{\gamma}, \dot{\gamma} \rangle| \; dt$$

Standard arguments with change of variables show that the result is independent of a reparameterization. It is also independent of the choice of Lorentz coordinates. We have skipped the physics that motivates the interpretation of the above integral as an elapsed time but we can render it plausible by observing that if γ is a straight line $t \mapsto (t, x_1 + v_1 t, x_2 + v_2 t, x_3 + v_3 t)$ which represents a uniform motion at constant speed then by a change to a new Lorentz coordinate system and a change of parameter the path will be described simply by $t \to (t, 0, 0, 0)$. The scientist is at rest in her own Lorentz frame. Now the integral reads $\tau(t) = \int_0^t 1 \, dt = t$. For any timelike curve γ the quantity $\tau(t) = \int_{t_1}^{t} |\langle \dot{\gamma}, \dot{\gamma} \rangle|^{1/2} \; dt$ is called the **proper time** of the curve from $\gamma(t_1)$ to $\gamma(t)$.

The famous twins paradox is not really a true paradox since an adequate explanation is available and the counter-intuitive aspects of the story are actually physically correct. In short the story goes as follows. Joe's twin Bill leaves in a spaceship and travels at say 98% of the speed of light to a distant star and then returns to earth after 100 years of earth time. Let use make two simplifying assumptions which will not change the validity of the discussion. The first is that the earth, contrary to fact, is at rest in some inertial coordinate system (replace the earth by a space station in deep space if you like). The second assumption is that Joe's twin Bill travels at constant velocity on the forward and return trip. This entails an unrealistic instant deceleration and acceleration at the star; the turn around but the essential result is the same if we make this part more realistic. Measuring time in the units where $c = 1$, the first half of Bill's trip is given $(t, .98t)$ and second half is given by $(t, -.98t)$. Of course, this entails that in the earth frame the distance to the star is $.98 \frac{lightyears}{year} \times 100$

years = 98 light-years. Using a coordinate system fixed relative to the earth we calculate the proper time experienced by Bill:

$$\int_0^{100} |\langle \dot\gamma, \dot\gamma\rangle| \, dt = \int_0^{50} \sqrt{|-1 + (.98)^2|} dt + \int_0^{50} \sqrt{|-1 + (-.98)^2|} dt$$
$$= 2 \times 9.949\,9 = 19.900$$

Bill is 19.9 years older when we returns to earth. On the other hand, Joe's has aged 100 years! One may wonder if there is not a problem here since one might argue that from Joe's point of view it was the earth that travelled away from him and then returned. This is the paradox but it is quickly resolved once it is pointed out that the is not symmetry between Joe's and Bill's situation. In order to return to earth Bill had to turn around which entail an acceleration and more importantly prevented a Bill from being stationary with respect to any single Lorentz frame. Joe, on the other hand, has been at rest in a single Lorentz frame the whole time. The age difference effect is real and a scaled down version of such an experiment involving atomic clocks put in relative motion has been carried out and the effect measured.

2.2.6 Hyperbolic Geometry

We have looked at affine spaces, Euclidean spaces, Minkowski space and Galilean spacetime. Each of these has an associated group and in each case straight lines are maps to straight lines. In a more general context the analogue of straight lines are called geodesics a topic we will eventually take up in depth. The notion of distance makes sense in a Euclidean space essentially because each tangent space is an inner product space. Now each tangent space of $\mathbf{E}^n$ is of the form $\{p\} \times \mathbb{R}^n$ for some point p. This is the set of tangent vectors at p. If we choose an orthonormal basis for $\mathbb{R}^n$, say $e_1, ..., e_n$ then we get a corresponding orthonormal basis in each tangent space which we denote by $e_{1,p}, ..., e_{n,p}$ and where $e_{i,p} := (p, e_i)$ for $i = 1, 2, ..., n$. With respect to this basis in the tangent space the matrix which represents the inner product is the identity matrix $I = (\delta_{ij})$. This is true uniformly for each choice of p. One possible generalization is to let the matrix vary with p. This idea eventually leads to Riemannian geometry. We can give an important example right now. We have already seen that the subgroup $\mathrm{SL}(2, \mathbb{R})$ of the group $\mathrm{SL}(2, \mathbb{C})$ also acts on the complex plane and in fact fixes the upper half plane $\mathbb{C}^+$. In each tangent space of the upper half plane we may put an inner product as follows. If $v_p = (p, v)$ and $w_p = (p, w)$ are in the tangent space of p then the inner product is

$$\langle v_p, w_p \rangle := \frac{v^1 w^1 + v^2 w^2}{y^2}$$

where $p = (x, y) \sim x + iy$ and $v = (v^1, v^2)$, $w = (w^1, w^2)$. In this context, the assignment of an inner product the tangent space at each point is called a metric. We are identifying $\mathbb{R}^2$ with $\mathbb{C}$. Now the length of a curve $\gamma : t \mapsto (x(t), y(t))$

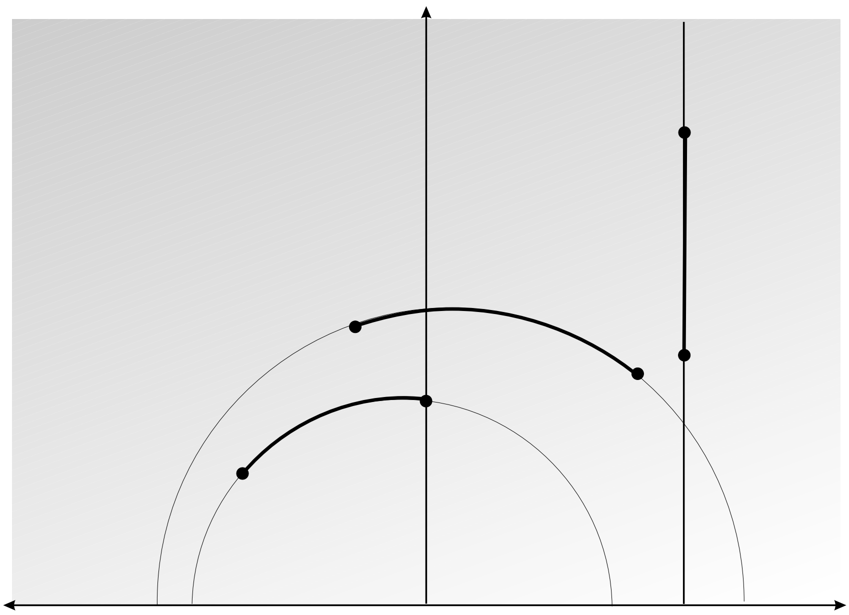
defined on $[a, b]$ and with image in the upper half-plane is given by

$$\int_a^b \|\dot{\gamma}(t)\| \, dt = \int_a^b \left(\frac{\dot{x}(t)^2 + \dot{y}(t)^2}{y(t)^2} \right)^{1/2} dt.$$

We may define arc length starting from some point $p = \gamma(c)$ along a curve $\gamma : [a, b] \to \mathbb{C}^+$ in the same way as was done in calculus of several variables:

$$s := l(t) = \int_a^t \left(\frac{\dot{x}(\tau)^2 + \dot{y}(\tau)^2}{y(\tau)^2} \right)^{1/2} d\tau$$

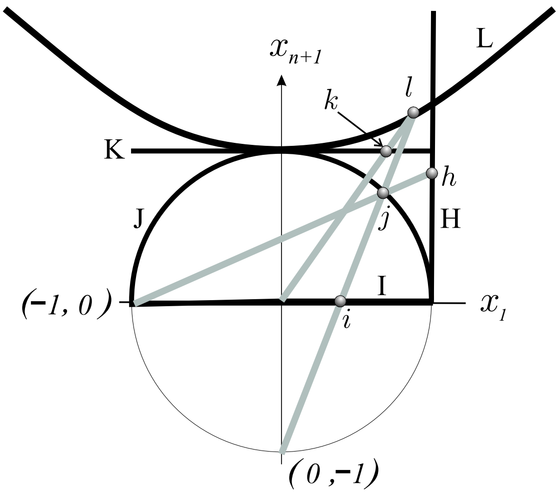
The function $l(t)$ is invertible and it is then possible to reparameterize the curve by arc length $\widetilde{\gamma}(s) := \gamma(l^{-1}(s))$.The distance between any two points in the upper half-plane is the length of the shortest curve that connects the two points. Of course, one must show that there is always a shortest curve. It turns out that the shortest curves, the geodesics, are curved segments lying on circles which meet the real axis normally, or are vertical line segments.

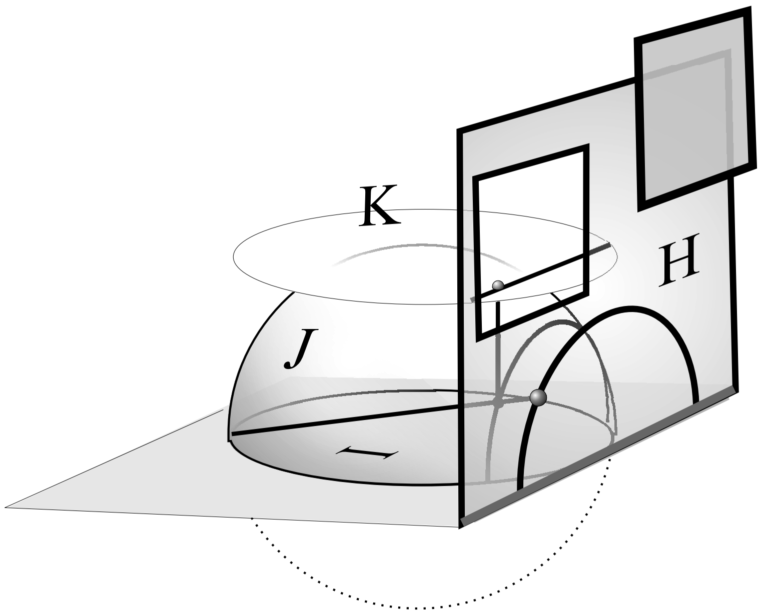


The upper half plane with this notion of distance is called the Poincaré upper half plane and is a realization of an abstract geometric space called the hyperbolic plane. The geodesics are the “straight lines” for this geometry.

2.2.7 Models of Hyperbolic Space

hjj (do this section-follow’flavors of geometry)


L
x_{n+1}
k
l
K
h
J
j
H
I
$(-1, 0)$
x_1
i
$(0, -1)$


K
H
J
I

$$
\begin{aligned}
H &= \{(1, x_2, ..., x_{n+1}) : x_{n+1} > 0\} \\
I &= \{(x_1, ..., x_n, 0) : x_1^2 + \cdots + x_n^2 < 1\} \\
J &= \{(x_1, ..., x_{n+1}) : x_1^2 + \cdots + x_{n+1}^2 = 1 \text{ and } x_{n+1} > 0\} \\
K &= \{(x_1, ..., x_n, 1) : x_1^2 + \cdots + x_n^2 < 1\} \\
L &= \{(x_1, ..., x_n, x_{n+1}) : x_1^2 + \cdots + x_n^2 - x_{n+1}^2 = -1 \text{ and } x_{n+1} > 0\}
\end{aligned}
$$

$$
\begin{aligned}
&\alpha : J \to H; \ (x_1, ..., x_{n+1}) \mapsto (1, 2x_2/(x_1+1), ..., 2x_{n+1}/(x_1+1)) \\
&\beta : J \to I; \ (x_1, ..., x_{n+1}) \mapsto (x_1/(x_{n+1}+1), ..., x_n/(x_{n+1}+1), 0) \\
&\gamma : K \to J; \ (x_1, ..., x_n, 1) \mapsto (x_1, ..., x_n, \sqrt{1 - x_1^2 - \cdots - x_n^2}) \\
&\delta : L \to J; \ (x_1, ..., x_{n+1}) \mapsto (x_1/x_{n+1}, ..., x_n/x_{n+1}, 1/x_{n+1})
\end{aligned}
$$

The map a is stereographic projection with focal point at $(-1, 0, ..., 0)$ and maps j to h in the diagrams. The map β is a stereographic projection with focal point at $(0, ..., 0, -1)$ and maps j to i in the diagrams. The map γ is vertical orthogonal projection and maps k to j in the diagrams. The map δ is stereographic projection with focal point at $(0, ..., 0, -1)$ as before but this time projecting onto the hyperboloid L.

$$
\begin{aligned}
ds_H^2 &= \frac{dx_2^2 + \cdots + dx_{n+1}^2}{x_{n+1}^2}; \\
ds_I^2 &= 4\frac{dx_1^2 + \cdots + dx_n^2}{(1 - x_1^2 - \cdots - x_n^2)^2}; \\
ds_J^2 &= \frac{dx_1^2 + \cdots + dx_n^2}{x_{n+1}^2} \\
ds_K^2 &= \frac{dx_1^2 + \cdots + dx_n^2}{1 - x_1^2 - \cdots - x_n^2} + \frac{(x_1 dx_1 + \cdots + x_n dx_n)^2}{(1 - x_1^2 - \cdots - x_n^2)^2} \\
dx_L^2 &= dx_1^2 + \cdots + dx_n^2 - dx_{n+1}^2
\end{aligned}
$$

To get a hint of the kind of things to come, notice that we can have two geodesics which start at nearby points and start of in the same direction and yet the distance between corresponding points increases. In some sense, the geometry acts like a force which (in this case) repels nearby geodesics. The specific invariant responsible is the curvature. Curvature is a notion that has a generalization to a much larger context and we will see that the identification of curvature with a force field is something that happens in both Einstein's general theory of relativity and also in gauge theoretic particle physics. One of the goals of this book is to explore this connection between force and geometry.

Another simple example of curvature acting as a force is the following. Imaging that the planet was completely spherical with a smooth surface. Now imagine the people a few miles apart but both on the equator. Give each person a pair of roller skates and once they have them on give them a simultaneous push toward the north. Notice that they start out with parallel motion. Since they will eventually meet at the north pole the distance between them must be shrinking. What is pulling them together. Regardless of whether one wants to call the cause of their coming together a force or not, it is clear that it is the curvature of the earth that is responsible. Readers familiar with the basic idea behind General Relativity will know that according to that theory, the "force" of gravity is due to the curved shape of 4-dimensional spacetime. The origin of the curvature is said to be due to the distribution of mass and energy in space.

2.2.8 The Möbius Group

Let $\mathbb{C}^{+\infty}$ denote the set $\mathbb{C} \cup \{\infty\}$. The topology we provide $\mathbb{C}^{+\infty}$ is generated by the open subsets of $\mathbb{C}$ together with sets of the form $O \cup \{\infty\}$ where O is the compliment of a compact subset of $\mathbb{C}$. This is the $1-$**point compactification** of $\mathbb{C}$. Topologically, $\mathbb{C}^{+\infty}$ is just the sphere S^2.

Now consider the group $\mathrm{SL}(2, \mathbb{C})$ consisting of all invertible 2×2 complex matrices with determinant 1. We have an action of $\mathrm{SL}(2, \mathbb{C})$ on $\mathbb{C}^{+\infty}$ given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az+b}{cz+d}.$$

For any fixed $A \in \mathrm{SL}(2, \mathbb{C})$ them map $z \mapsto A \cdot z$ is a homeomorphism (and much more as we shall eventually see). Notice that this is not the standard action of $\mathrm{SL}(2, \mathbb{C})$ on $\mathbb{C}^2$ by multiplication $\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \mapsto A \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$ but there is a relationship between the two actions. Namely, let

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

and define $z = z_1/z_2$ and $w = w_1/w_1$. Then $w = A \cdot z = \frac{az+b}{cz+d}$. The two component vectors $\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$ are sometimes called spinors in physics.

Notice is that for any $A \in \mathrm{SL}(2, \mathbb{C})$ the homeomorphisms $z \mapsto A \cdot z$ and $z \mapsto (-A) \cdot z$ are actually equal. Thus group theoretically, the set of all distinct transformations obtained in this way is really the quotient group $\mathrm{SL}(2, \mathbb{C})/\{I, -I\}$ and this is called the **Möbius group** or group of Möbius transformations.

There is quite a bit of geometry hiding in the group $\mathrm{SL}(2, \mathbb{C})$ and we will eventually discover a relationship between $\mathrm{SL}(2, \mathbb{C})$ and the Lorentz group.

When we add the notion of time into the picture we are studying spaces of "events" rather than "literal" geometric points. On the other hand, the spaces of evens might be considered to have a geometry of sorts and so in that sense the events are indeed points. An approach similar to how we handle Euclidean space will allow us to let spacetime be modeled by a Cartesian space $\mathbb{R}^4$; we find a family of coordinates related to the standard coordinates by the action of a group. Of course, in actual physics, the usual case is where space is 3-dimensional and spacetime is $4-$dimensional so lets restrict attention this case. But what is the right group? What is the right geometry? We now give two answers to this question. The first one corresponds to intuition quite well and is implicit in the way physics was done before the advent of special relativity. The idea is that there is a global measure of time that applies equally well to all points of $3-$dimensional space and it is unique up to an affine change of parameter $t \mapsto t' = at + b$. The affine change of parameter corresponds to a change in units.

2.3 Local classical fields

We know that the set of all $n \times n$ matrices is isomorphic as a vector space to $\mathbb{R}^{n^2}$ but there is a good reason that we don't usually think of matrices as $nm-$tuples. Rather, the nm numbers are arranged in a rectangle to facilitate the rule for matrix multiplication. The n^2 numbers that make up a matrix are ordered not only by the natural order of the indexing set $[n] = \{1, 2, ..., n\}$ but also by "position". The positioning of the indices is a device that, among other things, helps encode the transformational properties under the group $G\mathrm{L}(n, \mathbb{R})$. We shall now introduce a very classical but perfectly good definition of a tensor space as indexed arrays of numbers where the position of the index has significance. Let $T^k_l(\mathbb{R}^n)$ denote the vector space of functions $[n] \times ... \times [n] \times .. \times [n] \to \mathbb{R}$ of the form $t : i_1, ..., i_k, j_1, ..., j_l \mapsto t^{\ i_1 \ldots i_k}\big|_{j_1 \ldots j_l}$. These functions of the index set correspond to (the components of) classical tensors of type k, l. Notice that the "slots" are arranged not only by first, second, etc. but also by "up" and "down". Thus we can think of the tensor t as the set of numbers $\{t^{\ i_1 \ldots i_k}\big|_{j_1 \ldots j_l} : 1 \leq i_1, ..., i_k, j_1, ..., j_l \leq n\}$. Now let G be a subgroup of the group $G\mathrm{L}(n, \mathbb{R})$. The group acts on $T^k_l(\mathbb{R}^n)$ by the following standard recipe: If $g := (g^i_{\ j}) \in G$ then $g \cdot r = r'$ where

$$(t')^{\ i_1 \ldots i_k}\big|_{j_1 \ldots j_l} = g^{i_1}_{\ r_1} \cdots g^{i_k}_{\ r_l} t^{\ r_1 \ldots r_k}|_{s_1 \ldots s_l} \left(g^{-1}\right)^{s_1}_{\ j_1} \cdots \left(g^{-1}\right)^{s_k}_{\ j_l} \tag{2.1}$$

and the Einstein summation convention is in force so that all indices[7] $r_1 ... r_k$ and $s_1 ... s_k$ are summed over the fixed index set $[n] = \{1, 2, ..., n\}$. Now recall that if a matrix $(A^i_{\ j})$ represents a linear transformation $L : \mathbb{R}^n \to \mathbb{R}^n$ with respect to a basis $\{v_1, ..., v_n\}$ then the matrix (A'^i_j) represents the same linear transformation with respect to the new basis $\{v'_1, ..., v'_n\}$ if

$$v'_i = v_r g^r_{\ i}$$
$$\text{and}$$
$$A'^i_j = g^i_{\ r} A^r_{\ s} \left(g^{-1}\right)^s_{\ r}$$

[7]We have employed a little joke here: Since we sum over repeated indices we have used r and s for the dummy indices to sum over because of the association to "repeat" and "sum".

In the same way we say that if $(t^{\,i_1\ldots i_k}\big|_{j_1\ldots j_l})$ represents an "abstract tensor" with respect to the basis $\{v_1, ..., v_n\}$ then $(t'^{\,i_1\ldots i_k}\big|_{j_1\ldots j_l})$ represents the same abstract tensor with respect to $\{v'_1, ..., v'_n\}$ if $v'_i = v_r g^r{}_i$ and $(t')^{\,i_1\ldots i_k}\big|_{j_1\ldots j_l} = g^{i_1}{}_{r_1} \cdots g^{i_k}{}_{r_l} t^{\,r_1\ldots r_k}\big|_{s_1\ldots s_l} \left(g^{-1}\right)^{s_1}{}_{j_1} \cdots \left(g^{-1}\right)^{s_k}{}_{j_l}$ hold. For now we shall stick to one basis unless indication otherwise.

Definition 2.19 *A local expression for a* ***tensor field*** *of type k, l on an open set $U \subset \mathbb{R}^n$ is a smooth map $\phi : U \to T^k_l(\mathbb{R}^n)$. We write $\phi^{\,i_1\ldots i_k}\big|_{j_1\ldots j_l} : x \mapsto \phi^{\,i_1\ldots i_k}\big|_{j_1\ldots j_l}(x)$ for the individual component functions.*

In the sequel, we will give other more sophisticated and more general descriptions of tensor fields which will make sense in the setting of differentiable manifolds (our ultimate setting). But even these tensor fields do not include all the types of fields that are used in physical theories. Specifically, we have not yet even included the notion of spinor fields. In order to be as general as possible at this juncture we shall just define a (classical) field as a smooth map $\phi : U \to \mathrm{V}$ where V is a finite dimensional vector space that might support a right action by a matrix group which often has a definite physical or geometric significance. Later on we shall encounter the more general notion of a field as a map from a spacetime manifold into another manifold.

The derivative $D\phi : U \to L(\mathbb{R}^n, \mathrm{V})$ is a map that is determined by the set of all partial derivatives $\partial_i \phi := \frac{\partial \phi}{\partial x^i}$. Let $\partial\phi$ denote the $n-$tuple $(\partial_i\phi)$ of partial derivatives. Thus for a fixed $x \in U$ we have $\partial\phi(x) \in \mathrm{V}^n := \mathrm{V} \times \cdots \times \mathrm{V}$. Let us consider the important case of a field defined on $\mathbb{R}^4 = \mathbb{R} \times \mathbb{R}^3$. It is traditional to denote the variables in $\mathbb{R}^4$ as $x = (x^0, x^1, x^2, x^3)$ instead of (t, x, y, z). Notice that our index set is now $\{0, 1, 2, 3\}$. The $x^0 = t$ is to be thought of as time and it is sometimes useful to think of ϕ as giving a one-parameter family of fields ϕ_t on $\mathbb{R}^3$ which for each t is given by $\phi_t(x^1, x^2, x^3) := \phi(t, x^1, x^2, x^3)$. Let $\phi : \mathbb{R} \times \mathbb{R}^3 \to \mathrm{V}$ be such a field. As an example we might take $\mathrm{V} = T^k_l(\mathbb{R}, 4)$. A **Lagrangian density** is a function $\mathcal{L} : \mathrm{V} \times \mathrm{V}^4 \to \mathbb{R}$. Let The associated "action over $U \subset \mathbb{R}^4$" is the function defined by

$$S(\phi) = \int_U \mathcal{L}(\phi(x), \partial\phi(x)) d^4x$$

where d^4x denotes the volume element given by $dx^0dx^1dx^2dx^3$. Often $U = I \times O$ for some open subset $O \subset \mathbb{R}^3$ and where $I = (a, b)$ is a fixed interval. This is similar to the earlier situation where we integrated a Lagrangian over an interval. There the functional was defined on a space of curves $t \mapsto (q^i(t))$. What has happened now is that the index i is replaced by the space variable

$\vec{x} = (x^1, x^2, x^3)$ and we have the following dictionary:

$$\begin{array}{lll} i & \rightsquigarrow & \vec{x} \\ q & \rightsquigarrow & \phi \\ q^i & \rightsquigarrow & \phi(.,\vec{x}) \\ q^i(t) & \rightsquigarrow & \phi_t(\vec{x}) = \phi(t,\vec{x}) = \phi(x) \\ p^i(t) & \rightsquigarrow & \partial\phi(x) \end{array}$$

$$\begin{array}{lll} L(q,p) & \rightsquigarrow & \int_{V\subset\mathbb{R}^3} \mathcal{L}(\phi,\partial\phi)d^3x \\ S = \int L(\mathbf{q},\dot{\mathbf{q}})dt & \rightsquigarrow & S = \int\int \mathcal{L}(\phi,\partial\phi)d^3x dt \end{array}$$

where $\partial\phi = (\partial_0\phi, \partial_1\phi, \partial_2\phi, \partial_3\phi)$. So in a way, the mechanics of classical massive particles is classical field theory on the space with three points which is the set $\{1, 2, 3\}$. Or we can view field theory as infinitely many particle systems indexed by points of space. In other words, a system with an infinite number of degrees of freedom.

Actually, we have only set up the formalism of scalar and tensor fields and have not, for instance, set things up to cover internal degrees of freedom like spin. However, we will discuss spin later in this text. Let us look at the formal variational calculus of field theory. We let $\delta\phi$ be a variation which we might later assume to vanish on the boundary of some region in space-time $U = I \times O \subset \mathbb{R} \times \mathbb{R}^3 = \mathbb{R}^4$. We will now use notation in style common in the physics literature but let us be perfectly clear about what the notation means. Let us denote the variables in $\mathcal{L}$ by u, v_0, v_1, v_2, v_3 so that we write $\mathcal{L}(u, v_0, v_1, v_2, v_3)$. Then the meaning of $\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}$ is

$$\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}(x) = \frac{\partial\mathcal{L}}{\partial v_\mu}(\phi(x), \partial_0\phi(x), \partial_1\phi(x), \partial_2\phi(x), \partial_3\phi(x))$$

Similarly, $\frac{\partial\mathcal{L}}{\partial\phi}(x) = \frac{\partial\mathcal{L}}{\partial u}(\phi(x))$. We will also see the expression $\frac{\partial\mathcal{L}}{\partial\phi} \cdot \delta\phi$. Here $\delta\phi$ is just an arbitrary (variation) field $\delta\phi : U \to \mathrm{V}$ and since $\frac{\partial\mathcal{L}}{\partial\phi}(x)$ is now a linear map $\mathrm{V} \to \mathbb{R}$ we interpret $\frac{\partial\mathcal{L}}{\partial\phi} \cdot \delta\phi$ to mean the function $U \to \mathbb{R}$ given by

$$x \mapsto \frac{\partial\mathcal{L}}{\partial\phi}(x) \cdot \delta\phi(x)$$

Similarly we can make sense of expressions like $\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \cdot \delta\phi$. With this in mind we take the variation of the action functional S (defined above) in the direction $\delta\phi$. We have

$$\begin{aligned} \delta S &= \int_U \left(\frac{\partial\mathcal{L}}{\partial\phi} \cdot \delta\phi + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \cdot \partial_\mu\delta\phi \right) d^4x \\ &= \int_U \partial_\mu \left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \cdot \delta\phi \right) d^4x + \int_U \left(\frac{\partial\mathcal{L}}{\partial\phi} - \partial_\mu \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \right) \cdot \delta\phi d^4x \end{aligned}$$

Now the first term would vanish by the divergence theorem if $\delta\phi$ vanished on the boundary ∂U. If ϕ were a field that were stationary under such variations then $\delta S = 0$ for all $\delta\phi$ including those that vanish on ∂U. Thus

$$\delta S = \int_U \left(\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) \cdot \delta\phi d^4x = 0$$

for all $\delta\phi$ that vanish on ∂U so we can easily conclude that Lagrange's equation holds for these "stationary" fields ϕ

$$\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} = 0$$

These are the field Euler-Lagrange field equations.

Example 2.8 *Let $A : \mathbb{R}^4 \to T_1^0(\mathbb{R}^4) \cong \mathbb{R}^4$ and $J : \mathbb{R}^4 \to T_0^1(\mathbb{R}^4) \cong \mathbb{R}^4$ be smooth maps and define $F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}$. We think of $F_{\mu\nu}$ as defining a tensor field of type $(0, 2)$. If A is a the electromagnetic vector potential (in covariant form) then the $F_{\mu\nu}$ are the components of the electromagnetic field. In fact,*

$$(F_{\mu\nu}) = \begin{bmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & B_z & -B_y \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{bmatrix}.$$

(see appendix). The interpretation of J is that it is the "charge current density 4-vector". Let $F^{\nu\mu} = \eta^{\nu r}\eta^{\mu s}F_{rs}$ where $(\eta^{ij}) = diag(-1, 1, 1, 1)$. Let $\mathcal{L}(A, \partial A) = -\frac{1}{4}F^{\nu\mu}F_{\nu\mu} - J^\mu A_\mu$ so that

$$S(A) = \int_{\mathbb{R}^4} \left(-\frac{1}{4}F^{\nu\mu}F_{\nu\mu} - J^\mu A_\mu \right) \, dx^4$$

We calculate the Euler-Lagrange equations for $\mathcal{L}$. Assume $\delta A = (\delta A_\nu)$ is field with support inside a compact domain Ω then

$$\begin{aligned} \delta S|_A (\delta A) &= \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} S(A + \varepsilon \delta A) = \\ &= \int_{I\times\mathbb{R}^3} -\frac{1}{2}F^{\nu\mu}\delta F_{\nu\mu} - J^\mu \delta A_\mu \;\; dx^4 \\ &= \int_\Omega -\frac{1}{2}F^{\nu\mu}\left(\frac{\partial \delta A_\nu}{\partial x^\mu} - \frac{\partial \delta A_\mu}{\partial x^\nu}\right) - J^\mu \delta A_\mu \;\; dx^4 \\ &= \int_\Omega -F^{\nu\mu}\frac{\partial \delta A_\nu}{\partial x^\mu} - J^\mu \delta A_\mu \;\; dx^4 \\ &= \int_\Omega \left(\frac{\partial F^{\nu\mu}}{\partial x^\mu} - J^\mu\right) \delta A_\mu \;\; dx^4 \end{aligned}$$

Now since δA arbitrary we get $\frac{\partial F^{\nu\mu}}{\partial x^\mu} = J^\mu$. (These are the inhomogeneous Maxwell's equations.)

2.4 Problem Set

1. Calculate the distance between the points $z_1 = 1 + i$ and $z_2 = 1 - i$ in the Poincaré upper half plane.

Chapter 3

Differentiable Manifolds

An undefined problem has an infinite number of solutions.
-Robert A. Humphrey

3.1 Motivation

We have already seen that affine spaces and Euclidean spaces have families of coordinate systems. We may strip away the affine structure and (a fortiori) the metric structure of $\mathbb{R}^n$ leaving only a topological space that nevertheless supports various coordinates systems most of which are much more general that the affine coordinates. These are the curvilinear coordinates. Most general coordinate systems consist of functions which are not generally defined on all of the space but rather on a proper open subset (think of polar coordinates on the plane which has an ambiguity at least at the origin). Surfaces provide another motivation for the concept of a differentiable manifold. We are all familiar with spherical coordinates θ, ϕ on the unit sphere $S^2 := \{x \in \mathbb{R}^3 : x \cdot x = 1\}$. In fact, there are many such spherical coordinates since we are free to choose which point will play the role of the north pole. The various spherical coordinates can be expressed in terms of each other in much the same way that polar coordinates may be expressed in terms of rectangular coordinates (and visa versa) on the plane. The individual coordinate functions from one coordinate system will be differentiable functions of the coordinates functions from another coordinate system. This allows us to unambiguously talk about the differentiability of functions defined on the sphere. A function which appears differentiable when expressed in terms of one set of spherical coordinates will also be differentiable when expressed in another set of spherical coordinates. The idea of imposing a family of coordinate systems, also called coordinate charts, on a topological space that are mutually differentiable functions of each other is the basic idea behind the notion of a differentiable manifold. We shall see that the matrix groups that we have already encountered are also differentiable manifolds (they support coordinate systems). This already begins to show how general and useful is the notion of a differentiable manifold. We first introduce a the more general notion of a topological manifold where coordinates exist but are only assumed to be continuously related to each other.

3.2 Topological Manifolds

A **refinement** of an open cover $\{U_\beta\}_{\beta\in B}$ of a topological space is another open cover $\{V_i\}_{i\in I}$ such that every open set from the second cover is contain in at least one open set from the original cover. This means that means that if $\{U_\beta\}_{\beta\in B}$ is the given cover of X, then a refinement may be described as a cover $\{V_i\}_{i\in I}$ together with a set map $i \mapsto \beta(i)$ of the index sets $I \to B$ such that $V_i \subset U_{\beta(i)}$. We say that $\{V_i\}_{i\in I}$ is a **locally finite** cover if in every point of X has a neighborhood that intersects only a finite number of the sets from $\{V_i\}_{i\in I}$.

We recall a couple of concepts from point set topology: A topological space X is called **paracompact** if it is Hausdorff and if every open cover of X has a refinement to a locally finite cover. A **base** (or basis) for the topology of a topological space X is a collection of open $\mathfrak{B}$ sets such that all open sets from the topology $\mathfrak{T}$ are unions of open sets from the family $\mathfrak{B}$. A topological space is called **second countable** if its topology has a countable base.

Definition 3.1 *A **topological manifold** is a paracompact Hausdorff topological space M such that every point $p \in M$ is contained in some open set U_p that is the domain of a homeomorphism $\phi : U_p \to V$ onto an open subset of some Euclidean space $\mathbb{R}^n$.*

Thus we say that topological manifold M is "locally Euclidean". Many authors assume that a manifolds is second countable and then show that paracompactness follows. It seems to the author that paracompactness is the really important thing. In fact, it is surprising how far one can go without assuming second countability. Leaving paracompactness out of the definition has the advantage of making a foliation (defined later) just another manifold. Nevertheless we will follow tradition and assume second countability unless otherwise stated.

It might seem that the n in the definition might change from point to point or might not even be a well defined function on M depending essentially on the homeomorphism chosen. However, on a connected space, this in fact not possible. It is a consequence of a result of Brower called "invariance of domain" that the "dimension" n must be a locally constant function and therefore constant on connected manifolds. This result is rather easy to prove if the manifold has a differentiable structure (defined below) but more difficult in general. We shall simply record Brower's theorem:

Theorem 3.1 (Invariance of Domain) *The image of an open set $U \subset \mathbb{R}^n$ by a 1-1 continuous map $f : U \to \mathbb{R}^n$ is open. It follows that if $U \subset \mathbb{R}^n$ is homeomorphic to $V \subset \mathbb{R}^m$ then $m = n$.*

Each connected component of a manifold M could have a different dimension but we will restrict our attention to so called "pure manifolds" for which each component has the same dimension which we may then just refer to as the **dimension** of M. The latter is denoted $\dim(M)$. A **topological manifold with boundary** is a second countable Hausdorff topological space M such that point $p \in M$ is contained in some open set U_p that is the domain of a homeomorphism $\psi : U \to V$ onto an open subset V of some Euclidean half space $\mathbb{R}^n_- =: \{\vec{x} : x^1 \leq 0\}$[1]. A point that is mapped to the hypersurface $\partial\mathbb{R}^n_- = \mathbb{R}^n_0 =: \{\vec{x} : x^1 = 0\}$ under one of these homeomorphism is called a boundary point. As a corollary to Brower's theorem, this concept is independent of the homeomorphism used. The set of all boundary points of M is called the boundary of M and denoted ∂M. The interior is $\mathrm{int}(M) := M - \partial M$.

Topological manifolds, as we have defined them, are paracompact and also "normal" which means that given any pair of disjoint closed sets $F_1, F_2 \subset M$ there are open sets U_1 and U_2 containing F_1 and F_2 respectively such that U_1 and U_2 are also disjoint. The property of being paracompact may be combined with normality to show that topological manifolds support C^0-partitions of unity: Given any cover of M by open sets $\{U_\alpha\}$ there is a family of continuous functions $\{\beta_i\}$ called a C^0-partition of unity whose domains form a cover of M such that

(i) $\mathrm{supp}(\beta_i) \subset U_\alpha$ for some α,

(ii) each $p \in M$ has a neighborhood that intersects the support of only a finite number of the β_i.

(iii) we have $\sum \beta_i = 1$. (notice that the sum $\sum \beta_i(x)$ is finite for each $p \in M$ by (ii)).

Remark 3.1 *For differentiable manifolds we will be much more interested in the existence of $C^\infty-$partitions of unity. Finite dimensional differentiable manifolds always support smooth partitions of unity. This has been called smooth paracompactness.*

3.3 Differentiable Manifolds and Differentiable Maps

Definition 3.2 *Let M be a topological manifold. A pair $(U, \mathtt{x})$, where U is an open subset of M and $\mathtt{x} : U \to \mathbb{R}^n$ is a homeomorphism, is called a* ***chart*** *or* ***coordinate system*** *on M.*

If $(U, \mathtt{x})$ is a chart (with range in $\mathbb{R}^n$) then $\mathtt{x} = (x^1, ..., x^n)$ for some functions x^i $(i = 1, ..., n)$ defined on U which are called coordinate functions. To be

[1] Using $\mathbb{R}^n_+ =: \{x : x^1 \geq 0\}$ is equivalent at this point in the development and is actually the more popular choice. Later on when we define orientation on a (smooth) manifold this "negative" half space will be more convenient since we will be faced with less fussing over minus signs.

precise, we are saying that if $\mathrm{pr}_i : \mathbb{R}^n \to \mathbb{R}$ is the obvious projection onto the i-th factor of $\mathbb{R}^n := \mathbb{R} \times \cdots \times \mathbb{R}$ then $x^i := \mathrm{pr}_i \circ \mathtt{x}$ and so for $p \in M$ we have $\mathtt{x}(p) = (x^1(p), ..., x^n(p)) \in R^n$.

By the very definition of topological manifold we know that we may find a family of charts $\{(\mathtt{x}_\alpha, U_\alpha)\}_{\alpha \in A}$ whose domains cover M; that is $M = \cup_{\alpha \in A} U_\alpha$. Such a cover by charts is called an atlas for M. It is through the notion of change of coordinate maps (also called transition maps or overlap maps etc.) that we define the notion of a differentiable structure on a manifold.

Definition 3.3 *Let $\mathcal{A} = \{(\mathtt{x}_\alpha, U_\alpha)\}_{\alpha \in A}$ be an atlas on an n-dimensional topological manifold. Whenever the overlap $U_\alpha \cap U_\beta$ between two chart domains is nonempty we have the* ***change of coordinates*** *map $\mathtt{x}_\beta \circ \mathtt{x}_\alpha^{-1} : \mathtt{x}_\alpha(U_\alpha \cap U_\beta) \to \mathtt{x}_\beta(U_\alpha \cap U_\beta)$. If all such change of coordinates maps are C^r-diffeomorphisms then we call the atlas a* ***C^r-atlas***.

Now we might have any number of **C^r-atlases** on a topological manifold but we must have some notion of compatibility. A chart $(U, \mathtt{x})$ is **compatible** with some atlas $\mathcal{A} = \{(\mathtt{x}_\alpha, U_\alpha)\}_{\alpha \in A}$ on M if the maps $\mathtt{x} \circ \mathtt{x}_\alpha^{-1} : \mathtt{x}_\alpha(U_\alpha \cap U) \to \mathtt{x}(U_\alpha \cap U)$ are C^r-diffeomorphisms whenever defined. More generally, two **C^r-atlases** $\mathcal{A} = \{(\mathtt{x}_\alpha, U_\alpha)\}_{\alpha \in A}$ and $\mathcal{A}' = \{\mathtt{x}_{\alpha'}, U_{\alpha'}\}_{\alpha' \in A'}$ are said to be compatible if the union $\mathcal{A} \cup \mathcal{A}'$ is a C^r-atlas. It should be pretty clear that given a C^r-atlases on M there is a unique maximal atlas that contains $\mathcal{A}$ and is compatible with it. Now we are about to say that a C^r-atlas on a topological manifold elevates it to the status C^r-differentiable manifold by giving the manifold a so-called C^r-structure (smooth structure) but there is a slight problem with stating it in that way. First, if two different atlases are compatible then we don't really want to consider them to be giving different C^r-structures. To avoid this problem we will just use our observation about maximal atlases. The definition is as follows:

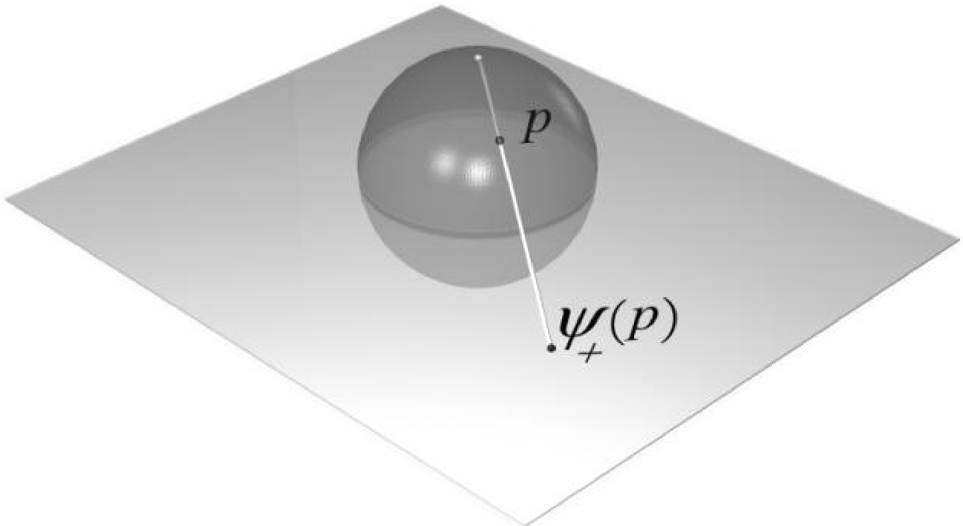
Definition 3.4 *A maximal C^r-atlas for a manifold M is called a C^r-differentiable structure. The manifold M together with this structure is called a C^r-differentiable manifold.*

Note well that any atlas determines a unique C^r-differentiable structure on M since it determine the unique maximal atlas that contains it. So in practice we just have to cover a space with mutually C^r-compatible charts in order to turn it into (or show that it has the structure of) a C^r-differentiable manifold. In other words, for a given manifold, we just need some atlas which we may enlarge with compatible charts as needed. For example, the space $\mathbb{R}^n$ is itself a C^∞ manifold (and hence a C^r-manifold for any $r \geq 0$) since we can take for an atlas for $\mathbb{R}^n$ the single chart $(\mathrm{id}, \mathbb{R}^n)$ where $\mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n$ is just the identity map $\mathrm{id}(x) = x$. Other atlases may be used in a given case and with experience it becomes more or less obvious which of the common atlases are mutually compatible and so the technical idea of a maximal atlas usually fades into the background. For example, once we have the atlas $\{(\mathrm{id}, \mathbb{R}^2)\}$ on the plane (consisting of the single chart) we have determined a differentiable structure on the plane. But then the chart given by polar coordinates is compatible with latter atlas and so we could throw this chart into the atlas and "fatten it up" a bit. Obviously, there are many more charts that could be thrown into the mix if needed because in this case any local diffeomorphism $U \subset \mathbb{R}^2 \to \mathbb{R}^2$ would be compatible with the "identity" chart $(\mathrm{id}, \mathbb{R}^2)$ and so would also be a chart within the same differentiable structure on $\mathbb{R}^2$. By the way, it is certainly possible for there to be two different differentiable structures on the same topological manifold. For example the chart given by the cubing function $(x \mapsto x^3, \mathbb{R}^1)$ is not compatible with the identity chart $(\mathrm{id}, \mathbb{R}^1)$ but since the cubing function also has domain all of $\mathbb{R}^1$ it too provides an atlas. But then this atlas cannot be compatible with the usual atlas $\{(\mathrm{id}, \mathbb{R}^1)\}$ and so they determine different maximal atlases. The problem is that the inverse of $x \mapsto x^3$ is not differentiable (in the usual sense) at the origin. Now we have two different differentiable structures on the line $\mathbb{R}^1$. Actually, the two atlases are equivalent in another sense that we will make precise below (they are diffeomorphic). We say that the two differentiable structures are different but equivalent or diffeomorphic. On the other hand, it is a deep result proved fairly recently that there exist infinitely many truly different non-diffeomorphic differentiable structures on $\mathbb{R}^4$. The reader ought to be wondering what is so special about dimension four. The next example generalizes these observations a bit:

Example 3.1 *For each positive integer n, the space $\mathbb{R}^n$ is a differentiable manifold in a simple way. Namely, there is a single chart that forms an atlas*[2] *which consists of the identity map $\mathbb{R}^n \to \mathbb{R}^n$. The resulting differentiable structure is called the standard differentiable structure on $\mathbb{R}^n$. Notice however that the map $\varepsilon : (x^1, x^2, ..., x^n) \mapsto ((x^1)^{1/3}, x^2, ..., x^n)$ is also a chart but not compatible with standard structure. Thus we seem to have two different differentiable structures and hence two different differentiable manifolds $\mathbb{R}^n, \mathcal{A}_1$ and $\mathbb{R}^n, \mathcal{A}_2$. This is true but they are equivalent in another sense. Namely, they are diffeomorphic via the map ε. See definition 3.13 below. Actually, if V is any vector space with a basis $(f_1, ..., f_n)$ and dual basis $(f_1^*, ..., f_n^*)$ then once again, we have an atlas consisting of just one chart defined on all of V which is the map $\mathrm{x} : \mathrm{v} \mapsto (f_1^* \mathrm{v}, ..., f_n^* \mathrm{v}) \in \mathbb{R}^n$. On the other hand V may as well be modelled (in a sense to be defined below) on itself using the identity map as the sole member of an atlas! The choice is a matter of convenience and taste.*

Example 3.2 *The sphere $S^2 \subset \mathbb{R}^3$. Choose a pair of antipodal points such as north and south poles where $z = 1$ and -1 respectively. Then off of these two pole points and off of a single half great circle connecting the poles we have the usual spherical coordinates. We actually have many such systems of spherical coordinates since we can re-choose the poles in many different ways. We can also use projection onto the coordinate planes as charts. For instance let U_z^+ be all $(x, y, z) \in S^2$ such that $z > 0$. Then $(x, y, z) \mapsto (x, y)$ provides a chart $U_z^+ \to \mathbb{R}^2$. The various transition functions can be computed explicitly and are clearly smooth. We can also use* ***stereographic projection*** *to give charts.*

[2]Of course there are many other compatible charts so this doesn't form a maximal atlas by a long shot.



More generally, we can provide the n−sphere $S^n \subset \mathbb{R}^{n+1}$ with a differentiable structure using two charts (U^+, ψ^+) and (U^-, ψ^-). Here,

$$U^{\pm} = \{\vec{x} = (x_1, ..., x_{n+1}) \in S^n : x_{n+1} \neq \pm 1\}$$

and ψ^+ (resp. ψ_-) is stereographic projection from the north pole $(0, 0....0, 1)$ (resp. south pole $(0, 0, ..., 0, -1)$). Explicitly we have

$$\psi^+(\vec{x}) = \frac{1}{(1 - x_{n+1})}(x_1, ..., x_n) \in \mathbb{R}^n$$
$$\psi_-(\vec{x}) = \frac{1}{(1 + x_{n+1})}(x_1, ..., x_n) \in \mathbb{R}^n$$

Exercise 3.1 *Compute $\psi_+ \circ \psi_-^{-1}$ and $\psi_-^{-1} \circ \psi_+$.*

Example 3.3 *The set of all lines through the origin in $\mathbb{R}^3$ is denoted $P_2(\mathbb{R})$ and is called the real projective plane . Let U_z be the set of all lines $\ell \in P_2(\mathbb{R})$ not contained in the x, y plane. Every line $\ell \in U_z$ intersects the plane $z = 1$ at exactly one point of the form $(x(\ell), y(\ell), 1)$. We can define a bijection $\psi_z : U_z \to \mathbb{R}^2$ by letting $\ell \mapsto (x(\ell), y(\ell))$. This is a chart for $P_2(\mathbb{R})$ and there are obviously two other analogous charts (ψ_x, U_x) and (ψ_y, U_y). These charts cover*

*$P_2(\mathbb{R})$ and so we have an atlas. More generally, the set of all lines through the origin in $\mathbb{R}^{n+1}$ is called **projective** $n-$**space** denoted $P_n(\mathbb{R})$ and can be given an atlas consisting of charts of the form (ψ_i, U_i) where*

$$U_i = \{\ell \in P_n(\mathbb{R}) : \ell \text{ is not contained in the hyperplane } x^i = 0$$
$$\psi_i(\ell) = \text{the unique coordinates } (u^1, ..., u^n) \text{ such that } (u^1, ..., 1, ..., u^n) \text{ is on the line } \ell.$$

Example 3.4 *If U is some open subset of a differentiable manifold M with atlas $\mathcal{A}_M$, then U is itself a differentiable manifold with an atlas of charts being given by all the restrictions $(\,\mathtt{x}_\alpha|_{U_\alpha \cap U}\,, U_\alpha \cap U\,)$ where $(\mathtt{x}_\alpha, U_\alpha) \in \mathcal{A}_M$. We shall refer to such an open subset $U \subset M$ with this differentiable structure as an **open submanifold** of M.*

Example 3.5 *The graph of a smooth function $f : \mathbb{R}^n \to \mathbb{R}$ is the subset of the Cartesian product $\mathbb{R}^n \times \mathbb{R}$ given by $\Gamma_f = \{(x, f(x)) : x \in \mathbb{R}^n\}$. The projection map $\Gamma_f \to \mathbb{R}^n$ is a homeomorphism and provides a global chart on Γ_f making it a smooth manifold. More generally, let $S \subset \mathbb{R}^{n+1}$ be a subset that has the property that for all $x \in S$ there is an open neighborhood $U \subset \mathbb{R}^{n+1}$ and some function $f :: \mathbb{R}^n \to \mathbb{R}$ such that $U \cap S$ consists exactly of the points of in U of the form*

$$(x^1, .., x^{j-1}, f(x^1, ..., \widehat{x^j}, .., x^{n+1}), x^{j+1}, ..., x^n).$$

Then on $U \cap S$ the projection

$$(x^1, .., x^{j-1}, f(x^1, ..., \widehat{x^j}, .., x^{n+1}), x^{j+1}, ..., x^n) \mapsto (x^1, .., x^{j-1}, x^{j+1}, ..., x^n)$$

is a chart for S. In this way, S is a differentiable manifold. Notice that S is a subset of the manifold $\mathbb{R}^{n+1}$ and the topology is the relative topology of S in $\mathbb{R}^{n+1}$.

Example 3.6 *The set of all $m \times n$ matrices $\mathbb{M}_{m \times n}$ (also written $\mathbb{R}^m_n$) is an $mn-$manifold modelled on $\mathbb{R}^{mn}$. We only need one chart again since it is clear that $\mathbb{M}_{m\times n}$ is in natural one to one correspondence with $\mathbb{R}^{mn}$ by the map $[a_{ij}] \mapsto (a_{11}, a_{12}, ..., a_{mn})$. Also, the set of all non-singular matrices $GL(n, \mathbb{R})$ is an open submanifold of $\mathbb{M}_{n\times n} \cong \mathbb{R}^{n^2}$.*

If we have two manifolds M_1 and M_2 of dimensions n_1 and n_2 respectively, we can form the topological Cartesian product $M_1 \times M_2$. We may give $M_1 \times M_2$ a differentiable structure in the following way: Let $\mathcal{A}_{M_1}$ and $\mathcal{A}_{M_2}$ be atlases for M_1 and M_2.Take as charts on $M_1 \times M_2$ the maps of the form

$$\mathtt{x}_\alpha \times \mathtt{y}_\gamma : U_\alpha \times V_\gamma \to \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$$

where $\mathtt{x}_\alpha, U_\alpha$ is a chart form $\mathcal{A}_{M_1}$ and $\mathtt{y}_\gamma, V_\gamma$ a chart from $\mathcal{A}_{M_2}$. This gives $M_1 \times M_2$ an atlas called the product atlas which induces a maximal atlas and hence a differentiable structure. With this product differentiable structure, $M_1 \times M_2$ is called a **product manifold**.

Example 3.7 *The circle is clearly a $C^\infty-$manifold and hence so is the product $T = S^1 \times S^1$ which is a torus.*

Example 3.8 *For any manifold M we can construct the "cylinder" $M \times I$ where I is some open interval in $\mathbb{R}$.*

Exercise 3.2 *$P(\mathbb{R}^2)$ is called the projective plane. Let V_z be the x, y plane and let V_x be the y, z plane. Let $U_{\mathrm{V}_x}, \mathtt{x}_x$ and $U_{\mathrm{V}_z}, \mathtt{x}_z$ be the corresponding coordinate charts. Find $\mathtt{x}_z \circ \mathtt{x}_x^{-1}$.*

3.4 Pseudo-Groups and Model Spaces*

Without much work we can generalize our definitions in such a way as to provide, as special cases, the definitions of some common notions such as that of *complex manifold* and *manifold with boundary.* If fact, we will also take this opportunity to include infinite dimensional manifolds. An infinite dimensional manifold is modelled on an infinite dimensional Banach space.. It is quite important for our purposes to realize that the spaces (so far just $\mathbb{R}^n$) that will be the model spaces on which we locally model our manifolds should have a distinguished family of local homeomorphisms. For example, C^r-differentiable manifolds are modelled on $\mathbb{R}^n$ where on the latter space we single out the local C^r-diffeomorphisms between open sets. But we will also study complex manifolds, foliated manifolds, manifolds with boundary, Hilbert manifolds and so on. Thus we need appropriate sets or spaces but also, significantly, we need a distinguished family of maps on the space. In this context the following notion becomes useful:

Definition 3.5 *A **pseudogroup of transformations**, say $\mathcal{G}$, of a topological space X is a family $\{\Phi_\gamma\}_{\gamma\in\Gamma}$ of homeomorphisms with domain U_γ and range V_γ, both open subsets of X, that satisfies the following properties:*

1) $\mathrm{id}_X \in \mathcal{G}$.

2) For all $\Phi_\gamma \in \mathcal{G}$ and open $U \subset U_\gamma$ the restrictions $\Phi_\gamma|_U$ are in $\mathcal{G}$.

3) $f_\gamma \in \mathcal{G}$ implies $f_\gamma^{-1} \in \mathcal{G}$.

4) The composition of elements of $\mathcal{G}$ are elements of $\mathcal{G}$ whenever the composition is defined with nonempty domain.

5) For any subfamily $\{\Phi_\gamma\}_{\gamma\in G_1} \subset \mathcal{G}$ such that $\Phi_\gamma|_{U_\gamma\cap U_\nu} = \Phi_\nu|_{U_\gamma\cap U_\nu}$ whenever $U_\gamma \cap U_\nu \neq \emptyset$ then the mapping defined by $\Phi : \bigcup_{\gamma\in G_1} U_\gamma \to \bigcup_{\gamma\in G_1} V_\gamma$ is an element of $\mathcal{G}$ if it is a homeomorphism.

Definition 3.6 *A **sub-pseudogroup** $\mathcal{S}$ of a pseudogroup is a subset of $\mathcal{G}$ that is also a pseudogroup (and so closed under composition and inverses).*

We will be mainly interested in what we shall refer to as C^r-pseudogroups and the spaces that support them. Our main example will be the set $\mathcal{G}^r_{\mathbb{R}^n}$ of all C^r maps between open subsets of $\mathbb{R}^n$. More generally, for a Banach space B we have the C^r-pseudogroup $\mathcal{G}^r_{\mathsf{B}}$ consisting of all C^r maps between open subsets of a Banach space B. Since this is our prototype the reader should spend some time thinking about this example.

Definition 3.7 *A C^r- pseudogroup of transformations of a subset M of Banach space B is a defined to be pseudogroup which results from restricting a sub-pseudogroup of $\mathcal{G}^r_{\mathsf{B}}$ to the subset M. The set M with the relative topology and this C^r- pseudogroup is called a* ***model space*** .

Example 3.9 *Recall that a map $U \subset \mathbb{C}^n \to \mathbb{C}^n$ is holomorphic if the derivative (from the point of view of the underlying real space $\mathbb{R}^{2n}$) is in fact complex linear. A holomorphic map with holomorphic inverse is called biholomorphic. The set of all biholomorphic maps between open subsets of $\mathbb{C}^n$ is a pseudogroup. This is a C^r-pseudogroup for all r including $r = \omega$. In this case the subset M referred to in the definition is just $\mathbb{C}^n$ itself.*

In the great majority of examples the subset $\mathsf{M} \subset \mathsf{V}$ is in fact equal to V itself. One important exception to this will lead us to a convenient formulation of manifold with boundary. First we need a definition:

Definition 3.8 *Let $\lambda \in \mathsf{M}^*$ be a continuous linear functional on a Banach space M. In the case of $\mathbb{R}^n$ it will be enough to consider projection onto the first coordinate x^1. Now let $\mathsf{M}^+_\lambda = \{x \in \mathsf{M}\colon \lambda(x) \geq 0\}$ and $\mathsf{M}^-_\lambda = \{x \in \mathsf{M}\colon \lambda(x) \leq 0\}$ and also let $\partial\mathsf{M}^+_\lambda = \partial\mathsf{M}^-_\lambda = \{x \in \mathsf{M}\colon \lambda(x) = 0\}$ (the kernel of λ). Clearly M^+_λ and M^-_λ are homeomorphic and $\partial\mathsf{M}^-_\lambda$ is a closed subspace.* [3] *The spaces M^+_λ and M^-_λ are referred to as* ***half spaces****.*

Example 3.10 *Let $\mathcal{G}^r_{\mathsf{M}^-_\lambda}$ be the restriction to M^-_λ of the set of C^r-diffeomorphisms ϕ from open subsets of M to open subsets of M that have the following property:*

**) If the domain U of $\phi \in \mathcal{G}^r_{\mathsf{M}}$ has nonempty intersection with $\mathsf{M}_0 := \{x \in \mathsf{M}\colon \lambda(x) = 0\}$ then $\phi|_{\mathsf{M}^-_\lambda \cap U}\left(\mathsf{M}^-_\lambda \cap U\right) \subset \mathsf{M}^-_\lambda \cap U$ and $\phi|_{\mathsf{M}_0 \cap U}\left(\mathsf{M}_0 \cap U\right) \subset \mathsf{M}_0 \cap U$.*

Notation 3.1 *It will be convenient to denote the model space for a manifold M (resp. N etc.) by M (resp. N etc.). That is, we use the same letter but use the sans serif font (this requires the reader to be tuned into font differences). There will be exceptions. One exception will be the case where we want to explicitly indicate that the manifold is finite dimensional and thus modelled on $\mathbb{R}^n$ for some n. Another exception will be when E is the total space of a vector bundle over M. In this case E will be modeled on a space of the form $\mathsf{M} \times \mathsf{E}$. This will be explained in detail when we study vector bundles.*

Let us now 'redefine' a few notions in greater generality. Let M be a topological space. An M**-chart** on M is a homeomorphism $\mathtt{x}$ whose domain is some subset $U \subset M$ and such that $\mathtt{x}(U)$ is an open subset of a fixed model space M.

[3] The reason we will use both M^+ and M^- in the following definition is a technical reason one having to do with the consistency of our definition of induced orientation of the boundary.

Definition 3.9 *Let $\mathcal{G}$ be a C^r-pseudogroup of transformations on a model space* M*. A $\mathcal{G}$-atlas for a topological space M is a family of charts $\{\mathtt{x}_\alpha, U_\alpha\}_{\alpha\in A}$ (where A is just an indexing set) that cover M in the sense that $M = \bigcup_{\alpha\in A} U_\alpha$ and such that whenever $U_\alpha \cap U_\beta$ is not empty then the map*

$$\mathtt{x}_\beta \circ \mathtt{x}_\alpha^{-1} : \mathtt{x}_\alpha(U_\alpha \cap U_\beta) \to \mathtt{x}_\beta(U_\alpha \cap U_\beta)$$

is a member of $\mathcal{G}$.

The maps $\mathtt{x}_\beta \circ \mathtt{x}_\alpha^{-1}$ are called various things by various authors including "transition maps", "coordinate change maps", and "overlap maps".

Now the way we set up the definitions, the model space M is a subset of a Banach space. If M is the whole Banach space (the most common situation) and if $\mathcal{G} = \mathcal{G}^r_{\mathsf{M}}$ (the whole pseudogroup of local C^rdiffeomorphisms) then, generalizing our former definition a bit, we call the atlas a C^r atlas.

Exercise 3.3 *Show that this definition of C^r atlas is the same as our original definition in the case where* M *is the finite dimensional Banach space $\mathbb{R}^n$.*

In practice, a $\mathcal{G}-$manifold is just a space M (soon to be a topological manifold) together with a $\mathcal{G}-$atlas $\mathcal{A}$ but as before we should tidy things up a bit for our formal definitions. First, let us say that a bijection onto an open set in a model space, say $\mathtt{x} : U \to \mathtt{x}(U) \subset \mathsf{M}$, is **compatible** with the atlas $\mathcal{A}$ if for every chart $\mathtt{x}_\alpha, U_\alpha$ from the atlas $\mathcal{A}$ we have that the composite map

$$\mathtt{x} \circ \mathtt{x}_\alpha^{-1} : \mathtt{x}_\alpha(U_\alpha \cap U) \to \mathtt{x}_\beta(U_\alpha \cap U)$$

is in $\mathcal{G}^r$. The point is that we can then add this map in to form a larger equivalent atlas: $\mathcal{A}' = \mathcal{A} \cup \{\mathtt{x}, U\}$. To make this precise let us say that two different C^r atlases, say $\mathcal{A}$ and $\mathcal{B}$ are equivalent if every map from the first is compatible (in the above sense) with the second and visa-versa. In this case $\mathcal{A}' = \mathcal{A} \cup \mathcal{B}$ is also an atlas. The resulting equivalence class is called a $\mathcal{G}^r-$**structure** on M.

As before, it is clear that every equivalence class of atlases contains a unique **maximal atlas** which is just the union of all the atlases in the equivalence class and every member of the equivalence class determines the maximal atlas –just toss in every possible compatible chart and we end up with the maximal atlas again. Thus the current definition of differentiable structure is equivalent to our previous definition.

Just as before, a topological manifold M is called a C^r-**differentiable manifold** (or just C^r manifold) if it comes equipped with a differentiable structure. Whenever we speak of a differentiable manifold we will have a fixed differentiable structure and therefore a maximal C^r-atlas $\mathcal{A}_M$ in mind. A chart from $\mathcal{A}_M$ will be called an **admissible chart**.

We started out with a topological manifold but if we had just started with a set M and then defined a chart to be a bijection $\mathtt{x} : U \to \mathtt{x}(U)$, only assuming $\mathtt{x}(U)$ to be open, then a maximal atlas $\mathcal{A}_M$ would generate a topology on M . Then the set U would be open. Of course we have to check that the result is a paracompact space but once that is thrown into our list of demands we have ended with the same notion of differentiable manifold. To see how this approach would go the reader should consult the excellent book [A,B,R].

Now from the vantage point of this general notion of model space we get a slick definition of manifold with boundary.

Definition 3.10 *A set M is called a C^r-**differentiable manifold with boundary** (or just C^r manifold with boundary) if it comes equipped with a $\mathcal{G}^r_{\mathsf{M}^-_\lambda}-$structure. M is given the topology induced by the maximal atlas for the given $\mathcal{G}^r_{\mathsf{M}^-_\lambda}$ structure.*

Whenever we speak of a C^r-manifold with boundary we will have a fixed $\mathcal{G}^r_{\mathsf{M}^-_\lambda}$ structure and therefore a maximal $\mathcal{G}^r_{\mathsf{M}^-_\lambda}$ atlas $\mathcal{A}_M$ in mind. A chart from $\mathcal{A}_M$ will be called an **admissible chart**.

Notice that the model spaces used in the definition of the charts were assumed to be a fixed space from chart to chart. We might have allowed for different model spaces but for topological reasons the model spaces must have constant dimension ($\leq \infty$) over charts with connected domain in a given connected component of M. In this more general setting if all charts of the manifold have range in a fixed M (as we usually assume) then the manifold is said to be a **pure manifold** and is said to be **modelled on** M.

Remark 3.2 *In the case of $\mathsf{M} = \mathbb{R}^n$ the chart maps $\mathtt{x}$ are maps into $\mathbb{R}^n$ and so projecting to each factor we have that $\mathtt{x}$ is comprised of n functions x^i and we write $\mathtt{x} = (x^1, ..., x^n)$. Because of this we sometimes talk about "x-coordinates versus y-coordinates" and so forth. Also, we use several rather self explanatory expressions such as "**coordinates**", "**coordinate charts**", "**coordinate systems**" and so on and these are all used to refer roughly to same thing as "chart" as we have defined the term. A chart $\mathtt{x}, U$ on M is said to be **centered at** p if $\mathtt{x}(p) = 0 \in \mathsf{M}$.*

Example 3.11 *Each Banach space M is a differentiable manifold in a trivial way. Namely, there is a single chart that forms an atlas[4] which is just the identity map $\mathsf{M} \to \mathsf{M}$. In particular $\mathbb{R}^n$ with the usual coordinates is a smooth manifold.*

If V is any vector space with a basis $(f_1, ..., f_n)$ and dual basis $(f_1^*, ..., f_n^*)$ then once again, we have an atlas consisting of just one chart define on all of V defined by $\mathtt{x} : \mathrm{v} \mapsto (f_1^*\mathrm{v}, ..., f_n^*\mathrm{v}) \in \mathbb{R}^n$. On the other hand V may as well be modelled on itself using the identity map! The choice is a matter of convenience and taste.

If we have two manifolds M_1 and M_2 we can form the topological Cartesian product $M_1 \times M_2$. The definition of a product proceeds just as in the finite

[4] Of course there are many other compatible charts so this doesn't form a maximal atlas by a long shot.

dimensional case: Let $\mathcal{A}_{M_1}$ and $\mathcal{A}_{M_2}$ be atlases for M_1 and M_2. Take as charts on $M_1 \times M_2$ the maps of the form

$$\mathtt{x}_\alpha \times \mathtt{y}_\gamma : U_\alpha \times V_\gamma \to \mathsf{M}_1 \times \mathsf{M}_2$$

where $\mathtt{x}_\alpha, U_\alpha$ is a chart form $\mathcal{A}_{M_1}$ and $\mathtt{y}_\gamma, V_\gamma$ a chart from $\mathcal{A}_{M_2}$. This gives $M_1 \times M_2$ an atlas called the product atlas which induces a maximal atlas and hence a differentiable structure.

It should be clear from the context that M_1 and M_2 are modelled on M_1 and M_2 respectively. Having to spell out what is obvious from context in this way would be tiring to both the reader and the author. Therefore, let us forgo such explanations to a greater degree as we proceed and depend rather on the common sense of the reader.

3.5 Smooth Maps and Diffeomorphisms

Functions.

A function defined on a manifold or on some open subset is differentiable by definition if it appears differentiable in every coordinate system that intersects the domain of the function. The definition will be independent of which coordinate system we use because that is exactly what the mutual compatibility of the charts in an atlas guarantees. To be precise we have

Definition 3.11 *Let M be a C^r-manifold modeled on the Banach space M (usually $\mathsf{M} = \mathbb{R}^n$ for some n). Let $f : O \subset M \to \mathbb{R}$ be a function on M with open domain O. We say that f is C^r-differentiable if and only if for every admissible chart $U, \mathtt{x}$ with $U \cap O \neq \emptyset$ the function*

$$f \circ \mathtt{x}^{-1} : \mathtt{x}(U \cap O) \to \mathsf{M}$$

is C^r-differentiable.

The reason that this definition works is because if $U, \mathtt{x}$, $\acute{U}, \acute{\mathtt{x}}$ are any two charts with domains intersecting O then we have

$$f \circ \mathtt{x}^{-1} = (f \circ \acute{\mathtt{x}}^{-1}) \circ (\acute{\mathtt{x}} \circ \mathtt{x}^{-1})$$

whenever both sides are defined and since $\acute{\mathtt{x}} \circ \mathtt{x}^{-1}$ is a C^r-diffeomorphism, we see that $f \circ \mathtt{x}^{-1}$ is C^r if and only if $f \circ \acute{\mathtt{x}}^{-1}$ is C^r. The chain rule is at work here of course.

Remark 3.3 *We have seen that when we compose various maps as above the domain of the result will in general be an open set that is the largest open set so that the composition makes sense. If we do not wish to write out explicitly what the domain is then will just refer to the* ***natural domain*** *of the composite map.*

Definition 3.12 *Let M and N be C^r manifolds with corresponding maximal atlases $\mathcal{A}_M$ and $\mathcal{A}_N$ and modelled on* M *and* F *respectively. A map $f : M \to N$ is said to be k* ***times continuously differentiable*** *or C^r if for every choice of charts $\mathtt{x}, U$ from $\mathcal{A}_M$ and $\mathtt{y}, V$ from $\mathcal{A}_N$ the composite map*

$$\mathtt{y} \circ f \circ \mathtt{x}^{-1} :: \mathsf{M} \to \mathsf{F}$$

is C^r on its natural domain (see convention 1.2.2). The set of all C^r maps $M \to N$ is denoted $C^r(M, N)$ or sometimes $C^r(M \to N)$.

Exercise 3.4 *Explain why this is a well defined notion. Hint: Think about the chart overlap maps.*

Sometimes we may wish to speak of a map being C^r at a point and for that we have a modified version of the last definition: Let M and N be C^r manifolds with corresponding maximal atlases $\mathcal{A}_M$ and $\mathcal{A}_N$ and modeled on M and F respectively. A (pointed) map $f : (M, p) \to (N, q)$ is said to be r **times continuously differentiable** or C^r at p if for every choice of charts $\mathtt{x}, U$ from $\mathcal{A}_M$ and $\mathtt{y}, V$ from $\mathcal{A}_N$ containing p and $q = f(p)$ respectively, the composite map

$$\mathtt{y} \circ f \circ \mathtt{x}^{-1} :: (\mathsf{M}, \mathtt{x}(p)) \to (\mathsf{F}, \mathtt{y}(q))$$

is C^r on some open set containing $\psi(p)$.

Just as for maps between open sets of Banach spaces we have

Definition 3.13 *A bijective map $f : M \to N$ such that f and f^{-1} are C^r with $r \geq 1$ is called a C^r-**diffeomorphism**. In case $r = \infty$ we shorten C^∞-diffeomorphism to just **diffeomorphism**. The group of all C^r diffeomorphisms of a manifold M onto itself is denoted $\mathrm{Diff}^r(M)$. In case $r = \infty$ we simply write $\mathrm{Diff}(M)$.*

We will use the convention that $\mathrm{Diff}^0(M)$ denotes the group of homeomorphisms of M onto itself.

Example 3.12 *The map $r_\theta : S^2 \to S^2$ given by $r_\theta(x, y, z) = (x \cos\theta - y \sin\theta, x \sin\theta + y \cos\theta, z)$ for $x^2 + y^2 + z^2 = 1$ is a diffeomorphism (and also an isometry).*

Example 3.13 *The map $f : S^2 \to S^2$ given by $f(x, y, z) = (x \cos((1 - z^2)\theta) - y \sin((1 - z^2)\theta), x \sin((1 - z^2)\theta) + y \cos((1 - z^2)\theta), z)$ is also a diffeomorphism (but not an isometry). Try to picture this map.*

Definition 3.14 *C^r differentiable manifolds M and N will be called C^r diffeomorphic and then said to be in the same C^r diffeomorphism class if and only if there is a C^r diffeomorphism $f : M \to N$.*

Recall the we have pointed out that we can put more than one differentiable structure on $\mathbb{R}$ by using the function $x^{1/3}$ as a chart. This generalizes in the obvious way: The map $\varepsilon : (x^1, x^2, ..., x^n) \mapsto ((x^1)^{1/3}, x^2, ..., x^n)$ is a chart for

$\mathbb{R}^n$ but not compatible with the standard (identity) chart. It induces the usual topology again but the resulting maximal atlas is different! Thus we seem to have two manifolds $\mathbb{R}^n, \mathcal{A}_1$ and $\mathbb{R}^n, \mathcal{A}_2$. This is true. They are different. But they are equivalent in another sense. Namely, they are diffeomorphic via the map ε. So it may be that the same underlying topological space M carries two different differentiable structures and so we really have two differentiable manifolds. Nevertheless it may still be the case that they are diffeomorphic. The more interesting question is whether a topological manifold can carry differentiable structures that are not diffeomorphic. It turns out that $\mathbb{R}^4$ carries infinitely many pairwise non-diffeomorphic structures (a very deep and difficult result) but $\mathbb{R}^k$ for $k \geq 5$ has only one diffeomorphism class.

Definition 3.15 *A map $f : M \to N$ is called a local diffeomorphism if and only if every point $p \in M$ is in an open subset $U_p \subset M$ such that $f|_{U_p} : U_p \to f(U)$ is a diffeomorphism.*

Example 3.14 *The map $\pi : S^2 \to P(\mathbb{R}^2)$ given by taking the point (x, y, z) to the line through this point and the origin is a local diffeomorphism but is not a diffeomorphism since it is 2-1 rather than 1-1.*

Example 3.15 *If we integrate the first order system of differential equations with initial conditions*

$$\begin{aligned} &y = x' \\ &y' = x \\ &x(0) = \xi \\ &y(0) = \theta \end{aligned}$$

we get solutions

$$\begin{aligned} x\,(t;\xi,\theta) &= \left(\tfrac{1}{2}\theta + \tfrac{1}{2}\xi\right)e^{t} - \left(\tfrac{1}{2}\theta - \tfrac{1}{2}\xi\right)e^{-t} \\ y\,(t;\xi,\theta) &= \left(\tfrac{1}{2}\theta + \tfrac{1}{2}\xi\right)e^{t} + \left(\tfrac{1}{2}\theta - \tfrac{1}{2}\xi\right)e^{-t} \end{aligned}$$

that depend on the initial conditions (ξ, θ). Now for any t the map $\Phi_t : (\xi, \theta) \mapsto (x(t, \xi, \theta), y(t, \xi, \theta))$ is a diffeomorphism $\mathbb{R}^2 \to \mathbb{R}^2$. The fact that we automatically get a diffeomorphism here follows from a moderately hard theorem proved later in the book.

Example 3.16 *The map $(x, y) \mapsto (\frac{1}{1-z(x,y)}x, \frac{1}{1-z(x,y)}y)$ where $z(x, y) = \sqrt{1 - x^2 - y^2}$ is a diffeomorphism from the open disk $B(0, 1) = \{(x, y) : x^2 + y^2 < 1\}$ onto the whole plane. Thus $B(0, 1)$ and $\mathbb{R}^2$ are diffeomorphic and in this sense the "same" differentiable manifold.*

We shall often need to consider maps that are defined on subsets $S \subset M$ that are not necessarily open. We shall call such a map f smooth (resp. C^r) if there is an open set O containing S and map $\widetilde{f}$ that is smooth (resp. C^r) on O and such that $\widetilde{f}\Big|_S = f$. In particular a curve defined on a closed interval $[a, b]$ is called **smooth** if it has a smooth extension to an open interval containing $[a, b]$. We will occasionally need the following simple concept:

Definition 3.16 *A continuous curve $c : [a, b] \to M$ into a smooth manifold is called* ***piecewise smooth*** *if there exists a partition $a = t_0 < t_1 < \cdots < t_k = b$ such that c restricted to $[t_i, t_{i+1}]$ is smooth for $0 \le i \le k-1$.*

3.6 Local expressions

Many authors seem to be over zealous and overly pedantic when it comes to the notation used for calculations in a coordinate chart. We will then make some simplifying conventions that are exactly what every student at the level of advanced calculus is already using anyway. Consider an arbitrary pair of charts $\mathtt{x}$ and $\mathtt{y}$ and the transition maps $\mathtt{y} \circ \mathtt{x}^{-1} : \mathtt{x}(U \cap V) \to \mathtt{y}(U \cap V)$. We write

$$\mathtt{y}(p) = \mathtt{y} \circ \mathtt{x}^{-1}(\mathtt{x}(p))$$

for $p \in U \cap V$. For finite dimensional manifolds we see this written as

$$y^i(p) = y^i(x^1(p), ..., x^n(p)) \tag{3.1}$$

which makes sense but in the literature we also see

$$y^i = y^i(x^1, ..., x^n). \tag{3.2}$$

In this last expression one might wonder if the x^i are functions or numbers. But this ambiguity is sort of purposeful for if 3.1 is true for all $p \in U \cap V$ then 3.2 is true for all $(x^1, ..., x^n) \in \mathtt{x}(U \cap V)$ and so we are unlikely to be led into error. This common and purposely ambiguous notational is harder to pull of in the case of infinite dimensional manifolds. We will instead write two different expressions in which the lettering and fonts are intended to be at least reminiscent of the classical notation:

$$\mathtt{y}(p) = \mathtt{y} \circ \mathtt{x}^{-1}(\mathtt{x}(p))$$

and

$$y = \mathtt{y} \circ \mathtt{x}^{-1}(x).$$

In the first case, $\mathtt{x}$ and $\mathtt{y}$ are functions on $U \cap V$ while in the second case, x and y are elements of $\mathtt{x}(U \cap V)$ and $\mathtt{y}(U \cap V)$ respectively[5]. In order not to interfere with our subsequent development let us anticipate the fact that this notational principle will be manifest later when we compare and make sense out of the following familiar looking expressions:

$$d\mathtt{y}(\xi) = \left.\frac{\partial \mathrm{y}}{\partial \mathrm{x}}\right|_{\mathtt{x}(p)} \circ d\mathtt{x}(\xi)$$

and

$$w = \left.\frac{\partial \mathrm{y}}{\partial \mathrm{x}}\right|_{\mathtt{x}(p)} v$$

[5]Notice the font differences.

which should be compared with the classical expressions

$$dy^i(\xi) = \frac{\partial y^i}{\partial x^k} dx^k(\xi)$$

and

$$w^i = \frac{\partial y^i}{\partial x^k} v^k.$$

3.7 Coverings and Discrete groups

3.7.1 Covering spaces and the fundamental group

In this section and later when we study fiber bundles many of the results are interesting and true in either the purely topological category or in the differentiable category. In order to not have to do things twice let us agree that a C^r-manifold is simply a paracompact Hausdorff topological space in case $r = 0$. Thus all relevant maps in this section are to be C^r where if $r = 0$ we just mean continuous and then only require that the spaces be sufficiently nice topological spaces (usually Hausdorff and paracompact). Also, "C^r diffeomorphism" just means homeomorphism in case $r = 0$.

"C^0-diffeomorphism"	=	C^0-isomorphism	=	homeomorphism
"C^0-manifold"	=	topological space		
C^0-group	=	topological group		

Much of what we do for C^0 maps works for very general topological spaces and so the word "manifold" could be replaced by topological space although the technical condition of being "semi-locally simply connected" (SLSC) is sometimes needed. All manifolds are SLSC.

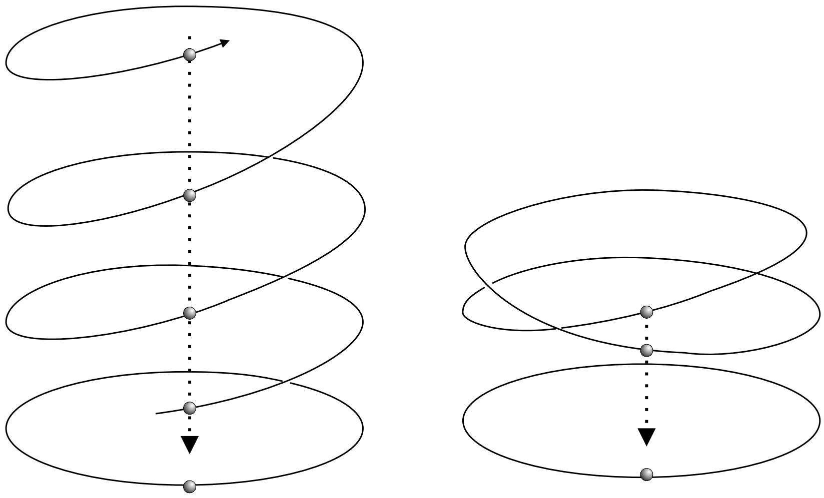
We may define a simple equivalence relation on a topological space by declaring

$$p \sim q \Leftrightarrow \text{ there is a continuous curve connecting } p \text{ to } q.$$

The equivalence classes are called path components and if there is only one such class then we say that M is **path connected**. The following exercise will be used whenever needed without explicit mention:

Exercise 3.5 *The path components of a manifold M are exactly the connected components of M. Thus, a manifold is connected if and only if it is path connected.*

Definition 3.17 *Let $\widetilde{M}$ and M be C^r-spaces. A surjective C^r map $\wp : \widetilde{M} \to M$ is called a C^r **covering map** if every point $p \in M$ has an open connected neighborhood U such that each connected component $\widetilde{U}_i$ of $\wp^{-1}(U)$ is C^r diffeomorphic to U via the restriction $\wp|_{\widetilde{U}_i} : \widetilde{U}_i \to U$. In this case we say that U is **evenly covered**. The triple $(\widetilde{M}, \wp, M)$ is called a covering space. We also refer to the space $\widetilde{M}$ (somewhat informally) as a **covering space** for M.*



We are mainly interested in the case where the spaces and maps are differentiable (C^r for $r > 1$).

Example 3.17 *The map* $\mathbb{R} \to S^1$ *given by* $t \mapsto e^{it}$ *is a covering. The set of points* $\{e^{it} : \theta - \pi < t < \theta + \pi\}$ *is an open set evenly covered by the intervals* I_n *in the real line given by* $I_n := (\theta - \pi + n2\pi, \theta + \pi + n2\pi)$.

Exercise 3.6 *Explain why the map* $(-2\pi, 2\pi) \to S^1$ *given by* $t \mapsto e^{it}$ *is* not *a covering map.*

The set of all C^r covering spaces are the objects of a category. A morphism between covering spaces, say $(\widetilde{M}_1, \wp_1, M_1)$ and $(\widetilde{M}_2, \wp_2, M_2)$ is a pair of maps $(\widetilde{f}, f)$ that give a commutative diagram

$$\begin{array}{ccc} \widetilde{M}_1 & \xrightarrow{\widetilde{f}} & \widetilde{M}_2 \\ \downarrow & & \downarrow \\ M_1 & \xrightarrow{f} & M_2 \end{array}$$

which means that $f \circ \wp_1 = \wp_2 \circ \widetilde{f}$. Similarly the set of coverings of a fixed space M are the objects of a category where the morphisms are maps $\Phi : \widetilde{M}_1 \to \widetilde{M}_2$ required to make the following diagram commute:

$$\begin{array}{ccc} \widetilde{M}_1 & \xrightarrow{\Phi} & \widetilde{M}_2 \\ \downarrow & & \downarrow \\ M & \xrightarrow{\mathrm{id}_M} & M \end{array}$$

so that $\wp_1 = \wp_2 \circ \Phi$. Now let $(\widetilde{M}, \wp, M)$ be a C^r covering space. The set of all C^r–diffeomorphisms Φ that are automorphisms in the above category; that is, diffeomorphisms for which $\wp = \wp \circ \Phi$, are called **deck transformations**. A deck transformation permutes the elements of each fiber $\wp^{-1}(p)$. In fact, it is not to hard to see that if $U \subset M$ is evenly covered then Φ permutes the connected components of $\wp^{-1}(U)$.

Proposition 3.1 *If $\wp : \widetilde{M} \to M$ is a C^r covering map with M being connected then the cardinality of $\wp^{-1}(p)$ is either infinite or is independent of p. In the latter case the cardinality of $\wp^{-1}(p)$ is the* ***multiplicity*** *of the covering.*

Proof. Fix $k \le \infty$. Let U_k be the set of all points such that $\wp^{-1}(p)$ has cardinality k. It is easy to show that U_k is both open and closed and so, since M is connected, U_k is either empty or all of M. ∎

Definition 3.18 *Let $\alpha : [0,1] \to M$ and $\beta : [0,1] \to M$ be two C^r maps (paths) both starting at $p \in M$ and both ending at q. A C^r* ***fixed end point homotopy*** *from α to β is a family of C^r maps $H_s : [0,1] \to M$ parameterized by $s \in [0,1]$ such that*

1) $H : [0,1] \times [0,1] \to M$ is continuous (or C^r) where $H(t,s) := H_s(t)$,
2) $H_0 = \alpha$ and $H_1 = \beta$,
3) $H_s(0) = p$ and $H_s(1) = q$ for all $s \in [0,1]$.

Definition 3.19 *If there is a C^r homotopy from α to β we say that α is C^r* ***homotopic*** *to β and write $\alpha \simeq \beta$ (C^r). If $r = 0$ we speak of paths being continuously homotopic.*

Remark 3.4 *It turns out that every continuous path on a C^r manifold may be uniformly approximated by a C^r path. Furthermore, if two C^r paths are continuously homotopic then they are C^r homotopic.*

It is easily checked that homotopy is an equivalence relation. Let $P(p,q)$ denote the set of all continuous paths from p to q defined on $[0,1]$. Every $\alpha \in P(p,q)$ has a unique inverse path $\alpha^{\leftarrow}$ defined by

$$\alpha^{\leftarrow}(t) := \alpha(1-t).$$

If p_1, p_2 and p_3 are three points in M then for $\alpha \in P(p_1,p_2)$ and $\beta \in P(p_2,p_3)$ we can "multiply" the paths to get a path $\alpha * \beta \in P(p_1,p_3)$ defined by

$$\alpha * \beta(t) := \begin{cases} \alpha(2t) & \text{for } 0 \le t < 1/2 \\ \beta(2t-1) & \text{for } 1/2 \le t < 1 \end{cases}.$$

An important observation is that if $\alpha_1 \simeq \alpha_2$ and $\beta_1 \simeq \beta_2$ then
$\alpha_1 * \beta_1 \simeq \alpha_2 * \beta_2$ where the homotopy between $\alpha_1 * \beta_1$ and $\alpha_2 * \beta_2$ is given in terms of the homotopy $H_\alpha : \alpha_1 \simeq \alpha_2$ and $H_\beta : \beta_1 \simeq \beta_2$ by

$$H(t,s) := \begin{cases} H_\alpha(2t,s) & \text{for } 0 \le t < 1/2 \\ H_\beta(2t-1,s) & \text{for } 1/2 \le t < 1 \end{cases} \quad \text{and } 0 \le s < 1$$

Similarly, if $\alpha_1 \simeq \alpha_2$ then $\alpha_1^{\leftarrow} \simeq \alpha_2^{\leftarrow}$. Using this information we can define a group structure on the set of homotopy equivalence classes of loops, that is, of paths in $P(p,p)$ for some fixed $p \in M$. First of all, we can always form $\alpha * \beta$ for any $\alpha, \beta \in P(p,p)$ since we are always starting and stopping at the same point p. Secondly we have the following

Proposition 3.2 *Let $\pi_1(M,p)$ denote the set of fixed end point (C^0) homotopy classes of paths from $P(p,p)$. For $[\alpha], [\beta] \in \pi_1(M,p)$ define $[\alpha] \cdot [\beta] := [\alpha * \beta]$. This is a well defined multiplication and with this multiplication $\pi_1(M,p)$ is a group. The identity element of the group is the homotopy class 1 of the constant map $1_p : t \to p$, the inverse of a class $[\alpha]$ is $[\alpha^{\leftarrow}]$.*

Proof. We have already shown that $[\alpha] \cdot [\beta] := [\alpha * \beta]$ is well defined. One must also show that

1) For any α, the paths $\alpha \circ \alpha^{\leftarrow}$ and $\alpha^{\leftarrow} \circ \alpha$ are both homotopic to the constant map 1_p.

2) For any $\alpha \in P(p,p)$ we have $1_p * \alpha \simeq \alpha$ and $\alpha * 1_p \simeq \alpha$.

3) For any $\alpha, \beta, \gamma \in P(p,p)$ we have $(\alpha * \beta) * \gamma \simeq \alpha * (\beta * \gamma)$.

The first two of these are straightforward and left as exercises. (JLfinish1)
■

The group $\pi_1(M,p)$ is called the **fundamental group** of M at p. If $\gamma : [0,1] \to M$ is a path from p to q then we have a group isomorphism $\pi_1(M,q) \to \pi_1(M,p)$ given by

$$[\alpha] \mapsto [\gamma * \alpha * \gamma^{\leftarrow}].$$

(One must check that this is well defined.) As a result we have

Proposition 3.3 *For any two points p, q in the same path component of M, the groups $\pi_1(M,p)$ and $\pi_1(M,q)$ are isomorphic (by the map described above).*

Corollary 3.1 *If M is connected then the fundamental groups based at different points are all isomorphic.*

Because of this last proposition, if M is connected we may refer to the fundamental group of M.

Definition 3.20 *A path connected topological space is called **simply connected** if $\pi_1(M) = \{1\}$.*

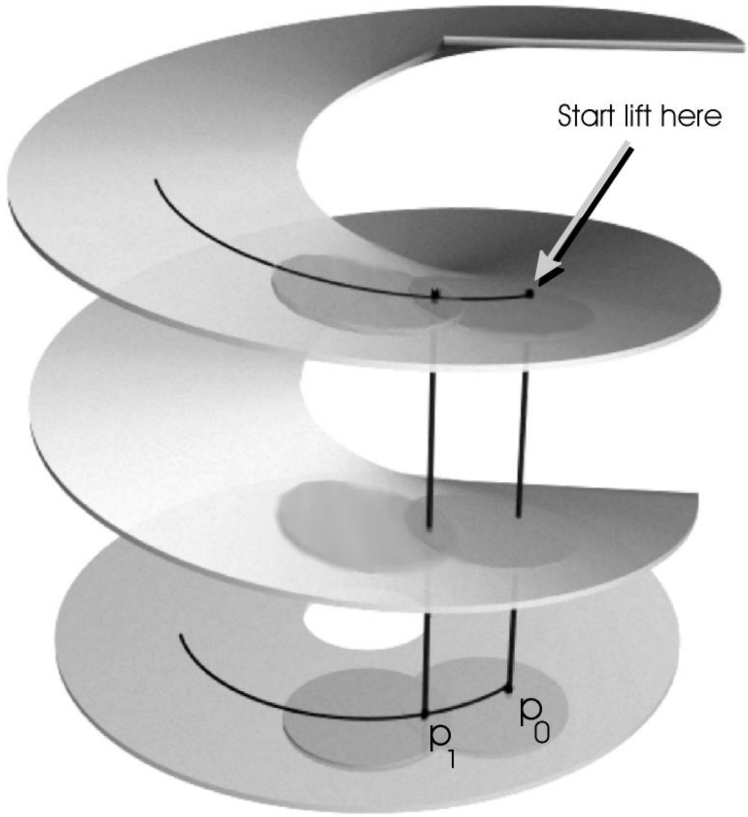
The fundamental group is actually the result of applying a functor (see B). To every pointed space (M,p) we assign the fundamental group $\pi_1(M,p)$ and to every base point preserving map (pointed map) $f : (M,p) \to (N, f(p))$ we may assign a group homomorphism $\pi_1(f) : \pi_1(M,p) \to \pi_1(N, f(p))$ by

$$\pi_1(f)([\alpha]) = [f \circ \alpha].$$

It is easy to check that this is a covariant functor and so for pointed maps f and g that can be composed $(M,x) \xrightarrow{f} (N,y) \xrightarrow{g} (P,z)$ we have $\pi_1(g \circ f) = \pi_1(g)\pi_1(f)$.

Notation 3.2 *To avoid notational clutter we will denote $\pi_1(f)$ by $f_\#$.*

Theorem 3.2 *Every connected manifold M has a simply connected covering space. Furthermore, if H is any subgroup of $\pi_1(M,p)$, then there is a connected covering $\wp : \widetilde{M} \to M$ and a point $\widetilde{p} \in \widetilde{M}$ such that $\wp_\#(\pi_1(\widetilde{M}, \widetilde{p})) = H$.*



Let $f : P \to M$ be a C^r map. A map $\widetilde{f} : P \to \widetilde{M}$ is said to be a **lift** of the map f if $\wp \circ \widetilde{f} = f$.

Theorem 3.3 *Let $\wp : \widetilde{M} \to M$ be a covering of C^r manifolds and $\gamma : [a,b] \to M$ a C^r curve and pick a point y in $\wp^{-1}(\gamma(a))$. Then there exists a unique C^r lift $\widetilde{\gamma} : [a,b] \to \widetilde{M}$ of γ such that $\widetilde{\gamma}(a) = y$. Thus the following diagram commutes.*

$$\begin{tikzcd} & \widetilde{M} \arrow[d, "\wp"] \\ {[a,b]} \arrow[ur, "\widetilde{\gamma}"] \arrow[r, "\gamma"] & M \end{tikzcd}$$

Similarly, if $h : [a,b] \times [c,d] \to M$ *is a* C^r-*map then it has a unique lift* $\widetilde{h} : [a,b] \times [c,d] \to \widetilde{M}$.

Proof. Figure 3.7.1 shows the way. Decompose the curve γ into segments that lie in evenly covered open sets. Lift inductively starting by using the inverse of $\wp$ in the first evenly covered open set. It is clear that in order to connect up continuously, each step is forced and so the lifted curve is unique. The proof of the second half is just slightly more complicated but the idea is the same and the proof is left to the curious reader. A tiny technicality in either case is the fact that for $r > 0$ a C^r-map on a closed set is defined to mean that there is a C^r-map on a slightly larger open set. For instance, for the curve γ we must lift an extension $\gamma_{ext} : (a - \varepsilon, b + \varepsilon) \to M$ but considering how the proof went we see that the procedure is basically the same and gives a C^r- extension $\widetilde{\gamma}_{ext}$ of the lift $\widetilde{\gamma}$. ∎

There are two important corollaries to this result. The first is that if $\alpha, \beta : [0,1] \to M$ are fixed end point homotopic paths in M and if $\widetilde{\alpha}$ and $\widetilde{\beta}$ are corresponding lifts with $\widetilde{\alpha}(0) = \widetilde{\beta}(0)$ then any homotopy $h_t : \alpha \simeq \beta$ lifts to a homotopy $\widetilde{h}_t : \widetilde{\alpha} \simeq \widetilde{\beta}$. The theorem then implies that $\widetilde{\alpha}(1) = \widetilde{\beta}(1)$. From this one easily proves the following

Corollary 3.2 *For every* $[\alpha] \in \pi_1(M,p)$ *there is a well defined map* $[\alpha]_\sharp : \wp^{-1}(p) \to \wp^{-1}(p)$ *given by letting* $[\alpha]_\sharp y$ *be* $\widetilde{\alpha}(1)$ *for the lift* $\widetilde{\alpha}$ *of* α *with* $\widetilde{\alpha}(0) = y$. *(Well defined means that any representative* $\alpha' \in [\alpha]$ *gives the same answer.)*

Now recall the notion of a deck transformation. Since we now have two ways to permute the elements of a fiber, on might wonder about their relationship. For instance, given a deck transformation Φ and a chosen fiber $\wp^{-1}(p)$ when do we have $\Phi|_{\wp^{-1}(p)} = [\alpha]_\sharp$ for some $[\alpha] \in \pi_1(M,p)$?

3.7.2 Discrete Group Actions

Let G be a group and endow G with the discrete topology so that, in particular, every point is an open set. In this case we call G a discrete group. If M is a topological space then so is $G \times M$ with the product topology. What does it mean for a map $\alpha : G \times M \to M$ to be continuous? The topology of $G \times M$ is clearly generated by sets of the form $S \times U$ where S is an arbitrary subset of G and U is open in M. The map $\alpha : G \times M \to M$ will be continuous if for any point $(g_0, x_0) \in G \times M$ and any open set $U \subset M$ containing $\alpha(g_0, x_0)$ we can find an open set $S \times V$ containing (g_0, x_0) such that $\alpha(S \times V) \subset U$. Since the topology of G is discrete, it is necessary and sufficient that there is an open V such that $\alpha(g_0 \times V) \subset U$. It is easy to see that a necessary and sufficient condition for α to be continuous on all of $G \times M$ is that the partial maps $\alpha_g(.) := \alpha(g, .)$ are continuous for every $g \in G$.

Recall the definition of a group action. In case G is discrete we have the special case of a discrete group action:

Definition 3.21 *Let G and M be as above. A (left) discrete group action is a map $\alpha : G \times M \to M$ such that for every $g \in G$ the partial map $\alpha_g(.) := \alpha(g, .)$ is continuous and such that the following hold:*
1) $\alpha(g_2, \alpha(g_1, x)) = \alpha(g_2 g_1, x)$ for all $g_1, g_2 \in G$ and all $x \in M$.
2) $\alpha(e, x) = x$ for all $x \in M$.

It follows that if $\alpha : G \times M \to M$ is a discrete action then each partial map $\alpha_g(.)$ is a homeomorphism with $\alpha_g^{-1}(.) = \alpha_{g^{-1}}(.)$. As before we write $g \cdot x$ or just gx in place of the more pedantic notation $\alpha(g, x)$. Using this notation we have $g_2(g_1 x) = (g_2 g_1)x$ and $ex = x$.

Definition 3.22 *A discrete group action is C^r if M is a C^r manifold and each $\alpha_g(.)$ is a C^r map.*

Let M/G be the set of orbits (cosets) and let $p : M \to M/G$ be the projection taking each x to Gx. If we give the space of orbits (cosets) M/G the quotient topology then p is forced to be continuous but more is true: The map $p : M \to M/G$ is an open map. To see this notice that if $U \subset M$ and we let $\widetilde{U} := p(U)$ then $p^{-1}(\widetilde{U})$ is open since

$$p^{-1}(\widetilde{U}) = \bigcup \{gU : g \in G\}$$

which is a union of open sets. Now since $p^{-1}(\widetilde{U})$ is open, $\widetilde{U}$ is open by definition of the quotient topology.

Example 3.18 *Let $\phi : M \to M$ be a diffeomorphism and let $\mathbb{Z}$ act on M by $n \cdot x := \phi^n(x)$ where*

$$\begin{aligned} \phi^0 &:= \mathrm{id}_M, \\ \phi^n &:= \phi \circ \cdots \circ \phi \text{ for } n > 0 \\ \phi^{-n} &:= (\phi^{-1})^n \text{ for } n > 0. \end{aligned}$$

This gives a discrete action of $\mathbb{Z}$ on M.

Definition 3.23 *A discrete group action $\alpha : G \times M \to M$ is said to act* ***properly*** *if for every $x \in M$ there is an open set $U \subset M$ containing x such that $gU \cap U = \emptyset$ for all $g \neq e$. We shall call such an open set* ***self avoiding****.*

It is easy to see that if $U \subset M$ is self avoiding then any open subset $V \subset U$ is also self avoiding. Thus every point $x \in M$ has a self avoiding neighborhood that is connected.

Proposition 3.4 *If $\alpha : G \times M \to M$ is a proper action and $U \subset M$ is self avoiding then p maps gU onto $p(U)$ for all $g \in G$ and the restrictions $p|_{gU} : gU \to p(U)$ are homeomorphisms. In fact, $p : M \to M/G$ is a covering map.*

Proof. Since $U \cong gU$ via $x \mapsto gx$ and since x and gx are in the same orbit, we see that gU and U both project to same set $p(U)$. Now if $x, y \in gU$ and $p|_{gU}(x) = p|_{gU}(y)$ then $y = hx$ for some $h \in G$. But also $x = ga$ (and $y = gb$) for some $a, b \in U$. Thus $h^{-1}gb = x$ so $x \in h^{-1}gU$. On the other hand we also know that $x \in gU$ so $h^{-1}gU \cap gU \neq \emptyset$ which implies that $g^{-1}h^{-1}gU \cap U \neq \emptyset$. Since U is self avoiding this means that $g^{-1}h^{-1}g = e$ and so $h = e$ from which we get $y = x$. Thus $p|_{gU} : gU \to p(U)$ is 1-1. Now since $p|_{gU}$ is clearly onto and since we also know that $p|_{gU}$ is an open continuous map the result follows. ∎

Example 3.19 *Fix a basis* $(\mathbf{f}_1, \mathbf{f}_2)$ *of* $\mathbb{R}^2$*. Let* $\mathbb{Z}^2$ *act on* $\mathbb{R}^2$ *by* $(m, n) \cdot (x, y) := (x, y) + m\mathbf{f}_1 + n\mathbf{f}_2$*. This action is easily seen to be proper.*

Example 3.20 *Recall the abelian group* $\mathbb{Z}_2$ *of two elements has both a multiplicative presentation and an additive presentation. In this example we take the multiplicative version. Let* $\mathbb{Z}_2 := \{1, -1\}$ *act on the sphere by* $(\pm 1) \cdot \vec{x} := \pm\vec{x}$*. Thus the action is generated by letting* -1 *send a point on the sphere to its antipode. This action is also easily seen to be proper.*

Exercise 3.7 (!) *Let* $\alpha : G \times M \to M$ *act properly on* M*. Show that if* U_1 *and* U_2 *are self avoiding and* $p(U_1) \cap p(U_2) \neq \emptyset$ *then there is a* $g \in G$ *such that* $\alpha_g(U_1) \cap U_2 \neq \emptyset$*. Show also that if* α_g *is* C^r *then* α_g *maps* $U_1 \cap \alpha_{g^{-1}}(U_2) := O_1$ *diffeomorphically onto* $\alpha_g(U_1) \cap U_2 := O_2$ *and in this case*

$$\alpha_g|_{O_1} = p|_{O_2}^{-1} \circ p|_{O_1}.$$

Hint: If U_1 *and* U_2 *are self avoiding then so are* O_1 *and* O_2 *and* $p(O_1) = p(O_2)$*.*

Proposition 3.5 *Let* $\alpha : G \times M \to M$ *act properly by* C^r*-diffeomorphisms on a* C^r*- manifold* M*. Then the quotient space* M/G *has a natural* C^r *structure such that the quotient map is a* C^r *local diffeomorphism. The quotient map is a covering map.*

Proof. We exhibit the atlas on M/G and then let the reader finish the (easy) proof. Let $\mathcal{A}_M$ be an atlas for M. Let $\bar{x} = Gx$ be a point in M/G and pick an open $U \subset M$ that contains a point, say x, in the orbit Gx that (as a set) is the preimage of $\bar{x}$ and such that unless $g = e$ we have $gU \cap U = \emptyset$. Now let $U_\alpha, \mathtt{x}_\alpha$ be a chart on M containing x. By replacing U and U_α by $U \cap U_\alpha$ we may assume that $\mathtt{x}_\alpha$ is defined on $U = U_\alpha$. In this situation, if we let $U^* := p(U)$ then each restriction $p|_U : U \to U^*$ is a homeomorphism. We define a chart map $\mathtt{x}_\alpha^*$ with domain U_α^* by

$$\mathtt{x}_\alpha^* := \mathtt{x}_\alpha \circ p|_{U_\alpha}^{-1} : U_\alpha^* \to \mathsf{M}.$$

Let $\mathtt{x}_\alpha^*$ and $\mathtt{x}_\beta^*$ be two such chart maps with domains U_α^* and U_β^*. If $U_\alpha^* \cap U_\beta^* \neq \emptyset$ then we have to show that $\mathtt{x}_\beta^* \circ (\mathtt{x}_\alpha^*)^{-1}$ is a C^r diffeomorphism. Let $\bar{x} \in U_\alpha^* \cap U_\beta^*$

and abbreviate $U^*_{\alpha\beta} = U^*_\alpha \cap U^*_\beta$. Since $U^*_\alpha \cap U^*_\beta \neq \emptyset$ there must be a $g \in G$ such that $\alpha_g(U_\alpha) \cap U_\beta \neq \emptyset$. Using exercise 3.7 and letting $\alpha_g(U_\alpha) \cap U_\beta := O_2$ and $O_1 := \alpha_{g^{-1}} O_2$ we have

$$\begin{aligned}
&\mathtt{x}^*_\beta \circ \mathtt{x}^*_\alpha|^{-1} \\
&= \mathtt{x}_\beta \circ p|^{-1}_{O_2} \circ (\mathtt{x}_\alpha \circ p|^{-1}_{O_1})^{-1} \\
&= \mathtt{x}_\beta \circ p|^{-1}_{O_2} \circ p|_{O_1} \circ \mathtt{x}^{-1}_\alpha \\
&= \mathtt{x}_\beta \circ \alpha_g \circ \mathtt{x}^{-1}_\alpha
\end{aligned}$$

which is C^r. The rest of the proof is straightforward and is left as an exercise.

Remark 3.5 *In the above, we have suppressed some of information about domains. For example, what is the domain and range of* $\mathtt{x}_\beta \circ p|^{-1}_{O_2} \circ p|_{O_1} \circ \mathtt{x}^{-1}_\alpha$ *? For completeness and in order to help the reader interpret the composition we write out the sequence of domains:*

$$\begin{aligned}
\mathtt{x}_\alpha(U_\alpha \cap \alpha_{g^{-1}} U_\beta) &\to U_\alpha \cap \alpha_{g^{-1}} U_\beta \\
&\to U^*_\alpha \cap U^*_\beta \to \alpha_g(U_\alpha) \cap U_\beta \to \mathtt{x}_\beta(\alpha_g(U_\alpha) \cap U_\beta).
\end{aligned}$$

■

Example 3.21 *We have seen the torus as a differentiable manifold previously presented as* $T^2 = S^1 \times S^1$. *Another presentation that emphasizes the symmetries is given as follows: Let the group* $\mathbb{Z} \times \mathbb{Z} = \mathbb{Z}^2$ *act on* $\mathbb{R}^2$ *by*

$$(m, n) \times (x, y) \mapsto (x + m, y + n).$$

It is easy to check that proposition 3.5 applies to give a manifold $\mathbb{R}^2/\mathbb{Z}^2$. *This is actually the torus in another guise and we have the diffeomorphism* $\phi : \mathbb{R}^2/\mathbb{Z}^2 \to S^1 \times S^1 = T^2$ *given by* $[(x, y)] \mapsto (e^{ix}, e^{iy})$. *The following diagram commutes:*

$$\begin{tikzcd}
\mathbb{R}^2 \arrow[r] \arrow[d] & S^1 \times S^1 \\
\mathbb{R}^2/\mathbb{Z}^2 \arrow[ur] &
\end{tikzcd}$$

Covering spaces $\wp : \widetilde{M} \to M$ that arise from a proper discrete group action are special in that if M is connected then the covering is a normal covering. (check this LPL)

Example 3.22 *Let* $\mathbb{Z}_2$ *act on* $S^{n-1} \subset \mathbb{R}^n$ *by* $1 \cdot p = p$ *and* $-1 \cdot p = -p$. *Show that this action is proper and show that* $S^{n-1}/\mathbb{Z}_2$ *is diffeomorphic to (or is just another realization of) projective space* $P(\mathbb{R}^n)$.

3.8 Grassmann Manifolds

A very useful generalization of the projective spaces is the Grassmann manifolds. Let $G_{n,k}$ denote the set of k-dimensional subspaces of $\mathbb{R}^n$. We will exhibit a natural differentiable structure on this set. The idea here is the following. An alternative way of defining the points of projective space is as equivalence classes of $n-$tuples $(v^1, ..., v^n) \in \mathbb{R}^n - \{0\}$ where $(v^1, ..., v^n) \sim (\lambda v^1, ..., \lambda v^n)$ for any nonzero. This is clearly just a way of specifying a line through the origin. Generalizing, we shall represent a $k-$plane as an $n \times k$ matrix whose row vectors span the $k-$ plane. Thus we are putting an equivalence relation on the set of $n \times k$ matrices where $A \sim Ag$ for any nonsingular $k \times k$ matrix g. Let $\mathbb{M}_{n \times k}$ of $n \times k$ matrices with rank $k < n$ (maximal rank). Two matrices from $\mathbb{M}_{n \times k}$ are equivalent exactly if their rows span the same k-dimensional subspace. Thus the set $G(k, n)$ of equivalence classes is in one to one correspondence with the set of real k dimensional subspaces of $\mathbb{R}^n$.

Now let U be the set of all $[A] \in G(k, n)$ such that A has its first k rows linearly independent. This property is independent of the representative A of the equivalence class $[A]$ and so U is a well defined set. This last fact is easily proven by a Gaussian column reduction argument. Now every element $[A] \in U \subset G(k, n)$ is an equivalence class that has a unique member A_0 of the form

$$\begin{pmatrix} I_{k \times k} \\ Z \end{pmatrix}.$$

Thus we have a map on U defined by $\Psi : [A] \mapsto Z \in \mathbb{M}_{(n-k) \times k} \cong \mathbb{R}^{k(n-k)}$. We wish to cover $G(k, n)$ with sets U_σ similar to U and define similar maps. Let $\sigma_{i_1 \ldots i_k}$ be the shuffle permutation that puts the k columns indexed by $i_1, ..., i_k$ into the positions $1, ..., k$ without changing the relative order of the remaining columns. Now consider the set $U_{i_1 \ldots i_k}$ of all $[A] \in G(k, n)$ such that any representative A is such that the k rows indexed by $i_1, ..., i_k$ are linearly independent. This characterization of $[A]$ is independent of the representative A. The the permutation induces an obvious 1-1 onto map $\widetilde{\sigma_{i_1 \ldots i_k}}$ from $U_{i_1 \ldots i_k}$ onto $U = U_{1 \ldots k}$. We now have maps $\Psi_{i_1 \ldots i_k} : U_{i_1 \ldots i_k} \to \mathbb{M}_{(n-k) \times k} \cong \mathbb{R}^{k(n-k)}$ given by composition $\Psi_{i_1 \ldots i_k} := \Psi \circ \widetilde{\sigma_{i_1 \ldots i_k}}$. These maps form an atlas $\{\Psi_{i_1 \ldots i_k}, U_{i_1 \ldots i_k}\}$ for $G(k, n)$ and gives it the structure of a smooth manifold called the **Grassmann manifold** of real k-planes in $\mathbb{R}^n$.

3.9 Submanifolds

Recall the simple situation from calculus where we have a C^∞ function $F(x, y)$ on the x, y plane. We know from the implicit function theorem that if $\frac{\partial F}{\partial y}(x_0, y_0) \neq 0$ then *near* (x_0, y_0) the level set $\{(x, y) : F(x, y) = F(x_0, y_0)\}$ is the graph of some function $y = g(x)$. The map $(x, y) \mapsto x$ is a homeomorphism onto an open subset of $\mathbb{R}$ and provides a chart. Hence, near this point, the level set is a 1-dimensional differentiable manifold. Now if either $\frac{\partial F}{\partial y}(x, y) \neq 0$ or $\frac{\partial F}{\partial x}(x, y) \neq 0$ at every point (x, y) on the level set, then we could cover the level set by these kind of charts (induced by coordinate projection) and since it turns out that overlap maps are C^∞ we would have a C^∞ differentiable 1-manifold. This idea generalizes nicely not only to higher dimensions but to manifolds in general and all we need is the local theory of differentiable maps as described by the inverse and implicit mapping theorems.

There is another description of a level set. Locally these are graphs of functions. But then we can also parameterize portions of the level sets by using this local graph structure. For example, in the simple situation just described we have the map $t \mapsto (t + x_0, g(t + y_0))$ which parameterizes a portion of the level set near (x_0, y_0). The inverse of this parameterization is just the chart.

First we define the notion of a submanifold and study some related generalities concerning maps. The reader should keep in mind the dual notions of level sets and parametrizations.

A subset S of a C^r-differentiable manifold M (modeled on M) is called a (**regular**) **submanifold** (of M) if there exists a Banach space decomposition of the model space $\mathsf{M} = \mathsf{M}_1 \times \mathsf{M}_2$ such that every point $p \in S$ is in the domain of an admissible chart $(\mathtt{x}, U)$ that has the following **submanifold property**:

$$\mathtt{x}(U \cap S) = \mathtt{x}(U) \cap (\mathsf{M}_1 \times \{0\})$$

Equivalently, we require that $\mathtt{x} : U \to V_1 \times V_2 \subset \mathsf{M}_1 \times \mathsf{M}_2$ is a diffeomorphism such that

$$\mathtt{x}(U \cap S) = V_1 \times \{0\}$$

for some open sets V_1, V_2. We will call such charts **adapted** (to S). The restrictions $\mathtt{x}|_{U \cap S}$ of adapted charts provide an atlas for S (called an induced submanifold atlas) making it a C^r differentiable manifold in its own right. If M_2 has finite dimension k then this k is called the **codimension** of S (in M) and we say that S is a submanifold of codimension k.

Exercise 3.8 *Show that S is a differentiable manifold and that a continuous map $f : N \to M$ that has its image contained in S is differentiable with respect to the submanifold atlas if and only if it is differentiable as a map in to M.*

When S is a submanifold of M then the tangent space T_pS at $p \in S \subset M$ is intuitively a subspace of T_pM. In fact, this is true as long as one is not bent on distinguishing a curve in S through p from the "same" curve thought of as a map into M. If one wants to be pedantic then we have the inclusion map $\iota : S \hookrightarrow M$ and if $c : I \to S$ is a curve into S then $\iota \circ c : I \to M$ is a map into M as such. At the tangent level this means that $c'(0) \in T_pS$ while $(\iota \circ c)'(0) \in T_pM$. Thus from this more pedantic point of view we have to explicitly declare $T_p\iota : T_pS \to T_p\iota(T_pS) \subset T_pM$ to be an identifying map. We will avoid the use of inclusion maps when possible and simply write $T_pS \subset T_pM$ and trust the intuitive notion that T_pS is indeed a subspace of T_pM.

3.10 Submanifolds of $\mathbb{R}^n$

If $M \subset \mathbb{R}^n$ is a regular k-dimensional submanifold then, by definition, for every $p \in M$ there is an open subset U of $\mathbb{R}^n$ containing p on which we have new coordinates $\mathtt{y} : U \to V \subset \mathbb{R}^n$ abbreviated by $(y_1, ..., y_n)$ such that $M \cap U$ is exactly given by $y_{k+1} = ... = y_n = 0$. On the other hand we have (U, id), i.e., the standard coordinates restricted to U, which we denote by $x_1, ..., x_n$. We see that $\mathrm{id} \circ \phi^{-1}$ is a diffeomorphism given by

$$\begin{gathered} x_1 = x_1(y_1, ..., y_n) \\ x_2 = x_2(y_1, ..., y_n) \\ \vdots \\ x_n = x_n(y_1, ..., y_n) \end{gathered}$$

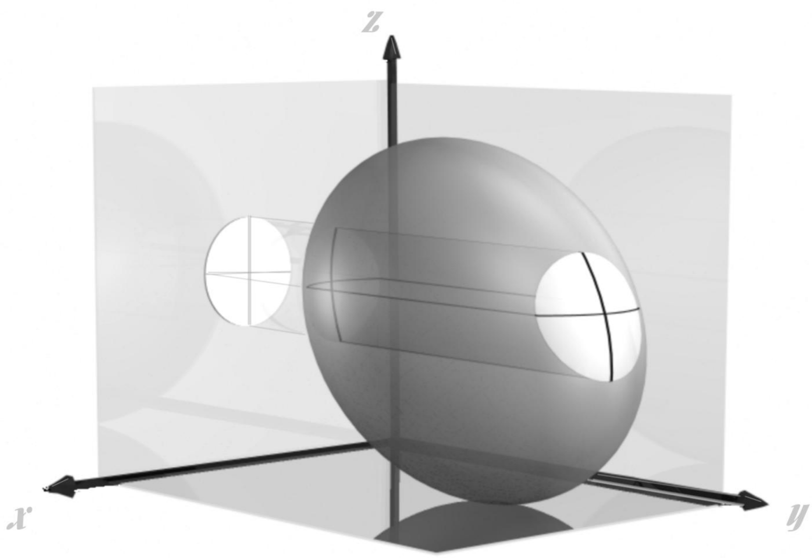
which in turn implies that the determinant $\det \frac{\partial(x_1,....,x_n)}{\partial(y_1,....,x_n)}$ must be nonzero throughout V. From a little linear algebra we conclude that for some renumbering of the coordinates the $x_1, ..., x_n$ the determinant $\det \frac{\partial(x_1,....,x_r)}{\partial(y_1,....,x_r)}$ must be nonzero at and therefore near $\phi(p) \in V \subset \mathbb{R}^n$. On this possibly smaller neighborhood V' we define a map F by

$$\begin{gathered} x_1 = x_1(y_1, ..., y_n) \\ x_2 = x_2(y_1, ..., y_n) \\ \vdots \\ x_r = x_r(y_1, ..., y_n) \\ x_{r+1} = y_{r+1} \\ \vdots \\ x_n = y_n \end{gathered}$$

then we have that F is a local diffeomorphism $V' \subset \mathbb{R}^n \to \mathbb{R}^n$. Now let $\mathtt{y}^{-1}U' = V'$ and form the composition $\psi := F \circ \phi$ which is defined on U' and must have the form

$$\begin{gathered} z_1 = x_1 \\ z_2 = x_2 \\ \vdots \\ z_r = x_r \\ z_{r+1} = \psi_{r+1}(x_1, ..., x_n) \\ \vdots \\ z_n = \psi_n(x_1, ..., x_n) \end{gathered}$$

From here is it is not hard to show that $z_{r+1} = \cdots = z_n = 0$ is exactly the set $\psi(M \cap U')$ and since $\mathtt{y}$ restricted to a $M \cap U'$ is a coordinate system so we see that ψ restricted to $M \cap U'$ is a coordinate system for M. Now notice that in fact ψ maps a point with standard (identity) coordinates $(a_1, ..., a_n)$ onto $(a_1, ..., a_r, 0, ..., 0)$. Now remembering that we renumbered coordinates in the middle of the discussion we have proved the following theorem.



Theorem 3.4 *If M is an r-dimensional regular submanifold of $\mathbb{R}^n$ then for every $p \in M$ there exists at least one r-dimensional coordinate plane P such that linear projection $P \to \mathbb{R}^n$ restricts to a coordinate system for M defined on some neighborhood of p.*

3.11 Manifolds with boundary.

For the general Stokes theorem where the notion of flux has its natural setting we will need to have a concept of a manifold with boundary . A basic example to keep in mind is the closed hemisphere S^2_+ which is the set of all $(x, y, z) \in S^2$ with $z \geq 0$.

Let $\lambda \in \mathsf{M}^*$ be a continuous form on a Banach space M. In the case of $\mathbb{R}^n$ it will be enough to consider projection onto the first coordinate x^1. Now let $\mathsf{M}^+_\lambda = \{x \in \mathsf{M}\colon\ \lambda(x) \geq 0\}$ and $\mathsf{M}^-_\lambda = \{x \in \mathsf{M}\colon\ \lambda(x) \leq 0\}$. The set $\partial\mathsf{M}^+_\lambda = \partial\mathsf{M}^-_\lambda = \{x \in \mathsf{M}\colon\ \lambda(x) = 0\}$ is the kernel of λ and is the boundary of both M^+_λ and M^-_λ. Clearly M^+_λ and M^-_λ are homeomorphic and $\partial\mathsf{M}^+_\lambda$ is a closed subspace. The space M^-_λ is the model space for a manifold with boundary and is called a (negative) **half space**.

Remark 3.6 *We have chosen the space M^-_λ rather than M^+_λ on purpose. The point is that later we will wish to have a prescription whereby one of the coordinate vectors $\frac{\partial}{\partial x^i}$ will always be outward pointing at $\partial\mathsf{M}^-_\lambda$ while the remaining* *coordinate vectors in their given order are positively oriented on $\partial\mathsf{M}_\lambda^-$ in a sense we will define later. Now, $\frac{\partial}{\partial x^1}$ is outward pointing for $\mathbb{R}^n_{x^1\leq 0}$ but not for $\mathbb{R}^n_{x^1\geq 0}$. One might be inclined to think that we should look at $\mathbb{R}^n_{x^j\geq 0}$ for some other choice of j - the most popular being the last coordinate x^n. Although this could be done (and is done in the literature) it would actually only make things more complicated. The problem is that if we declare $\frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, ..., \frac{\partial}{\partial x^{n-1}}$ to be positively oriented on $\mathbb{R}^{n-1}\times 0$ whenever$\frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, ..., \frac{\partial}{\partial x^n}$ is positively oriented on $\mathbb{R}^n$ we will introduce a minus sign into Stokes' theorem in every other dimension!*

We let M_λ^- (and M_λ^+) have the relative topology. Since $\mathsf{M}_\lambda^-\subset\mathsf{M}$ we already have a notion of differentiability on M_λ^- (and hence M_λ^+) via definition 3.5. The notions of C^r maps of open subsets of half space and diffeomorphisms etc. is now defined in the obvious way. For convenience let us define for an open set $U\subset\mathsf{M}_\lambda^-$ (always relatively open) the following slightly inaccurate notations; let ∂U denote $\partial\mathsf{M}_\lambda^-\cap U$ and $\mathrm{int}(U)$ denote $U\setminus\partial U$.

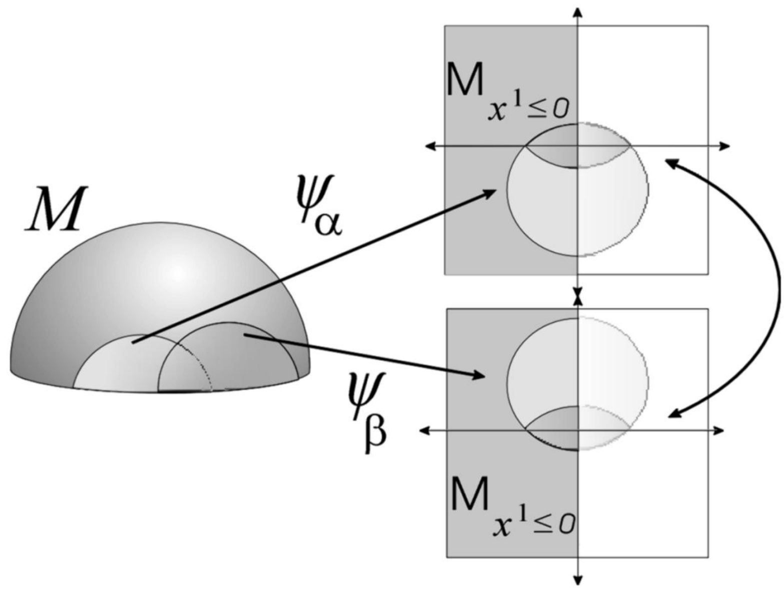
We have the following three facts:

1. First, it is an easy exercise to show that if $f: U\subset\mathsf{M}\to\mathsf{F}$ is C^r differentiable (with $r\geq 1$) and g is another such map with the same domain, then if $f=g$ on $\mathsf{M}_\lambda^-\cap U$ then $D_xf=D_xg$ for all $x\in\mathsf{M}_\lambda^-\cap U$.
2. Let F_ℓ^- be a half space in a Banach space F. If $f: U\subset\mathsf{M}\to\mathsf{F}_\alpha^-$ is C^r differentiable (with $r\geq 1$) and $f(x)\in\partial\mathsf{F}_\ell^-$ then $D_xf:\mathsf{M}\to\mathsf{F}$ must have its image in $\partial\mathsf{F}_\ell^-$.
3. Let $f: U_1\subset\mathsf{M}_\lambda^-\to U_2\subset\mathsf{F}_\ell^-$ be a diffeomorphism (in our new extended sense). Assume that $\mathsf{M}_\lambda^-\cap U_1$ and $\mathsf{F}_\ell^-\cap U_2$ are not empty. Then f induces diffeomorphisms $\partial U_1\to\partial U_2$ and $\mathrm{int}(U_1)\to\mathrm{int}(U_2)$.

These three claims are not exactly obvious but there are very intuitive. On the other hand, none of them are difficult to prove and we will leave these as exercises. These facts show that the notion of a boundary defined here and in general below is a well defined concept and is a natural notion in the context of differentiable manifolds; it is a "differentiable invariant".

We can now form a definition of manifold with boundary in a fashion completely analogous to the definition of a manifold without boundary. A **half space chart** $\mathtt{x}_\alpha$ for a set M is a bijection of some subset U_α of M onto an open subset of M_λ^- (or M_λ^+ for many authors). A C^r **half space atlas** is a collection $\mathtt{x}_\alpha, U_\alpha$ of such half space charts such that for any two, say $\mathtt{x}_\alpha, U_\alpha$ and $\mathtt{x}_\beta, U_\beta$, the map $\mathtt{x}_\alpha\circ\mathtt{x}_\beta^{-1}$ is a C^r diffeomorphism on its natural domain (if non-empty). Note: "Diffeomorphism" means diffeomorphism in the extended the sense of a being homeomorphism and such that both $\mathtt{x}_\alpha\circ\mathtt{x}_\beta^{-1} :: \mathsf{M}_\lambda^-\to\mathsf{M}$ and its inverse are C^r in the sense of definition 3.5. The reader may wish to review the definition of manifold with boundary given via pseudogroups of transformations 3.10.

Definition 3.24 *A C^r-**manifold with boundary** $(M,\mathcal{A})$ is a pair consisting of a set M together with a maximal atlas of half space charts $\mathcal{A}$. The manifold*



topology is that generated by the domains of all such charts. The boundary of M is denoted by ∂M and is the set of points that have their image in the boundary M_0 of M_λ^- under some and hence every chart.

Colloquially, one usually just refers to M as a manifold with boundary and forgoes the explicit reference the atlas.

Definition 3.25 *The interior of a manifold with boundary is $M \backslash \partial M$ a manifold without boundary and is denoted $\overset{\circ}{M}$.*

Exercise 3.9 *Show that $M \cup \partial M$ is closed and that $M \backslash \partial M$ is open.*

In the present context, a manifold without boundary that is compact (and hence closed in the usual topological sense if M is Hausdorff) is often referred to as a **closed manifold**. If no component of a some manifold without boundary is compact it is called an **open manifold**. For example, the "interior" $\overset{\circ}{M}$ of a connected manifold M with nonempty boundary is never compact an open manifold in the above sense.

So $\overset{\circ}{M}$ will be an open manifold if every component of M contained part of the boundary.

Remark 3.7 *The phrase "closed manifold" is a bit problematic since the word closed is acting as an adjective and so conflicts with the notion of closed in the ordinary topological sense. For this reason we will try to avoid this terminology and use instead the phrase "**compact manifold without boundary**".*

Remark 3.8 (Warning) *Some authors let M denote the interior, so that $M \cup \partial M$ is the closure and is the manifold with boundary in our sense.*

Theorem 3.5 *∂M is a C^r manifold (without boundary) with an atlas being given by all maps of the form $\mathtt{x}_\alpha|, U_\alpha \cap \partial M$. The manifold ∂M is called the* ***boundary of*** *M.*

Idea of Proof. The truth of this theorem becomes obvious once we recall what it means for a chart overlap map $\mathtt{y} \circ \mathtt{x}^{-1} : U \to V$ to be a diffeomorphism in a neighborhood a point $x \in U \cap \mathsf{M}^+_\lambda$. First of all there must be a set U' containing U that is open in M and an extension of $\mathtt{y} \circ \mathtt{x}^{-1}$ to a differentiable map on U'. But the same is true for $(\mathtt{y} \circ \mathtt{x}^{-1})^{-1} = \mathtt{x} \circ \mathtt{y}^{-1}$. The extensions are inverses of each other on U and V. But we must also have that the derivatives of the chart overlap maps are isomorphisms at all points up to and including ∂U and ∂V. But then the inverse function theorem says that there are neighborhoods of points in ∂U in M and ∂V in M such that these extensions are actually diffeomorphisms and inverses of each other. Now it follows that the restrictions $\mathtt{y} \circ \mathtt{x}^{-1}\big|_{\partial U} : \partial U \to \partial V$ are diffeomorphisms. In fact, this is the main point of the comment (3) above and we have now seen the idea of its proof also. ■

Example 3.23 *The closed ball $\overline{B}(p, R)$ in $\mathbb{R}^n$ is a manifold with boundary $\partial\overline{B}(p, R) = S^{n-1}$.*

Example 3.24 *The hemisphere $S^n_+ = \{x \in \mathbb{R}^{n+1} : x^{n+1} \geq 0\}$ is a manifold with boundary.*

Exercise 3.10 *Is the Cartesian product of two manifolds with boundary necessarily a manifold with boundary?*

3.12 Problem Set

1. Let M and N be C^∞ manifolds, and $f : M \to N$ a C^∞ map. Suppose that M is compact, N is connected. Suppose further that f is injective and that $T_x f$ is an isomorphism for each $x \in M$. Show that f is a diffeomorphism.

2. Let M_1 , M_2 and M_3 be C^∞ manifolds.

   (a) Show that $(M_1 \times M_2) \times M_3$ is diffeomorphic to $M_1 \times (M_2 \times M_3)$ in a natural way.

   (b) Show that $f : M \to M_1 \times M_2$ is C^∞ if and only if the composite maps $pr_1 \circ f : M \to M_1$ and $pr_2 \circ f : M \to M_2$ are both C^∞.

3. Show that a C^r a manifold M is connected as a topological space if and only it is C^r-path connected in the sense that for any two points $p_1, p_2 \in M$ there is a C^r map $c : [0, 1] \to M$ such that $c(0) = p_1$ and $c(1) = p_2$.

4. An affine map between two vector spaces is a map of the form $x \mapsto Lx + b$ for a fixed vector. An affine space is has the structure of a vector space except that only the difference of vectors rather than the sum is defined. For example,let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map and let $A_c = \{x \in \mathbb{R}^n : Lx = c\}$. Now A_c is not a vector subspace of $\mathbb{R}^n$ it is an affine space with "difference space" $\ker L$. This means that for any $p_1, p_2 \in A_c$ the difference $p_2 - p_1$ is an element of $\ker L$. Furthermore, for any fixed $p_0 \in A_c$ the map $x \mapsto p_0 + x$ is a bijection $\ker L \cong A_c$. Show that the set of all such bijections form an atlas for A_c such that for any two charts from this atlas the overlap map is affine isomorphism from $\ker L$ to itself. Develop a more abstract definition of topological affine space with a Banach space as difference space. Show show that this affine structure is enough to make sense out of the derivative via the difference quotient.

5. Show that $\mathrm{V} = \{c : [a, b] \to \mathbb{R}^n : c \text{ is } C^k,\ c(a) = c(b) = 0\ \}$ is a Banach space when endowed with the norm $\|c\| := \sum_{j \leq k} \sup_{t \in [a,b]} \left| c^{(j)}(t) \right|$. Give the set $M = \{c : [a, b] \to \mathbb{R}^n : c \text{ is } C^k,\ c(a) = p_0 \text{ and } c(b) = p_1\}$ of a smooth manifold modeled on the Banach space V.

6. A $k-$frame in $\mathbb{R}^n$ is a linearly independent ordered set of vectors $(v_1, ..., v_k)$. Show that the set of all $k-$frames in $\mathbb{R}^n$ can be given the structure of a differentiable manifold. This kind of manifold is called a **Stiefel manifold**.

7. Embed the Stiefel manifold of $k-$frames in $\mathbb{R}^n$ into a Euclidean space $\mathbb{R}^N$ for some large N.

Chapter 4

The Tangent Structure

4.1 Rough Ideas II

Let us suppose that we have two coordinate systems $\mathtt{x} = (x^1, x^2, ..., x^n)$ and $\bar{\mathtt{x}} = (\bar{x}^1, x^2, ..., \bar{x}^n)$ defined on some common open set $U \cap V$ of a finite dimensional differentiable manifold M. Let us also suppose that we have two lists of numbers $v^1, v^2, ..., v^n$ and $\bar{v}^1, \bar{v}^2, ...\bar{v}^n$ somehow coming from the respective coordinate systems and associated to a point p in the common domain of the two coordinate systems. Suppose that the lists are related to each other by

$$v^i = \sum_{k=1}^{n} \frac{\partial x^i}{\partial \bar{x}^k} \bar{v}^k$$

where the derivatives $\frac{\partial x^i}{\partial \bar{x}^k}$ are evaluated at the coordinates $\bar{x}^1(p), \bar{x}^2(p), ..., \bar{x}^n(p)$. Now if f is a function also defined in a neighborhood of p then the representative functions $f(x)$ and $f(\bar{x})$ for f in the respective systems are related by

$$\frac{\partial f(x)}{\partial x^i} = \sum_{k=1}^{n} \frac{\partial \bar{x}^k}{\partial x^i} \frac{\partial f(\bar{x})}{\partial \bar{x}^k}.$$

The chain rule then implies that

$$\frac{\partial f(x)}{\partial x^i} v^i = \frac{\partial f(\bar{x})}{\partial \bar{x}^i} \bar{v}^i. \tag{4.1}$$

Thus if we had a list $v^1, v^2, ..., v^n$ for every coordinate chart on the manifold whose domains contain the point p and related to each other as above then we say that we have a tangent vector v at $p \in M$. If we define the **directional derivative** of a function f at p in the direction of v by

$$vf := \frac{\partial f}{\partial x^i} v^i$$

then we are in business since by 4.1 it doesn't matter which coordinate system we use. Because of this we think of (v^i) and $(\bar{v}^i)$ as representing the same geometric object; a tangent vector at p, but with respect to two different coordinate systems. This way of thinking about tangent vectors is the basis of the first of several equivalent formal definitions of tangent vector which we present below. How do we get such vectors in a natural way? Well, one good way is from the notion of the velocity of a curve. A differentiable curve though $p \in M$ is a map $c : (-a, a) \to M$ with $c(0) = p$ such that the coordinate expressions for the curve $x^i(t) = (x^i \circ c)(t)$ are all differentiable. We then take

$$v^i := \frac{dx^i}{dt}(0)$$

for each coordinate system $\mathtt{x} = (x^1, x^2, ...x^n)$ with p in its domain. This gives a well defined tangent vector v at p called the velocity of c at $t = 0$. We denote this by $c'(0)$ or by $\frac{dc}{dt}(0)$. Of course we could have done this for each $t \in (-a, a)$ by defining $v^i(t) := \frac{dx^i}{dt}(t)$ and we would get a smoothly varying family of velocity vectors $c'(t)$ defined at the points $c(t) \in M$.

The set of all tangent vectors at a point $p \in M$ is a vector space since we can clearly choose a coordinate system in which to calculate and then the vectors just appear as $n-$tuples; that is, elements of R^n. The vector space operations (scaling and vector addition) remain consistently related when viewed in another coordinate system since the relations are linear. The set of all tangent vectors at $p \in M$ is called the tangent space at p. We will denote these by T_pM for the various points p in M. The tangent spaces combine to give another differentiable manifold of twice the dimension called the tangent bundle. The coordinates for the tangent bundle come from those that exist on M by combining "point coordinates" $(x^1, ..., x^n)$ with the components of the vectors with respect to the standard basis. Thus the coordinates of a tangent vector v at p are simply $(x^1, ..., x^n; v^1, ..., v^n)$. More on the tangent bundle below.

4.2 Tangent Vectors

For a submanifold S of $\mathbb{R}^n$ we have a good idea what a tangent vector ought to be. Let $t \mapsto c(t) = (x^1(t), ..., x^n(t))$ be a $C^\infty-$curve with image contained in S and passing through the point $p \in S$ at time $t = 0$. Then the vector $v = \dot{c}(t) = \frac{d}{dt}\big|_{t=0} c(t)$ is tangent to S at p. So to be **tangent** to S at p is to be the velocity at p of some curve in S through p. Of course, we must consider v to be *based* at $p \in S$ in order to distinguish it from parallel vectors of the same length that may be velocity vectors of curves going through other points. One way to do this is to write the tangent vector as a pair (p, v) where p is the base point. In this way we can construct the space TS of all vectors tangent to S as a subset of $\mathbb{R}^n \times \mathbb{R}^n$

$$TS = \{(p, v) \in \mathbb{R}^n \times \mathbb{R}^n : p \in S \text{ and } v \text{ tangent to } S \text{ at } p\}$$

This method will not work well for manifolds that are not given as submanifolds of $\mathbb{R}^n$. We will now give three methods of defining tangent vectors at a point of a differentiable manifold.

Definition 4.1 (Tangent vector via charts) *Consider the set of all admissible charts $(\mathtt{x}_\alpha, U_\alpha)_{\alpha\in A}$ on M indexed by some set A for convenience. Next consider the set T of all triples (p, v, α) such that $p \in U_\alpha$. Define an equivalence relation so that $(p, v, \alpha) \sim (q, w, \beta)$ if and only if $p = q$ and*

$$D(\mathtt{x}_\beta \circ \mathtt{x}_\alpha^{-1})\big|_{\mathtt{x}(p)} \cdot v = w.$$

In other words, the derivative at $\mathtt{x}(p)$ of the coordinate change $\mathtt{x}_\beta \circ \mathtt{x}_\alpha^{-1}$ "identifies" v with w. Tangent vectors are then equivalence classes. The tangent vectors at a point p are those equivalence classes represented by triples with first slot occupied by p. The set of all tangent vectors at p is written as T_pM and is called the tangent space at p. The ***tangent bundle*** *TM is the disjoint union of all the tangent spaces for all points in M.*

$$TM := \bigsqcup_{p\in M} T_pM$$

This viewpoint takes on a more familiar appearance in finite dimensions if we use a more classical notation; Let $\mathtt{x}, U$ and $\mathtt{y}, V$ be two charts containing p in their domains. If an $n-$tuple $(v^1, ..., v^n)$ represents a tangent vector at p from the point of view of $\mathtt{x}, U$ and if the $n-$tuple $(w^1, ..., w^n)$ represents the same vector from the point of view of $\mathtt{y}, V$ then

$$w^i = \sum_{j=1}^{n} \frac{\partial y^i}{\partial x^j}\bigg|_{\mathtt{x}(p)} v^j$$

where we write the change of coordinates as $y^i = y^i(x^1, ..., x^n)$ with $1 \le i \le n$.

We can get a similar expression in the infinite dimensional case by just letting $D(\mathtt{y} \circ \mathtt{x}^{-1})\big|_{\mathtt{x}(p)}$ be denoted by $\frac{\partial \mathtt{y}}{\partial \mathtt{x}}\Big|_{\mathtt{x}(p)}$ then writing

$$w = \frac{\partial \mathtt{y}}{\partial \mathtt{x}}\bigg|_{\mathtt{x}(p)} v.$$

Recall, that a manifold with boundary is modeled on a half space $\mathsf{M}_\lambda^- := \{x \in \mathsf{M} : \lambda(x) \le 0\}$ where M is some Banach space and λ is a continuous linear functional on M. The usual case is where $\mathsf{M} = \mathbb{R}^n$ for some n and with our conventions λ is then taken to be the first coordinate function x^1. If M is a manifold with boundary, the tangent bundle TM is defined as before. For instance, even if $p \in \partial M$ the fiber T_pM may still be thought of as consisting of equivalence classes where $(p, v, \alpha) \sim (p, w, \beta)$ if and only if $D(\mathtt{x}_\beta \circ \mathtt{x}_\alpha^{-1})\big|_{\mathtt{x}_\alpha(p)} \cdot v = w$. Notice that for a given chart $\mathtt{x}_\alpha$, the vectors v in the various (p, v, α) still run throughout M and so T_pM still has tangent vectors "pointing in all directions".

On the other hand, if $p \in \partial M$ then for any half-space chart $\mathtt{x}_\alpha : U_\alpha \to \mathsf{M}^-_\lambda$ with p in its domain, $T\mathtt{x}_\alpha^{-1}(\partial \mathsf{M}^-_\lambda)$ is a subspace of T_pM. This is the subspace of vectors tangent to the boundary and is identified with the tangent space to ∂M (also a manifold as well shall see) $T_p\partial M$.

Exercise 4.1 *Show that this subspace does not depend on the choice of* $\mathtt{x}_\alpha$.

Definition 4.2 *Let M be a manifold with boundary and suppose that $p \in \partial M$. A tangent vector $v = [(p, v, \alpha)] \in T_pM$ is said to be outward pointing if $\lambda(v) > 0$ and inward pointing if $\lambda(v) < 0$. Here $\alpha \in A$ indexes charts as before; $\mathtt{x}_\alpha : U_\alpha \to \mathsf{M}^-_\lambda$.*

Exercise 4.2 *Show that the above definition is independent of the choice of the half-space chart $\mathtt{x}_\alpha : U_\alpha \to \mathsf{M}^-_\lambda$.*

Definition 4.3 (Tangent vectors via curves) *Let p be a point in a C^r manifold with $r > 1$. Suppose that we have C^r curves c_1 and c_2 mapping into the manifold M, each with open domains containing $0 \in \mathbb{R}$ and with $c_1(0) = c_2(0) = p$. We say that c_1 is tangent to c_2 at p if for all C^r functions $f : M \to \mathbb{R}$ we have $\frac{d}{dt}\big|_{t=0} f \circ c_1 = \frac{d}{dt}\big|_{t=0} f \circ c_2$. This is an equivalence relation on the set of all such curves. Define a **tangent vector at** p to be an equivalence class $X_p = [c]$ under this relation. In this case we will also write $\dot{c}(0) = X_p$. The **tangent space** T_pM is defined to be the set of all tangent vectors at $p \in M$. The **tangent bundle** TM is the disjoint union of all the tangent spaces for all points in M.*

$$TM := \bigsqcup_{p \in M} T_pM$$

Remark 4.1 (Notation) *Let $X_p \in T_pM$ for p in the domain of an admissible chart $(U_\alpha, \mathtt{x}_\alpha)$. Under our first definition, in this chart, X_p is represented by a triple (p, v, α). We denote by $[X_p]_\alpha$ the principle part v of the representative of X_p. Equivalently, $[X_p]_\alpha = D(\mathtt{x}_\alpha \circ c)|_0$ for any c with $c'(0) = X_p$ i.e. $X_p = [c]$ as in definition 4.3.*

For the next definition of tangent vector we need to think about the set of real valued functions defined near a some fixed point p. We want a formal way of considering two functions that agree on some open set containing a point as being locally the same at that point. To this end we take the set F_p of all smooth functions with open domains of definition containing $p \in M$. Define two such functions to be equivalent if they agree on some small open set containing p. The equivalence classes are called **germs of smooth functions at** p and the set of all such is denoted $\mathcal{F}_p = F_p/\sim$. It is easily seen that $\mathcal{F}_p$ is naturally a vector space and we can even multiply germs in the obvious way (just pick representatives for the germs of f and g, take restrictions of these to a common domain, multiply and then take the germ of the result). This makes $\mathcal{F}_p$ a ring (and an algebra over the field $\mathbb{R}$). Furthermore, if f is a representative for the equivalence class $\breve{f} \in \mathcal{F}_p$ then we can unambiguously define the value of $\breve{f}$ at p

by $\breve{f}(p) = f(p)$. Thus we have an **evaluation map** $ev_p : F_p \to \mathbb{R}$. We are really just thinking about functions defined near a point and the germ formalism is convenient whenever we do something where it only matters what is happening near p. We will thus sometimes abuse notation and write f instead of $\breve{f}$ to denote the germ represented by a function f. In fact, we don't really absolutely need the germ idea for the following kind of definition to work so we could put the word "germ" in parentheses.

Remark 4.2 *We have defined $\mathcal{F}_p$ using smooth functions but we can also define in an obvious way $\mathcal{F}_p^r$ using C^r functions.*

Definition 4.4 *Let $\breve{f}$ be the germ of a function $f :: M \to \mathbb{R}$. Let us define the **differential of** f **at** p to be a map $df(p) : T_pM \to \mathbb{R}$ by simply composing a curve c representing a given vector $X_p = [c]$ with f to get $f \circ c :: \mathbb{R} \to \mathbb{R}$. Then define*

$$df(p) \cdot X_p = \left.\frac{d}{dt}\right|_{t=0} f \circ c \in \mathbb{R}.$$

Clearly we get the same answer if we use another function with the same germ at p. The differential at p is also often written as $df|_p$. More generally, if $f :: M \to \mathsf{E}$ for some Banach space E then $df(p) : T_pM \to \mathsf{E}$ is defined by the same formula.

Remark 4.3 (Very useful notation) *This use of the "differential" notation for maps into vector spaces is useful for coordinates expressions. Let $p \in U$ where $\mathtt{x}, U$ is a chart and consider again a tangent vector X_p at p. Then the local representative of X_p (or principal part of X_p) in this chart is exactly $d\mathtt{x}(X_p)$.*

Definition 4.5 *A **derivation** of the algebra $\mathcal{F}_p$ is a map $\mathcal{D} : \mathcal{F}_p \to \mathbb{R}$ such that $\mathcal{D}(\breve{f}\breve{g}) = \breve{f}(p)\mathcal{D}\breve{g} + \breve{g}(p)\mathcal{D}\breve{f}$ for all $\breve{f}, \breve{g} \in \mathcal{F}_p$.*

Notation 4.1 *The set of all derivations on $\mathcal{F}_p$ is easily seen to be a real vector space and we will denote this by $\mathsf{Der}(\mathcal{F}_p)$.*

We will now define the operation of a tangent vector on a function or more precisely, on germs of functions at a point.

Definition 4.6 *Let $\mathcal{D}_{X_p} : \mathcal{F}_p \to \mathbb{R}$ be given by the rule $\mathcal{D}_{X_p}\breve{f} = df(p) \cdot X_p$.*

Lemma 4.1 *$\mathcal{D}_{X_p}$ is a derivation of the algebra $\mathcal{F}_p$. That is $\mathcal{D}_{X_p}$ is $\mathbb{R}-$linear and we have $\mathcal{D}_{X_p}(\breve{f}\breve{g}) = \breve{f}(p)\mathcal{D}_{X_p}\breve{g} + \breve{g}(p)\mathcal{D}_{X_p}\breve{f}$.*

A basic example of a derivation is the partial derivative operator $\left.\frac{\partial}{\partial x^i}\right|_{x_0}$: $f \mapsto \frac{\partial f}{\partial x^i}(x_0)$. We shall show that for a smooth n-manifold these form a basis for the space of all derivations at a point $x_0 \in M$. This vector space of all derivations is naturally isomorphic to the tangent space at x_0. Since this is certainly a local problem it will suffice to show this for $x_0 \in \mathbb{R}^n$. In the literature

$\mathcal{D}_{X_p}\breve{f}$ is written $X_p f$ and we will also use this notation. As indicated above, if M is finite dimensional and C^∞ then all derivations of $\mathcal{F}_p$ are given in this way by tangent vectors. Thus in this case we can and will abbreviate $\mathcal{D}_{X_p} f = X_p f$ and actually *define* tangent vectors to be derivations. For this we need a couple of lemmas:

Lemma 4.2 *If c is (the germ of) a constant function then $\mathcal{D}c = 0$.*

Proof. Since $\mathcal{D}$ is $\mathbb{R}-$linear this is certainly true if $c = 0$. Also, linearity shows that we need only prove the result for $c = 1$. Then

$$\begin{aligned} \mathcal{D}1 &= \mathcal{D}(1^2) \\ &= (\mathcal{D}1)1 + 1\mathcal{D}1 = 2\mathcal{D}1 \end{aligned}$$

and so $\mathcal{D}1 = 0$. ■

Lemma 4.3 *Let $f :: (\mathbb{R}^n, x_0) \to (\mathbb{R}, f(x_0))$ be defined and C^∞ in a neighborhood of x_0. Then near x_0 we have*

$$f(x) = f(x_0) + \sum_{1 \le i \le n} (x^i - x_0^i) \left[\frac{\partial f}{\partial x^i}(x_0) + a^i(x) \right]$$

for some smooth functions $a^i(x)$ with $a^i(x_0) = 0$.

Proof. Write $f(x) - f(x_0) = \int_0^1 \frac{\partial}{\partial t} [f(x_0 + t(x - x_0))] \, dt = \sum_{i=1}^n (x^i - x_0^i) \int_0^1 \frac{\partial f}{\partial x^i}(x_0 + t(x - x_0)) dt$. Integrate the last integral by parts to get

$$\begin{aligned} &\int_0^1 \frac{\partial f}{\partial x^i} [(x_0 + t(x - x_0))] \, dt \\ &= t \frac{\partial f}{\partial x^i} [(x_0 + t(x - x_0))] \bigg|_0^1 - \int_0^1 t \sum_{i=1}^n (x^i - x_0^i) \frac{\partial^2}{\partial x^i \partial x^j} (x_0 + t(x - x_0)) dt \\ &= \frac{\partial f}{\partial x^i}(x_0) + a^i(x) \end{aligned}$$

where the term $a^i(x)$ clearly satisfies the requirements. ■

Proposition 4.1 *Let $\mathcal{D}_{x_0}$ be a derivation on $\mathcal{F}_{x_0}$ where $x_0 \in \mathbb{R}^n$. Then*

$$\mathcal{D}_{x_0} = \sum_{i=1}^n \mathcal{D}_{x_0}(x^i) \left. \frac{\partial}{\partial x^i} \right|_{x_0}.$$

In particular, $\mathcal{D}$ corresponds to a unique vector at x_0 and by the association $(\mathcal{D}_{x_0}(x^1), ..., \mathcal{D}_{x_0}(x^n)) \mapsto \mathcal{D}_{x_0}$ we get an isomorphism of $\mathbb{R}^n$ with $\mathsf{Der}(\mathcal{F}_p)$.

Proof. Apply $\mathcal{D}$ to both sides of

$$f(x) = f(x_0) + \sum_{1 \le i \le n} (x^i - x_0^i) \left[\frac{\partial f}{\partial x^i}(x_0) + a^i(x) \right].$$

and use 4.2 to get the formula of the proposition. The rest is easy but it is important to note that $a^i(x)$ is in the domain of $\mathcal{D}$ and so must be C^∞ near x_0. In fact, a careful examination of the situation reveals that we need to be in the C^∞ category for this to work and so on a C^r manifold for $r < \infty$, the space of all derivations of germs of C^r functions is not the same as the tangent space. ■

An important point is that the above construction carries over via charts to manifolds. The reason for this is that if $\mathtt{x}, U$ is a chart containing a point p in a smooth n manifold then we can define an isomorphism between $\mathsf{Der}(\mathcal{F}_p)$ and $\mathsf{Der}(\mathcal{F}_{\mathtt{x}(p)})$ by the following simple rule:

$$\mathcal{D}_{\mathtt{x}(p)} \mapsto \mathcal{D}_p$$
$$\text{where } \mathcal{D}_p f = \mathcal{D}_{\mathtt{x}(p)}(f \circ \mathtt{x}^{-1}).$$

The one thing that must be noticed is that the vector $(\mathcal{D}_{\mathtt{x}(p)}(x^1), ..., \mathcal{D}_{\mathtt{x}(p)}(x^n))$ transforms in the proper way under change of coordinates so that the correspondence induces a well defined 1-1 linear map between T_pM and $\mathsf{Der}(\mathcal{F}_p)$. So using this we have one more possible definition of tangent vectors that works on finite dimensional C^∞ manifolds:

Definition 4.7 (Tangent vectors as derivations) *Let M be a C^∞ manifold of dimension $n < \infty$. Consider the set of all (germs of) C^∞ functions $\mathcal{F}_p$ at $p \in M$. A **tangent vector** at p is a linear map $X_p : \mathcal{F}_p \to \mathbb{R}$ that is also a derivation in the sense that for $f, g \in \mathcal{F}_p$*

$$X_p(fg) = g(p)X_pf + f(p)X_pg.$$

Once again, the tangent space at p is the set of all tangent vectors at p and the tangent bundle is defined as a disjoint union of tangent spaces as before.

In any event, even in the general case of a C^r Banach manifold with $r \geq 1$ a tangent vector determines a derivation written $X_p : f \mapsto X_pf$. However, in this case the derivation maps $\mathcal{F}_p^r$ to $\mathcal{F}_p^{r-1}$. Also, on infinite dimensional manifolds, even if we consider only the C^∞ case, there may be derivations not coming from tangent vectors as given in definition 4.3 or in definition 4.1. At least we have not shown that this cannot happen.

4.3 Interpretations

We will now show how to move from one definition of tangent vector to the next. For simplicity let us assume that M is a smooth (C^∞) n-manifold.

1. Suppose that we think of a tangent vector X_p as an equivalence class of curves represented by $c : I \to M$ with $c(0) = p$. We obtain a derivation by defining
$$X_pf := \left.\frac{d}{dt}\right|_{t=0} f \circ c$$

We can define a derivation in this way even if M is infinite dimensional but the space of derivations and the space of tangent vectors may not match up. We may also obtain a tangent vector in the sense of definition4.1 by letting X_p be associated to the triple (p, v, α) where $v^i := \frac{d}{dt}\big|_{t=0} x^i_\alpha \circ c$ for a chart $\mathtt{x}_\alpha, U_\alpha$ with $p \in U_\alpha$.

2. If X_p is a derivation at p and $U_\alpha, \mathtt{x}_\alpha = (x^1, ..., x^n)$ an admissible chart with domain containing p, then X_p, as a tangent vector a la definition 4.1, is represented by the triple (p, v, α) where $v = (v^1, ...v^n)$ is given by

$$v^i = X_p x^i \text{ (acting as a derivation)}$$

3. Suppose that, a la definition 4.1, a vector X_p at $p \in M$ is represented by (p, v, α) where $v \in \mathsf{M}$ and α names the chart $\mathtt{x}_\alpha, U_\alpha$. We obtain a derivation by defining

$$X_p f = D(f \circ \mathtt{x}_\alpha^{-1})\big|_{\mathtt{x}_\alpha(p)} \cdot v$$

In case the manifold is modeled on $\mathbb{R}^n$ we have the more traditional notation

$$X_p f = \sum v^i \left.\frac{\partial}{\partial x^i}\right|_p f.$$

for $v = (v^1, ...v^n)$.

The notation $\frac{\partial}{\partial x^i}\big|_p$ is made precise by the following:

Definition 4.8 *For a chart $\mathtt{x} = (x^1, ..., x^n)$ with domain U containing a point p we define a tangent vector $\frac{\partial}{\partial x^i}\big|_p \in T_pM$ by*

$$\left.\frac{\partial}{\partial x^i}\right|_p f = D_i(f \circ \mathtt{x}^{-1})(\mathtt{x}(p))$$

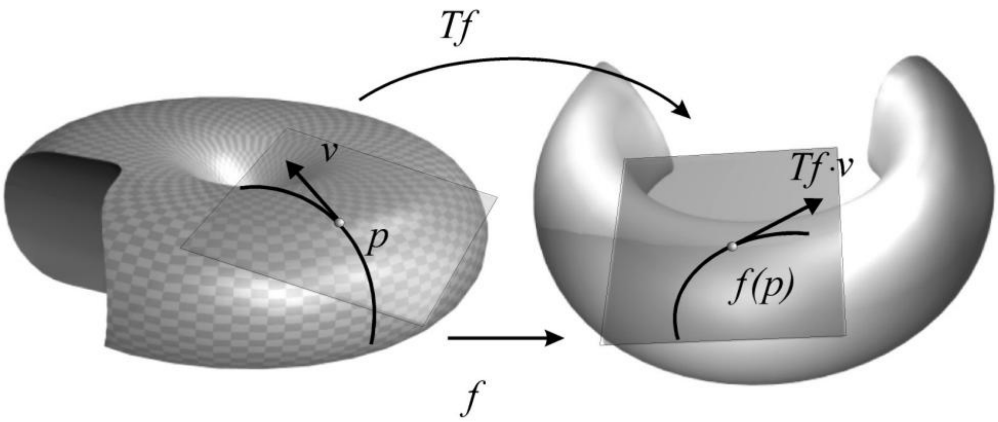
Alternatively, we may take $\frac{\partial}{\partial x^i}\big|_p$ to be the equivalence class of a coordinate curve. In other words, $\frac{\partial}{\partial x^i}\big|_p$ is the velocity at $\mathtt{x}(p)$ of the curve $t \mapsto \mathtt{x}^{-1}(x^1(p), ..., x^i(p)+ t, ..., x^n(p))$ defined for sufficiently small t.

We may also identify $\frac{\partial}{\partial x^i}\big|_p$ as the vector represented by the triple $(p, \mathsf{e}_i, \alpha)$ where e_i is the i-th member of the standard basis for $\mathbb{R}^n$ and α refers to the current chart $\mathtt{x} = \mathtt{x}_\alpha$.

Exercise 4.3 *For a finite dimensional $C^\infty-$manifold M and $p \in M$, let $\left(\mathtt{x}_\alpha = (x^1, ..., x^n), U_\alpha\right)$ be a chart whose domain contains p. Show that the vectors $\frac{\partial}{\partial x^1}\big|_p, ..., \frac{\partial}{\partial x^n}\big|_p$ (using our last definition of tangent vector) are a basis for the tangent space T_pM.*

4.4 The Tangent Map

The first definition given below of the tangent map of a smooth map $f : M, p \to N, f(p)$ will be considered our main definition but the others are actually equivalent at least for finite dimensional manifolds. Given f and p as above wish to define a linear map $T_p f : T_p M \to T_{f(p)} N$



Definition 4.9 *If we have a smooth function between manifolds*

$$f : M \to N$$

and we consider a point $p \in M$ and its image $q = f(p) \in N$. Choose any chart $(\mathtt{x}, U)$ containing p and a chart $(\mathtt{y}, V)$ containing $q = f(p)$ so that for any $v \in T_p M$ we have the representative $d\mathtt{x}(v)$ with respect to $\mathtt{x}, U$. Then the ***tangent map*** *$T_p f : T_p M \to T_{f(p)} N$ is defined by letting the representative of $T_p f \cdot v$ in the chart $(\mathtt{y}, V)$ be given by*

$$d\mathtt{y}(T_p f \cdot v) = D(\mathtt{y} \circ f \circ \mathtt{x}^{-1}) \cdot d\mathtt{x}(v).$$

This uniquely determines $T_p f \cdot v$ and the chain rule guarantees that this is well defined (independent of the choice of charts).

Since we have several definitions of tangent vector we expect to see several equivalent definitions of the tangent map. Here is another:

Definition 4.10 (Tangent map II) *If we have a smooth function between manifolds*

$$f : M \to N$$

and we consider a point $p \in M$ and its image $q = f(p) \in N$ then we define the ***tangent map*** *at p*

$$T_p f : T_p M \to T_q N$$

in the following way: Suppose that $v \in T_pM$ and we pick a curve c with $c(0) = p$ so that $v = [c]$, then by definition

$$T_pf \cdot v = [f \circ c] \in T_qN$$

where $[f \circ c] \in T_qN$ is the vector represented by the curve $f \circ c$.

Another alternative definition of tangent map that works for finite dimensional smooth manifolds is given in terms of derivations:

Definition 4.11 (Tangent Map III) *Let M be a smooth n-manifold. View tangent vectors as derivations as explained above. Then continuing our set up above and letting g be a smooth germ at $q = f(p) \in N$ we define $T_pf \cdot v$ as a derivation by*

$$(T_pf \cdot v)g = v(f \circ g)$$

It is easy to check that this defines a derivation on the (germs) of smooth functions at q and so is also a tangent vector in T_qM. This map is yet another version of the ***tangent map*** *T_pf.*

It is easy to check that for a smooth $f : M \to \mathsf{E}$ the differential $df(p) : T_pM \to \mathsf{E}$ is the composition of the tangent map T_pf and the canonical map $T_y\mathsf{E} \to \mathsf{E}$ where $y = f(p)$. Diagrammatically we have

$$df(p) : T_pM \overset{Tf}{\to} T\mathsf{E} = \mathsf{E} \times \mathsf{E} \overset{pr_1}{\to} \mathsf{E}.$$

4.5 The Tangent and Cotangent Bundles

4.5.1 Tangent Bundle

We have defined the **tangent bundle** of a manifold as the disjoint union of the tangent spaces $TM = \bigsqcup_{p\in M} T_pM$. We also gave similar definition of cotangent bundle. We show in proposition 4.2 below that TM is itself a differentiable manifold but first we record the following two definitions.

Definition 4.12 *Given a smooth map $f : M \to N$ as above then the tangent maps on the individual tangent spaces combine to give a map $Tf : TM \to TN$ on the tangent bundles that is linear on each fiber called the* ***tangent lift*** *or sometimes the* ***tangent map****.*

Definition 4.13 *The map $\tau_M : TM \to M$ defined by $\tau_M(v) = p$ for every $p \in T_pM$ is called the (tangent bundle) projection map. (The set TM together with the map $\tau_M : TM \to M$ is an example of a vector bundle defined in the sequel.)*

Proposition 4.2 *TM is a differentiable manifold and $\tau_M : TM \to M$ is a smooth map. Furthermore, for a smooth map $f : M \to N$ the tangent lift Tf is smooth and the following diagram commutes.*

$$\begin{array}{ccccc} & TM & \overset{Tf}{\to} & TN & \\ \tau_M & \downarrow & & \downarrow & \tau_N \\ & M & \overset{f}{\to} & N & \end{array}$$

Now for every chart $(\mathtt{x}, U)$ let $TU = \tau_M^{-1}(U)$. Charts on TM are defined using charts from M as follows

$$T\mathtt{x} : TU \to T\mathtt{x}(TU) \cong \mathtt{x}(U) \times \mathsf{M}$$
$$T\mathtt{x} : \xi \mapsto (\mathtt{x} \circ \tau_M(\xi), v)$$

where $v = d\mathtt{x}(\xi)$ is the principal part of ξ in the $\mathtt{x}$ chart. The chart $T\mathtt{x}, TU$ is then described by the composition

$$\xi \mapsto (\tau_M(\xi), \xi) \mapsto (\mathtt{x} \circ \tau_M(\xi), d\mathtt{x}(\xi))$$

but $\mathtt{x} \circ \tau_M(\xi)$ is usually abbreviated to just $\mathtt{x}$ so we may write the chart in the handy form $(\mathtt{x}, d\mathtt{x})$.

$$\begin{array}{ccc} TU & \overset{(\mathtt{x}, d\mathtt{x})}{\to} & \mathtt{x}(U) \times \mathsf{M} \\ \downarrow & & \downarrow \\ U & \overset{\mathtt{x}}{\to} & \mathtt{x}(U) \end{array}$$

For a finite dimensional manifold and a chart $\left(\mathtt{x} = (x^1, ..., x^n), U\right)$ any vector $\xi \in \tau_M^{-1}(U)$ can be written

$$\xi = \sum v^i(\xi) \left. \frac{\partial}{\partial x^i} \right|_{\tau_M(\xi)}$$

for some $v^i(\xi) \in \mathbb{R}$ depending on ξ. So in the finite dimensional case the chart is just written $(x^1, ..., x^n, v^1, ..., v^n)$.

Exercise 4.4 *Test your ability to interpret the notation by checking that each of these statements makes sense and is true:*

1) If $\xi = \xi^i \left.\frac{\partial}{\partial x^i}\right|_p$ and $\mathtt{x}_\alpha(p) = (a^1, ..., a^n) \in \mathtt{x}_\alpha(U_\alpha)$ then $T\mathtt{x}_\alpha(\xi) = (a^1, ..., a^n, \xi^1, ..., \xi^n) \in U_\alpha \times \mathbb{R}^n$.

2) If $v = [c]$ for some curve with $c(0) = p$ then

$$T\mathtt{x}_\alpha(v) = (\mathtt{x}_\alpha \circ c(0), \left.\frac{d}{dt}\right|_{t=0} \mathtt{x}_\alpha \circ c) \in U_\alpha \times \mathbb{R}^n$$

Suppose that $(U, T\mathtt{x})$ and $(V, T\mathtt{y})$ are two such charts constructed as above from two charts $U, \mathtt{x}$ and $V, \mathtt{y}$ and that $U \cap V \neq \emptyset$. Then $TU \cap TV \neq \emptyset$ and on the overlap we have the coordinate transitions $T\mathtt{y} \circ T\mathtt{x}^{-1}:(x, v) \mapsto (y, w)$ where

$$y = \mathtt{y} \circ \mathtt{x}^{-1}(x)$$
$$w = \sum_{k=1}^{n} D(\mathtt{y} \circ \mathtt{x}^{-1})\big|_{\mathtt{x}(p)}\, v$$

so the overlaps will be C^{r-1} whenever the $\mathtt{y} \circ \mathtt{x}^{-1}$ are C^r. Notice that for all $p \in \mathtt{x}(U \cap V)$ we have

$$\mathtt{y}(p) = \mathtt{y} \circ \mathtt{x}^{-1}(\mathtt{x}(p))$$
$$d\mathtt{y}(\xi) = D(\mathtt{y} \circ \mathtt{x}^{-1})\big|_{\mathtt{x}(p)}\, d\mathtt{x}(\xi)$$

or with our alternate notation

$$d\mathtt{y}(\xi) = \left.\frac{\partial \mathtt{y}}{\partial \mathtt{x}}\right|_{\mathtt{x}(p)} \circ d\mathtt{x}(\xi)$$

and in finite dimensions the classical notation

$$y^i = y^i(x^1, ..., x^n)$$
$$dy^i(\xi) = \frac{\partial y^i}{\partial x^k} dx^k(\xi)$$
$$or$$
$$w^i = \frac{\partial y^i}{\partial x^k} v^k$$

This classical notation may not be logically precise but it is easy to read and understand. In any case one could perhaps write

$$y = \mathtt{y} \circ \mathtt{x}^{-1}(x)$$
$$w = \left.\frac{\partial \mathtt{y}}{\partial \mathtt{x}}\right|_{\mathtt{x}(p)} v.$$

Exercise 4.5 *If M is actually equal to an open subset U of a Banach space M then we defined TU to be $U \times \mathsf{M}$. How should this be reconciled with the definitions of this chapter? Show that once this reconciliation is in force the tangent bundle chart map $T\mathtt{x}$ really is the tangent map of the coordinate map $\mathtt{x}$.*

4.5.2 The Cotangent Bundle

For each $p \in M$, T_pM has a dual space T_p^*M called the cotangent space at p.

Definition 4.14 *Define the **cotangent bundle** of a manifold M to be the set*

$$T^*M := \bigsqcup_{p \in M} T_p^*M$$

and define the map $\pi_M : \bigsqcup_{p \in M} T_p^*M \to M$ to be the obvious projection taking elements in each space T_p^*M to the corresponding point p.

Let $\{U_\alpha, \mathtt{x}_\alpha\}_{\alpha\in A}$ be an atlas of admissible charts on M. Now endow T^*M with the smooth structure given by the charts

$$T^*\mathtt{x}_\alpha : T^*U_\alpha = \pi_M^{-1}(U_\alpha) \to T^*\mathtt{x}_\alpha(T^*U_\alpha) \cong \mathtt{x}_\alpha(U_\alpha) \times \mathsf{M}^*$$

where the map $T^*\mathtt{x}_\alpha$ restricts to $T_p^*M \subset T^*U_\alpha$ as the **contragradient** of $T\mathtt{x}_\alpha$:

$$T^*\mathtt{x}_\alpha := (T\mathtt{x}_\alpha^{-1})^* : T_p^*M \to (\{\mathtt{x}_\alpha(p)\} \times \mathsf{M})^* = \mathsf{M}^*$$

If M is a smooth n dimensional manifold and $x^1, ..., x^n$ are coordinate functions coming from some chart on M then the "differentials" $dx^1\big|_p, ..., dx^n\big|_p$ are a basis of T_p^*M dual to $\frac{\partial}{\partial x^1}\big|_p, ..., \frac{\partial}{\partial x^n}\big|_p$. If $\theta \in T^*U_\alpha$ then we can write

$$\theta = \sum \xi_i(\theta)\; dx^i\big|_{\pi_M(\theta)}$$

for some numbers $\xi_i(\theta)$ depending on θ. In fact, we have

$$\theta(\frac{\partial}{\partial x^i}\bigg|_{\pi(\theta)}) = \sum \xi_j(\theta) dx^j(\frac{\partial}{\partial x^i}\bigg|_{\pi(\theta)}) = \sum_j \xi_j(\theta)\delta_i^j = \xi_i(\theta).$$

Thus we see that

$$\xi_i(\theta) = \theta(\frac{\partial}{\partial x^i}\bigg|_{\pi(\theta)}).$$

So if $U_\alpha, \mathtt{x}_\alpha = (x^1, ..., x^n)$ is a chart on an n-manifold M, then the natural chart $(TU_\alpha, T^*\mathtt{x}_\alpha)$ defined above is given by

$$\theta \mapsto (x^1 \circ \pi_M(\theta), ..., x^n \circ \pi_M(\theta), \xi_1(\theta), ..., \xi_n(\theta))$$

and abbreviated to $(x^1, ..., x^n, \xi_1, ..., \xi_n)$.

Suppose that $(x^1, ..., x^n, \xi_1, ..., \xi_n)$ and $(\overline{x}^1, ..., \overline{x}^n, \overline{\xi}_1, ..., \overline{\xi}_n)$ are two such charts constructed in this way from two charts on $(U_\alpha, \mathtt{x}_\alpha)$ and $(U_\beta, \mathtt{x}_\beta)$ respectively with $U_\alpha \cap U_\beta \neq \emptyset$. We are writing $\mathtt{x}_\alpha = (x^1, ..., x^n)$ and $\mathtt{x}_\beta = (\overline{x}^1, ..., \overline{x}^n)$ for notational convenience. Then $T^*U_\alpha \cap T^*U_\beta \neq \emptyset$ and on the overlap we have

$$T^*\mathtt{x}_\beta \circ T^*\mathtt{x}_\alpha^{-1} : \mathtt{x}_\alpha(U_\alpha \cap U_\beta) \times \mathsf{M}^* \to \mathtt{x}_\beta(U_\alpha \cap U_\beta) \times \mathsf{M}^*.$$

Notation 4.2 *With $\mathtt{x}_{\beta\alpha} = \mathtt{x}_\beta \circ \mathtt{x}_\alpha^{-1}$ the contragradient of $D\mathtt{x}_{\beta\alpha}$ at $x \in \mathtt{x}_\alpha(U_\alpha \cap U_\beta)$ is a map $D^*\mathtt{x}_{\beta\alpha}(x) : \mathsf{M}^* \to \mathsf{M}^*$ defined by*

$$D^*\mathtt{x}_{\beta\alpha}(x) \cdot v = \left(D\mathtt{x}_{\beta\alpha}^{-1}(\mathtt{x}_{\beta\alpha}(x))\right)^t \cdot v$$

With this notation we can write coordinate change maps as $(\mathtt{x}_{\beta\alpha}, D^*\mathtt{x}_{\beta\alpha})$ or to be exact

$$\left(T^*\mathtt{x}_\beta \circ T^*\mathtt{x}_\alpha^{-1}\right)(x, v) = \left(\mathtt{x}_{\beta\alpha}(x),\, D^*\mathtt{x}_{\beta\alpha}(x) \cdot v\right).$$

In case $\mathsf{M} = \mathbb{R}^n$ we write $pr_i \circ \mathtt{x}_{\beta\alpha} = \mathtt{x}^i_{\beta\alpha}$ and then

$$\overline{x}^i \circ \pi(\theta) = \mathtt{x}^i_{\beta\alpha}(x^1 \circ \pi(\theta), ..., x^n \circ \pi(\theta))$$
$$\overline{\xi}_i(\theta) = \sum_{k=1}^{n} (D\mathtt{x}^{-1}_{\beta\alpha})^k_i (x^1 \circ \pi(\theta), ..., x^n \circ \pi(\theta)) \cdot \xi_k(\theta)$$

This pedantic mess should be abbreviated. So let us write $x^i \circ \pi$ as just x^i and suppress the argument θ. Thus we may write

$$y = \mathtt{y} \circ \mathtt{x}^{-1}(x)$$
$$\bar{\xi} = \left.\frac{\partial \mathtt{x}}{\partial \mathtt{y}}\right|_{\mathtt{x}(p)} \xi$$

which may be further written in a classical style:

$$y^i = y^i(x^1, ..., x^n)$$
$$\overline{\xi}_i = \xi_k \frac{\partial x^k}{\partial y^i}.$$

4.6 Important Special Situations.

Let V be either a finite dimensional vector space or a Banach space. If a manifold under consideration is an open subset U of a the vector space V then the tangent space at any $x \in \mathsf{V}$ is canonically isomorphic with V itself. This was clear when we defined the tangent space at x as $\{x\} \times \mathsf{V}$ (by we now have several different but equivalent definitions of $T_x\mathsf{V}$). Then the identifying map is just $v \mapsto (x, v)$. Now one may convince oneself that the new more abstract definition of T_xU is essentially the same thing but we will describe the canonical map in another way: Let $v \in \mathsf{V}$ and define a curve $c_v : \mathbb{R} \to U \subset \mathsf{V}$ by $c_v(t) = x + tv$. Then $T_0c_v \cdot 1 = \dot{c}_v(0) \in T_xU$. The map $v \mapsto \dot{c}_v(0)$ is then our identifying map. Depending on how one defines the tangent bundle TU, the equality $TU = U \times \mathsf{V}$ is either a definition or a natural identification. The fact that there are these various identifications and that some things have several "equivalent" definitions is somewhat of a nuisance and can be confusing to the novice (occasionally to the expert also). The important thing is to think things through carefully, draw a few pictures, and most of all, try to think geometrically. One thing to notice is that for a vector spaces the derivative rather than the tangent map is all one needs in most cases. For example, if one wants to study a map $f : U \subset \mathsf{V} \to \mathsf{W}$ and if $v_p = (p, v) \in T_pU$ then $T_pf \cdot v_p = T_pf \cdot (p, v) = (p, Df|_p \cdot v)$. In other words the tangent map is $(p, v) \mapsto (p, Df|_p \cdot v)$ and so one might as well just think about the ordinary derivative $Df|_p$. In fact, in the case of a vector space some authors actually identify T_pf with $Df|_p$ as they also identify T_pU with V. There is usually no harm in this and it actually streamlines the calculations a bit.

The identifications of $T_x\mathsf{V}$ with V and TU with $U \times \mathsf{V}$ will be called canonical identifications. Whenever we come across a sufficiently natural isomorphism, then that isomorphism could be used to identify the two spaces. We will see cases where there are several different natural isomorphisms which compete for use as identifications. This arises when a space has more than one structure.

An example of this situation is the case of a manifold of matrices such as $GL(n,\mathbb{R})$. Here $GL(n,\mathbb{R})$ is actually an open subset of the set of all $n \times n$-matrices $\mathbb{M}_{n\times n}(\mathbb{R})$. The latter is a vector space so all our comments above apply so that we can think of $\mathbb{M}_{n\times n}(\mathbb{R})$ as any of the tangent spaces $T_A GL(n,\mathbb{R})$. Another interesting fact is that many important maps such as $c_Q : A \mapsto Q^t AQ$ are actually linear so with the identifications $T_A\mathrm{GL}(n,\mathbb{R}) = \mathbb{M}_{n\times n}(\mathbb{R})$ we have

$$T_A c_Q \text{ “}=\text{” } Dc_Q|_A = c_Q : \mathbb{M}_{n\times n}(\mathbb{R}) \to \mathbb{M}_{n\times n}(\mathbb{R}).$$

Whenever we come across a sufficiently natural isomorphism, then that isomorphism could be used to identify the two spaces.

Definition 4.15 (Partial Tangential) *Suppose that $f : M_1 \times M_2 \to N$ is a smooth map. We can define the partial maps as before and thus define* ***partial tangent maps****:*

$$(\partial_1 f)(x,y) : T_x M_1 \to T_{f(x,y)}N$$
$$(\partial_2 f)(x,y) : T_y M_2 \to T_{f(x,y)}N$$

Next we introduce another natural identification. It is obvious that a curve $c : I \to M_1 \times M_2$ is equivalent to a pair of curves

$$c_1 : I \to M_1$$
$$c_2 : I \to M_2$$

The infinitesimal version of this fact gives rise to a natural identification

$$T_{(x,y)}(M_1 \times M_2) \cong T_x M_1 \times T_y M_2$$

This is perhaps easiest to see if we view tangent vectors as equivalence classes of curves (tangency classes). Then if $c(t) = (c_1(t), c_2(t))$ and $c(0) = (x,y)$ then the map $[c] \mapsto ([c_1],[c_2])$ is a natural isomorphism which we use to simply identify $[c] \in T_{(x,y)}(M_1 \times M_2)$ with $([c_1],[c_2]) \in T_x M_1 \times T_y M_2$. For another view, consider the insertion maps $\iota_x : y \mapsto (x,y)$ and $\iota^y : x \mapsto (x,y)$.

$$(M_1,x) \underset{\iota^y}{\overset{pr_1}{\leftrightarrows}} (M_1 \times M_2,(x,y)) \underset{\iota_x}{\overset{pr_2}{\rightleftarrows}} (M_2,x)$$

$$T_x M_1 \underset{T_x\iota^y}{\overset{T_{(x,y)}pr_1}{\leftrightarrows}} T_{(x,y)}M_1 \times M_2 \underset{T_y\iota_x}{\overset{T_{(x,y)}pr_2}{\rightleftarrows}} T_x M_2$$

We have linear monomorphisms $T\iota^y(x) : T_x M_1 \to T_{(x,y)}(M_1 \times M_2)$ and $T\iota_x(y) : T_y M_2 \to T_{(x,y)}(M_1 \times M_2)$. Let us temporarily denote the isomorphic images of T_xM_1 and T_yM_2 in $T_{(x,y)}(M_1 \times M_2)$ under these two maps by the symbols $(T_xM)_1$ and $(T_yM)_2$. We then have the internal direct sum decomposition $(T_xM)_1 \oplus (T_yM)_2 = T_{(x,y)}(M_1 \times M_2)$ and the isomorphism

$$T\iota^y \times T\iota_x : T_xM_1 \times T_yM_2 \to (T_xM)_1 \oplus (T_yM)_2 = T_{(x,y)}(M_1 \times M_2).$$

The inverse of this isomorphism is

$$T_{(x,y)}pr_1 \times T_{(x,y)}pr_2 : T_{(x,y)}(M_1 \times M_2) \to T_xM_1 \times T_yM_2$$

which is then taken as an identification and, in fact, this is none other than the map $[c] \mapsto ([c_1], [c_2])$. Let us say a bit about the naturalness of the identification of $[c] \in T_{(x,y)}(M_1 \times M_2)$ with $([c_1], [c_2]) \in T_xM_1 \times T_yM_2$. In the smooth category there is a product operation. The essential point is that for any two manifolds M_1 and M_2 the manifold $M_1 \times M_2$ together with the two projection maps serves as the product in the technical sense that for any smooth maps $f : N \longrightarrow M_1$ and $g : N \longrightarrow M_2$ we always have the unique map $f \times g$ which makes the following diagram commute:

$$\begin{tikzcd}
 & N \arrow[dl, "f"'] \arrow[d, "f\times g"] \arrow[dr, "g"] & \\
M_1 & M_1 \times M_2 \arrow[l, "pr_1"'] \arrow[r, "pr_2"] & M_2
\end{tikzcd}$$

Now for a point $x \in N$ write $p = f(x)$ and $p = g(x)$. On the tangent level we have

$$\begin{tikzcd}
 & T_xN \arrow[dl, "T_xf"'] \arrow[d, "T_x(f\times g)"] \arrow[dr, "T_xg"] & \\
T_pM_1 & T_{(p,q)}(M_1 \times M_2) \arrow[l, "T_{(p,q)}pr_1"'] \arrow[r, "T_{(p,q)}pr_2"] & T_qM_2
\end{tikzcd}$$

which is a diagram in the vector space category. In the category of vector spaces the product of T_pM_1 and T_pM_2 is $T_pM_1 \times T_pM_2$ (outer direct sum) together with the projections onto the two factors. It is then quite reassuring to notice that under the identification introduced above this diagram corresponds to

$$\begin{tikzcd}
 & T_xN \arrow[dl, "T_xf"'] \arrow[d, "T_xf\times T_xg"] \arrow[dr, "T_xg"] & \\
T_pM_1 & T_pM_1 \times T_qM_2 \arrow[l, "Pr_1"'] \arrow[r, "Pr_2"] & T_qM_2
\end{tikzcd}$$

Notice that we have $f \circ \iota_y = f_{,y}$ and $f \circ \iota_x = f_x$.

Looking again at the definition of partial tangential one arrives at

Lemma 4.4 (partials lemma) *For a map $f : M_1 \times M_2 \to N$ we have*

$$T_{(x,y)}f \cdot (v, w) = (\partial_1 f)(x, y) \cdot v + (\partial_2 f)(x, y) \cdot w.$$

where we have used the aforementioned identification $T_{(x,y)}(M_1 \times M_2) = T_x M_1 \times T_y M_2$

Proving this last lemma is much easier and more instructive than reading the proof so we leave it to the reader in good conscience.

The following diagram commutes:

$$\begin{tikzcd}
T_{(x,y)}(M_1 \times M_2) \arrow[d, "T_{(x,y)}pr_1 \times T_{(x,y)}pr_2"'] \arrow[dr] & \\
T_x M_1 \times T_y M_2 \arrow[r] & T_{f(x,y)}N
\end{tikzcd}$$

Essentially, both diagonal maps refer to $T_{(x,y)}f$ because of our identification.

4.7 Vector fields and Differential 1-forms

Definition 4.16 *A smooth* ***vector field*** *is a smooth map $X : M \to TM$ such that $X(p) \in T_pM$ for all $p \in M$. In other words, a vector field on M is a smooth section of the tangent bundle $\tau_M : TM \to M$. We often write $X(p) = X_p$.*

The map X being smooth is equivalent to the requirement that the function $Xf : M \to \mathbb{R}$ given by $p \mapsto X_p f$ is smooth whenever $f : M \to \mathbb{R}$ is smooth.

If $\mathtt{x}, U$ is a chart and X a vector field defined on U then the local representation of X is $x \mapsto (x, X_U(x))$ where the **local representative** (or **principal part**) X_U is given by projecting $T\mathtt{x} \circ X \circ \mathtt{x}^{-1}$ onto the second factor in $T\mathsf{E} = \mathsf{E} \times \mathsf{E}$:

$$\begin{aligned} x \mapsto \mathtt{x}^{-1}(x) = p \mapsto X(p) \mapsto T\mathtt{x} \cdot X(p) \\ = (\mathtt{x}(p), X_U(\mathtt{x}(p))) = (x, X_U(x)) \mapsto X_U(x) \end{aligned}$$

In finite dimensions one can write $X_U(x) = (v_1(x), ..., v_n(x))$.

Notation 4.3 *The set of all smooth vector fields on M is denoted by $\Gamma(M, TM)$ or by the common notation $\mathfrak{X}(M)$. Smooth vector fields may at times be defined only on some open set so we also have the notation $\mathfrak{X}(U) = \mathfrak{X}_M(U)$ for these fields.*

A smooth (resp C^r) section of the cotangent bundle is called a smooth (resp C^r) **covector field** or also a smooth (resp C^r) **1-form** . The set of all C^r 1-forms is denoted by $\mathfrak{X}^{r*}(M)$ and the smooth 1-forms are denoted by $\mathfrak{X}^*(M)$. For any open set $U \subset M$, the set $C^\infty(U)$ of smooth functions defined on U is an algebra under the obvious linear structure $(af + bg)(p) := af(p) + bg(p)$ and obvious multiplication; $(fg)(p) := f(p)g(p)$. When we think of $C^\infty(U)$ in this way we sometimes denote it by $\mathcal{C}^\infty(U)$.

Remark 4.4 *The assignment* $U \mapsto \mathfrak{X}_M(U)$ *is a presheaf of modules over* $\mathcal{C}^\infty$ *and* $\mathfrak{X}^*(M)$ *is a module over the ring of functions* $C^\infty(M)$ *and a similar statement holds for the* C^r *case. (Jump forward to sections 6.3 and 6.6 for definitions.)*

Definition 4.17 *Let* $f : M \to \mathbb{R}$ *be a* C^r *function with* $r \geq 1$*. The map* $df : M \to T^*M$ *defined by* $p \mapsto df(p)$ *where* $df(p)$ *is the differential at* p *as defined in 4.4. is a 1-form called the differential of* f*.*

If $U, \mathtt{x}$ is a chart on M then we also have the following familiar looking formula in the finite dimensional case

$$df = \sum \frac{\partial f}{\partial x^i} dx^i$$

which is interpreted to mean that at each $p \in U_\alpha$ we have

$$df(p) = \sum \left. \frac{\partial f}{\partial x^i} \right|_p dx^i\big|_p .$$

In general, if we have a chart $U, \mathtt{x}$ on a possibly infinite dimensional manifold then we may write

$$df = \frac{\partial f}{\partial \mathtt{x}} d\mathtt{x}$$

We have really seen this before. All that has happened new is that p is allowed to vary so we have a field.

There is a slightly different way to view a 1-form that is often useful. Namely, we may think of $\alpha \in \mathfrak{X}^*(M)$ as a map $TM \to \mathbb{R}$ given simply by $\alpha(v) = \alpha(p)(v) = \alpha_p(v)$ whenever $v \in T_pM \subset TM$.

If $\phi : M \to N$ is a C^∞ map and $f : N \to \mathbb{R}$ a C^∞ function we define the **pullback** of f by ϕ as

$$\phi^* f = f \circ \phi$$

and the **pullback** of a 1-form $\alpha \in \mathfrak{X}^*(N)$ by $\phi^*\alpha = \alpha \circ T\phi$. To get a clearer picture of what is going on we could view things at a point and then we have $\phi^*\alpha\big|_p \cdot v = \alpha\big|_{\phi(p)} \cdot (T_p\phi \cdot v)$.

Next we describe the local expression for the pull-back of a 1-form. Let $U, \mathtt{x}$ be a chart on M and $V, \mathtt{y}$ be a coordinate chart on N with $\phi(U) \subset V$. A typical 1-form has a local expression on V of the form $\alpha = \sum a_i dy^i$ for $a_i \in C^\infty(V)$. The local expression for $\phi^*\alpha$ on U is $\phi^*\alpha = \sum a_i \circ \phi \, d\left(y^i \circ \phi\right) = \sum a_i \circ \phi \frac{\partial\left(y^i \circ \phi\right)}{\partial x^i} dx^i$.

The pull-back of a function or 1-form is defined whether $\phi : M \to N$ happens to be a diffeomorphism or not. On the other hand, when we define the pull-back of a vector field in a later section we will only be able to do this if the map that we are using is a diffeomorphism. Push-forward is another matter.

Definition 4.18 *Let* $\phi : M \to N$ *be a* C^∞ *diffeomorphism with* $r \geq 1$*. The* ***push-forward*** *of a function* $f \in C^\infty(M)$ *is denoted* $\phi_* f$ *and defined by* $\phi_* f(p) := f(\phi^{-1}(p))$*. We can also define the push-forward of a 1-form as* $\phi_*\alpha = \alpha \circ T\phi^{-1}$*.*

Exercise 4.6 *Find the local expression for $\phi_* f$. Explain why we need ϕ to be a diffeomorphism.*

It should be clear that the pull-back is the more natural of the two when it comes to forms and functions but in the case of vector fields this is not true.

Lemma 4.5 *The differential is natural with respect to pullback. In other words, if $\phi : N \to M$ is a C^∞ map and $f : M \to \mathbb{R}$ a C^∞ function with $r \geq 1$ then $d(\phi^* f) = \phi^* df$. Consequently, the differential is also natural with respect to restrictions*

Proof. Let v be a curve such that $\dot{c}(0) = v$. Then

$$\begin{aligned} d(\phi^* f)(v) &= \left.\frac{d}{dt}\right|_0 \phi^* f(c(t)) = \left.\frac{d}{dt}\right|_0 f(\phi(c(t))) \\ &= df \left.\frac{d}{dt}\right|_0 \phi(c(t)) = df(T\phi \cdot v) \end{aligned}$$

As for the second statement (besides being obvious from local coordinate expressions) notice that if U is open in M and $\iota : U \hookrightarrow M$ is the inclusion map (i.e. identity map id_M restricted to U) then $f|_U = \iota^* f$ and $df|_U = \iota^* df$ so the statement about restrictions is just a special case. ■

4.8 Moving frames

If $U, \mathtt{x}$ is a chart on a smooth n-manifold then writing $\mathtt{x} = (x^1, ..., x^n)$ we have vector fields defined on U by

$$\frac{\partial}{\partial x^i} : p \mapsto \left.\frac{\partial}{\partial x^i}\right|_p$$

such that the $\frac{\partial}{\partial x^i}$ form a basis at each tangent space at point in U. The set of fields $\frac{\partial}{\partial x^1}, ..., \frac{\partial}{\partial x^n}$ is called a **holonomic moving frame** over U or, more commonly, a **coordinate frame**. If X is a vector field defined on some set including this local chart domain U then for some smooth functions X^i defined on U we have

$$X(p) = \sum X^i(p) \left.\frac{\partial}{\partial x^i}\right|_p$$

or in other words

$$X|_U = \sum X^i \frac{\partial}{\partial x^i}.$$

Notice also that $dx^i : p \mapsto dx^i\big|_p$ defines a field of covectors such that $dx^1\big|_p, ..., dx^n|_p$ forms a basis of $T_p^* M$ for each $p \in U$. These covector fields form what is called a **holonomic**[1] **coframe** or **coframe field** over U. In fact, the functions X^i are given by $X^i = dx^i(X) : p \mapsto dx^i\big|_p (X_p)$ and so we write

$$X|_U = \sum dx^i(X) \frac{\partial}{\partial x^i}.$$

Notation 4.4 *We will not usually bother to distinguish X from its restrictions and so we just write* $X = \sum X^i \frac{\partial}{\partial x^i}$ *or using the Einstein summation convention* $X = X^i \frac{\partial}{\partial x^i}$.

It is important to realize that we can also get a family of vector (resp. covector) fields that are linearly independent at each point in their mutual domain and yet are not necessarily of the form $\frac{\partial}{\partial x^i}$ (resp. dx^i) for any coordinate chart:

Definition 4.19 *Let* $E_1, E_2, ..., E_n$ *be smooth vector fields defined on some open subset* U *of a smooth* n*-manifold* M. *If* $E_1(p), E_2(p), ..., E_n(p)$ *form a basis for* T_pM *for each* $p \in U$ *then we say that* $E_1, E_2, ..., E_n$ *is a (non-holonomic)* ***moving frame*** *or a* ***frame field*** *over* U.

If $E_1, E_2, ..., E_n$ is moving frame over $U \subset M$ and X is a vector field defined on U then we may write

$$X = \sum X^i E_i \text{ on } U$$

for some functions X^i defined on U. Taking the dual basis in T_p^*M for each $p \in U$ we get a (non-holonomic) moving coframe field $\theta^1, ..., \theta^n$ and then $X^i = \theta^i(X)$ so

$$X = \sum \theta^i(X) E_i \text{ on } U.$$

Definition 4.20 *A* ***derivation*** *on* $\mathcal{C}^\infty(U)$ *is a linear map* $\mathcal{D} : \mathcal{C}^\infty(U) \to \mathcal{C}^\infty(U)$ *such that*

$$\mathcal{D}(fg) = \mathcal{D}(f)g + f\mathcal{D}(g).$$

A C^∞ vector field on U may be considered as a derivation on $\mathfrak{X}(U)$ where we view $\mathfrak{X}(U)$ as a module[2] over the ring of smooth functions $\mathcal{C}^\infty(U)$.

Definition 4.21 *To a vector field X on U we associate the map* $\mathcal{L}_X : C^\infty(U) \to \mathfrak{X}_M(U)$ *defined by*

$$(\mathcal{L}_X\ f)(p) := X_p \cdot f$$

and called the ***Lie derivative*** *on functions.*

It is easy to see, based on the Leibniz rule established for vectors X_p in individual tangent spaces, that $\mathcal{L}_X$ is a derivation on $\mathcal{C}^\infty(U)$. We also define the symbolism "Xf", where $X \in \mathfrak{X}(U)$, to be an abbreviation for the function $\mathcal{L}_X\ f$.

[1]The word holonomic comes from mechanics and just means that the frame field derives from a chart. A related fact is that $[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}] = 0$.

[2]See section 6.7.

Remark 4.5 *We often leave out parentheses and just write $Xf(p)$ instead of the more careful $(Xf)(p)$.*

In summary, we have the **derivation law** (Leibniz rule) for vector fields:

$$X(fg) = fXg + gXf.$$

or in other notation $\mathcal{L}_X\,(fg) = f\mathcal{L}_X\,g + g\mathcal{L}_X\,f$.

Let us now consider an important special situation. If $M \times N$ is a product manifold and $(U, \mathtt{x})$ is a chart on M and $(V, \mathtt{y})$ is a chart on N then we have a chart $(U \times V, \mathtt{x} \times \mathtt{y})$ on $M \times N$ where the individual coordinate functions are $x^1 \circ pr_1, ..., x^m \circ pr_1$ $y^1 \circ pr_2, ..., y^n \circ pr_2$ which we temporarily denote by $\widetilde{x}^1, ..., \widetilde{x}^m, \widetilde{y}^1, ..., \widetilde{y}^n$. Now what is the relation between the coordinate frame fields $\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^i}$ and $\frac{\partial}{\partial \widetilde{x}^i}, \frac{\partial}{\partial \widetilde{y}^i}$? Well, the latter set of $n+m$ vector fields is certainly a linearly independent set at each point $(p, q) \in U \times V$. The crucial relations are $\frac{\partial}{\partial \widetilde{x}^i} f = \frac{\partial}{\partial x^i}\,(f \circ pr_1)$ and $\frac{\partial}{\partial \widetilde{y}^i} = \frac{\partial}{\partial y^i}\,(f \circ pr_2)$.

Exercise 4.7 *Show that $\left.\frac{\partial}{\partial \widetilde{x}^i}\right|_{(p,q)} = Tpr_1 \left.\frac{\partial}{\partial x^i}\right|_p$ for all q and that $\frac{\partial}{\partial \widetilde{y}^i} = Tpr_2 \left.\frac{\partial}{\partial y^i}\right|_q$.*

Remark 4.6 *In some circumstances it is safe to abuse notation and denote $x^i \circ pr_1$ by x^i and $y^i \circ pr_2$ by y^i. Of course we then are denoting $\frac{\partial}{\partial \widetilde{x}^i}$ by $\frac{\partial}{\partial x^i}$ and so on.*

4.9 Partitions of Unity

A smooth partition of unity is a technical tool that is used quite often in connection with constructing tensor fields, connections, metrics and other objects, out of local data. We will not meet tensor fields for a while and the reader may wish to postpone a detailed reading of the proofs in this section until we come to our first use of partitions of unity and/or the so called "bump functions" and "cut-off functions". Partitions of unity are also used in proving the existence of immersions and embeddings; topics we will also touch on later.

It is often the case that we are able to define some object or operation locally and we wish to somehow "glue together" the local data to form a globally defined object. The main and possibly only tool for doing this is the partition of unity. For differential geometry, what we actually need it is a *smooth* partition of unity. Banach manifolds do not always support smooth partitions of unity. We will state and prove the existence of partitions of unity for finite dimensional manifolds and refer the reader to [A,B,R] for the situation in Banach spaces. A bump function is basically a smooth function with support inside some prescribed open set O and that is nonzero on some prescribed set inside O. Notice that this would not in general be possible for a complex analytic function. A slightly stronger requirement is the existence of cut-off functions. A cut-off function is a bump function that is equal to unity in a prescribed region around a given point:

Definition 4.22 *A* ***spherical cut-off function*** *of class C^r for the nested pair of balls $B(r,p) \subset B(R,p)$ $(R > r)$is a C^r function $\beta : M \to \mathbb{R}$ such that $\beta|_{\overline{B(r,p)}} \equiv 1$ and $\beta|_{M \backslash B(R,p)} \equiv 0$.*

More generally we have the following

Definition 4.23 *Let K be a closed subset of M contained in an open subset $U \subset M$. A* ***cut-off function*** *of class C^r for a nested pair $K \subset U$ is a C^r function $\beta : M \to \mathbb{R}$ such that $\beta|_K \equiv 1$ and $\beta|_{M \backslash U} \equiv 0$.*

Definition 4.24 *A manifold M is said to* ***admit cut-off functions*** *if given any point $p \in M$ and any open neighborhood U of p, there is another neighborhood V of p such that $\overline{V} \subset U$ and a cut-off function $\beta_{\overline{V},U}$ for the nested pair $\overline{V} \subset U$. A manifold M is said to* ***admit spherical cut-off functions*** *if given any point $p \in M$ and any open ball U of p, there is another open ball V of p such that $\overline{V} \subset U$ and a cut-off function $\beta_{\overline{V},U}$ for the nested pair $\overline{V} \subset U$*

For most purposes, the existence of spherical cut off functions will be sufficient and in the infinite dimensional case it might be all we can get.

Definition 4.25 *Let E be a Banach space and suppose that the norm on E is smooth (resp.C^r) on the open set $\mathsf{E} \backslash \{0\}$. The we say that E is a* ***smooth*** *(resp. C^r)* ***Banach space****.*

Proposition 4.3 *If E is a* smooth *(resp. C^r) Banach space and $B_r \subset B_R$ nested open balls then there is a smooth (resp. C^r) cut-off function β for the pair $B_r \subset B_R$. Thus β is defined on all of E , identically equal to 1 on the closure $\overline{B_r}$ and zero outside of B_R.*

Proof. We assume with out loss of generality that the balls are centered at the origin $0 \in \mathsf{E}$. Let

$$\phi_1(s) = \frac{\int_{-\infty}^{s} g(t)dt}{\int_{-\infty}^{\infty} g(t)dt}$$

where

$$g(t) = \begin{cases} \exp(-1/(1-|t|^2) & \text{if} \quad |t| < 1 \\ \\ 0 & \text{otherwise} \end{cases}.$$

This is a smooth function and is zero if $s < -1$ and 1 if $s > 1$ (verify). Now let $\beta(x) = g(2 - |x|)$. Check that this does the job using the fact that $x \mapsto |x|$ is assumed to be smooth (resp. C^r). ■

Corollary 4.1 *If a manifold M is modeled on a smooth (resp. C^r) Banach space M (for example, if M is a finite dimensional smooth manifold) then for every $\alpha_p \in T^*M$, there is a (global) smooth (resp. C^r) function f such that $Df|_p = \alpha_p$.*

Proof. Let $x_0 = \psi(p) \in \mathsf{M}$ for some chart ψ, U. Then the local representative $\bar{\alpha}_{x_0} = (\psi^{-1})^* \alpha_p$ can be considered a linear function on M since we have the canonical identification $\mathsf{M} \cong \{x_0\} \times \mathsf{M} = \mathsf{T}_{x_0}\mathsf{M}$. Thus we can define

$$\varphi(x) = \begin{cases} \beta(x)\bar{\alpha}_{x_0}(x) & \text{for} \quad x \in B_R(x_0) \\ 0 & \quad\text{otherwise} \end{cases}$$

and now making sure that R is small enough that $B_R(x_0) \subset \psi(U)$ we can transfer this function back to M via ψ^{-1} and extend to zero outside of U get f. Now the differential of φ at x_0 is $\bar{\alpha}_{x_0}$ and so we have for $v \in T_pM$

$$\begin{aligned} df(p) \cdot v &= d(\psi^* \varphi)(p) \cdot v \\ &= (\psi^* d\varphi)(p) v \\ & d\varphi(T_p\psi \cdot v) \\ &= \bar{\alpha}_{x_0}(T_p\psi \cdot v) = (\psi^{-1})^* \alpha_p(T_p\psi \cdot v) \\ &= \alpha_p(T\psi^{-1} T_p\psi \cdot v) = \alpha_p(v) \end{aligned}$$

so $df(p) = \alpha_p$ ∎

It is usually taken for granted that derivations on smooth functions are vector fields and that all C^∞ vector fields arise in this way. In fact, this not true in general. It is true however, for finite dimensional manifold. More generally, we have the following result:

Proposition 4.4 *The map from $\mathfrak{X}(M)$ to the vector space of derivations $\mathsf{Der}(M)$ given by $X \mapsto \mathcal{L}_X$ is a linear monomorphism if M is modeled on a C^∞ Banach space.*

Proof. The fact that the map is linear is straightforward. We just need to get the injectivity. For that, suppose $\mathcal{L}_X\, f = 0$ for all $f \in \mathcal{C}^\infty(M)$. Then $Df|_p\, X_p = 0$ for all $p \in M$. Thus by corollary 4.1 $\alpha_p(X_p) = 0$ for all $\alpha_p \in T_p^*M$. By the Hahn-Banach theorem this means that $X_p = 0$. Since p was arbitrary we concluded that $X = 0$. ∎

Another very useful result is the following:

Theorem 4.1 *Let $L : \mathfrak{X}(M) \to \mathcal{C}^\infty(M)$ be a $\mathcal{C}^\infty(M)-$linear function on vector fields. If M admits (spherical?)cut off functions then $L(X)(p)$ depends only on the germ of X at p.*
If M is finite dimensional then $L(X)(p)$ depends only on the value of X at p.

Proof. Suppose $X = 0$ in a neighborhood U and let $p \in U$ be arbitrary. Let O be a smaller open set containing p and with closure inside U. Then letting β be a function that is identically 1 on a neighborhood of p contained in O and identically zero outside of O then $(1 - \beta)X = X$. Thus we have

$$\begin{aligned} L(X)(p) &= L((1-\beta)X)(p) \\ &= (1 - \beta(p))L(X)(p) = 0 \times L(X)(p) \\ &= 0. \end{aligned}$$

Applying this to $X - Y$ we see that if two fields X and Y agree in an open set then $L(X) = L(Y)$ on the same open set. The result follows from this.

Now suppose that M is finite dimensional and suppose that $X(p) = 0$. Write $X = X^i \frac{\partial}{\partial x^i}$ in some chart domain containing p with smooth function X^i satisfying $X^i(p) = 0$. Letting β be as above we have

$$\beta^2 L(X) = \beta X^i L(\beta \frac{\partial}{\partial x^i})$$

which evaluated at p gives

$$L(X)(p) = 0$$

since $\beta(p) = 1$. Applying this to $X - Y$ we see that if two fields X and Y agree at p then $L(X)(p) = L(Y)(p)$. ■

Corollary 4.2 *If M is finite dimensional and $L : \mathfrak{X}(M) \to \mathcal{C}^\infty(M)$ is a $\mathcal{C}^\infty(M)-$linear function on vector fields then there exists an element $\alpha \in \mathfrak{X}^*(M)$ such that $\alpha(X) = L(X)$ for all $X \in \mathfrak{X}(M)$.*

Definition 4.26 *The support of a smooth function is the closure of the set in its domain where it takes on nonzero values. The support of a function f is denoted* supp(f).

For finite dimensional manifolds we have the following stronger result.

Lemma 4.6 (Existence of cut-off functions) *Let K be a compact subset of $\mathbb{R}^n$ and U an open set containing K. There exists a smooth function β on $\mathbb{R}^n$ that is identically equal to 1 on K, has compact support in U and $0 \le \beta \le 1$.*

Proof. Special case: Assume that $U = B(0, R)$ and $K = \overline{B}(0, r)$. In this case we may take

$$\phi(x) = \frac{\int_{|x|}^{R} g(t)dt}{\int_{r}^{R} g(t)dt}$$

where

$$g(t) = \begin{cases} e^{-(t-r)^{-1}} e^{-(t-R)^{-1}} & \text{if } 0 < t < R \\ 0 & \text{otherwise.} \end{cases}$$

This is the circular cut-off that always exists for smooth Banach spaces.

General case: Let $K \subset U$ be as in the hypotheses. Let $K_i \subset U_i$ be concentric balls as in the special case above but now with various choices of radii and such that $K \subset \cup K_i$. The U_i are chosen small enough that $U_i \subset U$. Let ϕ_i be the corresponding functions provided in the proof of the special case. By compactness there are only a finite number of pairs $K_i \subset U_i$ needed so assume that this reduction to a finite cover has been made. Examination of the following function will convince the reader that it is well defined and provides the needed cut-off function;

$$\beta(x) = 1 - \prod_i (1 - \phi_i(x)).$$

■

Definition 4.27 *A topological space is called* ***locally convex*** *if every point has a neighborhood with compact closure.*

Note that a finite dimensional differentiable manifold is always locally compact and we have agreed that a finite dimensional manifold should be assumed to be Hausdorff unless otherwise stated. The following lemma is sometimes helpful. It shows that we can arrange to have the open sets of a cover and a locally refinement of the cover to be indexed by the same set in a consistent way:

Lemma 4.7 *If X is a paracompact space and $\{U_i\}_{i\in I}$ is an open cover, then there exists a locally finite refinement $\{O_i\}_{i\in I}$ of $\{U_i\}_{i\in I}$ with $O_i \subset U_i$.*

Proof. Let $\{V_k\}_{i\in K}$ be a locally finite refinement of $\{U_i\}_{i\in I}$ with the index map $k \mapsto i(k)$. Let O_i be the union of all V_k such that $i(k) = k$. Notice that if an open set U intersects an infinite number of the O_i then it will meet an infinite number of the V_k. It follows that $\{O_i\}_{i\in I}$ is locally finite. ■

Theorem 4.2 *A second countable, locally compact Hausdorff space X is paracompact.*

Sketch of proof. If follows from the hypotheses that there exists a sequence of open sets $U_1, U_2, ...$that cover X and such that each U_i has compact closure $\overline{U_i}$. We start an inductive construction: Set $V_n = U_1 \cup U_2 \cup ... \cup U_n$ for each positive integer n. Notice that $\{V_n\}$ is a new cover of X and each V_n has compact closure. Now let $O_1 = V_1$. Since $\{V_n\}$ is an open cover and $\overline{O_1}$ is compact we have

$$\overline{O_1} \subset V_{i_1} \cup V_{i_2} \cup ... \cup V_{i_k}.$$

Next put $O_2 = V_{i_1} \cup V_{i_2} \cup ... \cup V_{i_k}$ and continue the process. Now we have that X is the countable union of these open sets $\{O_i\}$ and each O_{i-1} has compact closure in O_i. Now we define a sequence of compact sets; $K_i = \overline{O_i} \setminus O_{i-1}$.
Now if $\{W_\beta\}_{\beta\in B}$ is *any* open cover of X we can use those W_β that meet K_i to cover K_i and then reduce to a finite subcover since K_i is compact. We can arrange that this cover of K_i consists only of sets each of which is contained in one of the sets $W_\beta \cap O_{i+1}$ and disjoint from O_{i-1}. Do this for all K_i and collect all the resulting open sets into a countable cover for X. This is the desired locally finite refinement. ■

Definition 4.28 *A C^r partition of unity on a C^r manifold M is a collection $\{V_i, \rho_i\}$ where*
(i) $\{V_i\}$ is a locally finite cover of M;
(ii) each ρ_i is a C^r function with $\rho_i \geq 0$ and compact support contained in V_i;
(iii) for each $x \in M$ we have $\sum \rho_i(x) = 1$ (This sum is finite since $\{V_i\}$ is locally finite).
If the cover of M by chart domains $\{U_\alpha\}$ of some atlas $\mathcal{A} = \{U_\alpha, \mathtt{x}_\alpha\}$ for M has a partition of unity $\{V_i, \rho_i\}$ such that each V_i is contained in one of the chart domains $U_{\alpha(i)}$ (locally finite refinement), then we say that $\{V_i, \rho_i\}$ is

subordinate to $\mathcal{A}$. *We will say that a manifold* ***admits a smooth partition of unity*** *if every atlas has a subordinate smooth partition of unity.*

Smooth (C^r, $r > 0$) partitions of unity do not necessarily exist on a Banach space and less so for manifolds modeled on such Banach spaces. On the other hand, some Banach spaces do admit partitions of unity. It is a fact that all separable Hilbert spaces admit partitions of unity. For more information see [A,B,R]. We will content ourselves with showing that all finite dimensional manifolds admit smooth partitions of unity.

Notice that in theorem 4.2 we have proven a bit more than is part of the definition of paracompactness. Namely, the open sets of the refinement $V_i \subset U_{\beta(i)}$ have compact closure in $U_{\beta(i)}$. This is actually true for any paracompact space but we will not prove it here. Now for the existence of a smooth partition of unity we basically need the paracompactness but since we haven't proved the above statement about compact closures (shrink?ing lemma) we state the theorem in terms of second countability:

Theorem 4.3 *Every second countable finite dimensional* C^r *manifold admits a* C^r*−partition of unity.*

Let M be the manifold in question. We have seen that the hypotheses imply paracompactness and that we may choose our locally finite refinements to have the compact closure property mentioned above. Let $\mathcal{A} = \{U_i, \mathtt{x}_i\}$ be an atlas for M and let $\{W_i\}$ be a locally finite refinement of the cover $\{U_i\}$ with $\overline{W}_i \subset U_i$. By lemma 4.6 above there is a smooth cut-off function β_i with $\text{supp}(\beta_i) = \overline{W}_i$. For any $x \in M$ the following sum is finite and defines a smooth function:

$$\beta(x) = \sum_i \beta_i(x).$$

Now we normalize to get the functions that form the partition of unity:

$$\rho_i = \frac{\beta_i}{\beta}.$$

It is easy to see that $\rho_i \geq 0$, and $\sum \rho_i = 1$.

4.10 Problem Set

1. Find a concrete description of the tangent bundle for each of the following manifolds:
   (a) Projective space $P(\mathbb{R}^n)$
   (b) The Grassmann manifold $Gr(k, n)$

Chapter 5

Lie Groups I

5.1 Definitions and Examples

One approach to geometry is to view geometry as the study of invariance and symmetry. In our case we are interested in studying symmetries of differentiable manifolds, Riemannian manifolds, symplectic manifolds etc. Now the usual way to talk about symmetry in mathematics is by the use of the notion of a transformation group. We have already seen groups in action (chapter 2). The wonderful thing for us is that the groups that arise in the study of geometric symmetries are often themselves differentiable manifolds. Such "group manifolds" are called Lie groups.

In physics, Lie groups play a big role in classical mechanics in connection with physical symmetries and conservation laws (Noether's theorem). Also, infinite dimensional Lie groups arise in fluid mechanics. Within physics, perhaps the most celebrated role played by Lie groups is in particle physics and gauge theory. In mathematics, Lie groups play a prominent role in Harmonic analysis (generalized Fourier theory) and group representations as well as in virtually every branch of geometry including Riemannian geometry, Cartan geometry, algebraic geometry, Kähler geometry, symplectic geometry and also topology.

Definition 5.1 *A C^∞ differentiable manifold G is called a* ***Lie group*** *if it is a group (abstract group) such that the multiplication map $\mu : G \times G \to G$ and the inverse map $\nu : G \to G$ given by $\mu(g, h) = gh$ and $\nu(g) = g^{-1}$ are C^∞ functions.*

Example 5.1 $\mathbb{R}$ *is a one-dimensional (abelian Lie group) were the group multiplication is addition. Similarly, any vector space, e.g. $\mathbb{R}^n$, is a Lie group and the vector addition. The circle $S^1 = \{z \in \mathbb{C} : |z|^2 = 1\}$ is a 1-dimensional (abelian) Lie group under complex multiplication. It is also traditional to denote this group by $U(1)$.*

If G and H are Lie groups then so is the **product group** $G \times H$ where multiplication is $(g_1, h_1) \cdot (g_2, h_2) = (g_1 g_2, h_1 h_2)$. Also, if H is a subgroup of a Lie group G that is also a regular closed submanifold then H is a Lie group itself and we refer to H as a **closed (Lie) subgroup**. It is a nontrivial fact that an abstract subgroup of a Lie group that is also a closed subset is automatically a Lie subgroup (see 12.4).

Example 5.2 *The product group $S^1 \times S^1$ is called the 2-torus. More generally, the torus groups are defined by $T^n = \underbrace{S^1 \times \ldots \times S^1}_{n\text{-times}}$.*

Example 5.3 *S^1 embedded as $S^1 \times \{1\}$ in the torus $S^1 \times S^1$ is a closed subgroup.*

Now we have already introduced the idea group action. For convenience we give the definition in the context of Lie groups.

Definition 5.2 *A left action of a Lie group G on a manifold M is a smooth map $\lambda : G \times M \to M$ such that $\lambda(g_1, \lambda(g_2, m)) = \lambda(g_1 g_2, m))$ for all $g_1, g_2 \in G$. Define the partial map $\lambda_g : M \to M$ by $\lambda_g(m) = \lambda(g, m)$ and then the requirement is that $\widetilde{\lambda} : g \mapsto \lambda_g$ is a group homomorphism $G \to \mathrm{Diff}(M)$. We often write $\lambda(g, m)$ as $g \cdot m$.*

The definition of a right Lie group action should be obvious and so we omit it.

5.2 Linear Lie Groups

The group $Aut(\mathrm{V})$ of all linear automorphisms of a vector space V over a field $\mathbb{F}$ ($= \mathbb{R}$ or $\mathbb{C}$) is an open submanifold of the vector space $End(\mathrm{V})$ (the space of linear maps $\mathrm{V} \to \mathrm{V}$. This group is easily seen to be a Lie group and in that context it is usually denoted by $GL(\mathrm{V}, \mathbb{F})$ and referred to as the real **general linear group** in case $\mathbb{F} = \mathbb{R}$ or as the complex **general linear group** in case $\mathbb{F} = \mathbb{C}$. We may easily obtain a large family of Lie groups by considering closed subgroups of $GL(\mathrm{V}, \mathbb{F})$. These are the linear Lie groups and by choosing a basis we always have the associated group of matrices. These matrix groups are not always distinguished from the abstract linear group with which they are isomorphic.

Recall that the determinant of a linear transformation from a vector space to itself is define independent of any choice of basis.

Theorem 5.1 *Let V be an $n-$dimensional vector space over the field $\mathbb{F}$ which we take to be either $\mathbb{R}$ or $\mathbb{C}$. Let b be a nondegenerate $\mathbb{R}-$bilinear form on V. Each of the following groups is a closed (Lie) subgroup of $GL(\mathrm{V}, \mathbb{F})$:*

1. $SL(\mathrm{V}, \mathbb{F}) := \{A \in GL(\mathrm{V}, \mathbb{F}) : \det(A) = 1\}$

2. $Aut(\mathrm{V}, b, \mathbb{F}) := \{A \in GL(\mathrm{V}, \mathbb{F}) : b(Av, Aw) = b(v, w) \text{ for all } v, w \in \mathrm{V}\}$

3. $SAut(\mathrm{V}, b, \mathbb{F}) = Aut(\mathrm{V}, b, \mathbb{F}) \cap SL(\mathrm{V}, \mathbb{F})$

Proof. They are clearly closed subgroups. The fact that they are Lie subgroups follows from theorem 12.4 from the sequel. However, as we shall see most of the specific cases arising from various choices of b are easily proved to be Lie groups by a direct argument. ■

Depending on whether $\mathbb{F} = \mathbb{C}$ or $\mathbb{R}$ and on whether b is symmetric, Hermitian, or skewsymmetric the notation for the linear groups take on a special conventional forms. Also, when choosing a basis in order to represent the group in its matrix version it is usually the case that on uses a basis under which b takes on a canonical form. Let us look at the usual examples. Let $\dim \mathrm{V} = n$.

Example 5.4 *After choosing a basis the groups $GL(\mathrm{V}, \mathbb{F})$ and $SL(\mathrm{V}, \mathbb{F})$ become the matrix groups*

$$\begin{aligned} GL(n, \mathbb{F}) &:= \{A \in M_{n\times n}(\mathbb{F}) : \det A \neq 0\} \\ SL(n, \mathbb{F}) &:= \{A \in M_{n\times n}(\mathbb{F}) : \det A = 1\} \end{aligned}$$

Example 5.5 (The (semi) Orthogonal Groups) *Here we consider the case where $\mathbb{F} = \mathbb{R}$ and the bilinear form is symmetric and nondegenerate. Then (V, b) is a general scalar product space. In this case we write $Aut(\mathrm{V}, b, \mathbb{F})$ as $O(\mathrm{V}, b, \mathbb{R})$ and refer to it as the* ***semi-orthogonal*** *group associated to b. By Sylvester's law of inertia we may choose a basis so that b is represented by a diagonal matrix of the form*

$$\eta_{p,q} = \begin{bmatrix} 1 & 0 & \cdots & & \cdots & 0 \\ 0 & \ddots & 0 & & & \vdots \\ \vdots & 0 & 1 & \ddots & & \\ & & \ddots & -1 & 0 & \vdots \\ \vdots & & & 0 & \ddots & 0 \\ 0 & \cdots & & \cdots & 0 & -1 \end{bmatrix}$$

where there are p ones and q minus ones down the diagonal. The group of matrices arising from $O(\mathrm{V}, b, \mathbb{R})$ with such a choice of basis is denoted $O(p, q)$ and consists exactly of the real matrices A satisfying $A\eta_{p,q}A^t = \eta_{p,q}$. These are called the semi-orthogonal matrix groups. With orthonormal choice of basis the bilinear form (scalar product) is given as a canonical from on $\mathbb{R}^n$ $(p + q = n)$:

$$\langle x, y\rangle := \sum_{i=1}^{p} x^i y^i - \sum_{i=p+1}^{n} x^i y^i$$

and we have the alternative description

$$O(p, q) = \{A : \langle Ax, Ay\rangle = \langle x, y\rangle \text{ for all } x, y \in \mathbb{R}^n\}$$

We write $O(n, 0)$ as $O(n)$ are refer to it as the ***orthogonal*** *(matrix)* ***group****.*

Example 5.6 *In the example we consider the situation where $\mathbb{F} = \mathbb{C}$ and where b is complex linear in one argument (say the first one) and conjugate linear in the other. Thus b is a sesquilinear form and (V, b) is a complex scalar product space. In this case we write $Aut(\mathrm{V}, b, \mathbb{F})$ as $U(\mathrm{V}, b)$ and refer to it as the **semi-unitary group** associated to the sesquilinear form b. Again we may choose a basis for V such that b is represented by the canonical sesquilinear form on $\mathbb{C}^n$*

$$\langle x, y\rangle := \sum_{i=1}^{p} \bar{x}^i y^i - \sum_{i=p+1}^{p+q=n} \bar{x}^i y^i$$

We then obtain the semi-unitary matrix group

$$U(p,q) = \{A \in M_{n\times n}(\mathbb{C}) : \langle Ax, Ay\rangle = \langle x, y\rangle \text{ for all } x, y \in \mathbb{R}^n\}$$

*We write $U(n,0)$ as $U(n)$ are refer to it as the **unitary** (matrix) **group**.*

Example 5.7 (Symplectic groups) *We will describe both the real and the complex symplectic groups. Suppose now that b is a skewsymmetric $\mathbb{C}-$bilinear (resp. $\mathbb{R}-$bilinear) and V is a $2k$ dimensional complex (resp. real) vector space. The group $Aut(\mathrm{V}, b, \mathbb{C})$ (resp. $Aut(\mathrm{V}, b, \mathbb{R})$) is called the complex (resp. real) symplectic group and denoted by $Sp(\mathrm{V}, \mathbb{C})$ (resp. $Sp(\mathrm{V}, \mathbb{R})$). There exists a basis $\{f_i\}$ for V such that b is represented in canonical form by*

$$(v, w) = \sum_{i=1}^{k} v^i w^{k+i} - \sum_{j=1}^{k} v^{k+j} w^j.$$

and the symplectic matrix groups are given by

$$Sp(n, \mathbb{C}) = \{A \in M_{2k\times 2k}(\mathbb{C}) : (Av, Aw) = (v, w)\}$$
$$Sp(n, \mathbb{R}) = \{A \in M_{2k\times 2k}(\mathbb{R}) : (Av, Aw) = (v, w)\}$$

The groups of the form $SAut(\mathrm{V}, b, \mathbb{F}) = Aut(\mathrm{V}, b, \mathbb{F}) \cap SL(\mathrm{V}, \mathbb{F})$ are usually designated by use of the word "special". We have the **special orthogonal** and **special semi-orthogonal** groups $SO(n)$ and $SO(p,q)$, the **special unitary** and **special semi-unitary** groups $SU(n)$and $SU(p,q)$ etc.

Exercise 5.1 *Show that $SU(2)$ is simply connected.*

Exercise 5.2 *Let $g \in G$ (a Lie group). Show that each of the following maps $G \to G$ is a diffeomorphism:*

1) $L_g : x \mapsto gx$ (left translation)
2) $R_g : x \mapsto xg$ (right translation)
3) $C_g : x \mapsto gxg^{-1}$ (conjugation).
4) $inv : x \mapsto x^{-1}$ (inversion)

5.3 Lie Group Homomorphisms

The following definition is manifestly natural:

Definition 5.3 *A smooth map $f : G \to H$ is called a* ***Lie group homomorphism*** *if*

$$\begin{aligned} f(g_1 g_2) &= f(g_1) f(g_2) \text{ for all } g_1, g_2 \in G \text{ and} \\ f(g^{-1}) &= f(g)^{-1} \text{ for all } g \in G. \end{aligned}$$

and an ***Lie group isomorphism*** *in case it has an inverse that is also a Lie group homomorphism. A Lie group isomorphism $G \to G$ is called a* ***Lie group automorphism.***

Example 5.8 *The inclusion $S\mathrm{O}(n, \mathbb{R}) \hookrightarrow \mathrm{GL}(n, \mathbb{R})$ is a Lie group homomorphism.*

Example 5.9 *The circle $S^1 \subset \mathbb{C}$ is a Lie group under complex multiplication and the map*

$$z = e^{i\theta} \to \begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & I_{n-2} \end{bmatrix}$$

is a Lie group homomorphism into $S\mathrm{O}(n)$.

Example 5.10 *The conjugation map $C_g : G \to G$ is a Lie group automorphism.*

Exercise 5.3 (*) *Show that the multiplication map $\mu : G \times G \to G$ has tangent map at $(e, e) \in G \times G$ given as $T_{(e,e)}\mu(v, w) = v + w$. Recall that we identify $T_{(e,e)}(G \times G)$ with $T_e G \times T_e G$.*

Exercise 5.4 $\mathrm{GL}(n, \mathbb{R})$ *is an open subset of the vector space of all $n \times n$ matrices $M_{n\times n}(\mathbb{R})$. Using the natural identification of $T_e\mathrm{GL}(n, \mathbb{R})$ with $M_{n\times n}(\mathbb{R})$ show that as a map $M_{n\times n}(\mathbb{R}) \to M_{n\times n}(\mathbb{R})$ we have*

$$T_e C_g : x \mapsto g x g^{-1}$$

where $g \in \mathrm{GL}(n, \mathbb{R})$ and $x \in M_{n\times n}(\mathbb{R})$.

Example 5.11 *The map $t \mapsto e^{it}$ is a Lie group homomorphism from $\mathbb{R}$ to $S^1 \subset \mathbb{C}$.*

Definition 5.4 *An* ***immersed subgroup*** *of a Lie group G is a smooth injective immersion that is also a Lie group homomorphism. The image of such a homomorphism is also referred to as an* ***immersed subgroup*** *and is an abstract subgroup of G though not necessarily a regular submanifold of G.*

Remark 5.1 *It is an unfortunate fact that in this setting a map itself is sometimes referred to as a "subgroup". We will avoid this terminology as much as possible preferring to call maps like those in the last definition simply **injective homomorphisms** and reserving the term subgroup for the image set. Term one-parameter subgroup from the next definition is a case where we give in to tradition.*

The image of an immersed subgroup is a Lie subgroup but not necessarily closed. If $H \subset G$ is such a subgroup then the inclusion map is an $H \hookrightarrow G$ immersed subgroup as in definition 5.4.

Definition 5.5 *A homomorphism from the additive group $\mathbb{R}$ into a Lie group is called a **one-parameter subgroup**.*

Example 5.12 *We have seen that the torus $S^1 \times S^1$ is a Lie group under multiplication given by $(e^{i\tau_1}, e^{i\theta_1})(e^{i\tau_2}, e^{i\theta_2}) = (e^{i(\tau_1+\tau_2)}, e^{i(\theta_1+\theta_2)})$. Every homomorphism of $\mathbb{R}$ into $S^1 \times S^1$, that is, every one parameter subgroup of $S^1 \times S^1$ is of the form $t \mapsto (e^{tai}, e^{tbi})$ for some pair of real numbers $a, b \in \mathbb{R}$.*

Example 5.13 *The map $R : \mathbb{R} \to \mathrm{SO}(3)$ given by*

$$\theta \mapsto \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

is a one parameter subgroup.

Recall that an $n \times n$ complex matrix A is called Hermitian (resp. skew-Hermitian) if $\bar{A}^t = A$ (resp. $\bar{A}^t = -A$).

Example 5.14 *Given an element g of the group $SU(2)$ we define the map $Ad_g : \mathfrak{su}(2) \to \mathfrak{su}(2)$ by $Ad_g : x \mapsto gxg^{-1}$. Now the skew-Hermitian matrices of zero trace can be identified with $\mathbb{R}^3$ by using the following matrices as a basis:*

$$\begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}.$$

*These are just $-i$ times the **Pauli matrices** $\sigma_1, \sigma_2, \sigma_3$ [1] and so the correspondence $\mathfrak{su}(2) \to \mathbb{R}^3$ is given by $-xi\sigma_1 - yi\sigma_2 - iz\sigma_3 \mapsto (x, y, z)$. Under this correspondence the inner product on $\mathbb{R}^3$ becomes the inner product $(A, B) = trace(AB)$. But then*

$$\begin{aligned} (\mathrm{Ad}_g A, \mathrm{Ad}_g B) &= \mathrm{trace}(gAgg^{-1}Bg^{-1}) \\ &= \mathrm{trace}(AB) = (A, B) \end{aligned}$$

so actually Ad_g can be thought of as an element of $O(3)$. More is true; Ad_g acts as an element of $SO(3)$ and the map $g \mapsto Ad_g$ is then a homomorphism from $SU(2)$ to $SO(\mathfrak{su}(2)) \cong SO(3)$. This is a special case of the adjoint map studied later.

[1] check this

The set of all Lie groups together with Lie group homomorphisms forms an important category that has much of the same structure as the category of all groups. The notions of kernel, image and quotient show up as expected.

Theorem 5.2 *Let $h : G \to H$ be a Lie group homomorphism. The subgroups $Ker(h) \subset G$, $Img(h) \subset H$ are Lie subgroups. $Ker(h)$ is in fact a closed subgroup.*

Definition 5.6 *If a Lie group homomorphism $\wp : \widetilde{G} \to G$ is also a covering $\wp : \widetilde{G} \to G$ then we say that $\widetilde{G}$ is a covering group and $\wp$ is a covering homomorphism. If $\widetilde{G}$ is simply connected then $\widetilde{G}$ (resp. $\wp$) is called the universal covering group (resp. universal covering homomorphism) of G.*

Exercise 5.5 *Show that if $\wp : \widetilde{M} \to G$ is a smooth covering map and G is a Lie group then $\widetilde{M}$ can be given a unique Lie group structure such that $\wp$ becomes a covering homomorphism.*

Example 5.15 *The group Mob of transformations of the complex plane given by $T_A : z \mapsto \frac{az+b}{cz+d}$ for $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{C})$ can be given the structure of a Lie group. The map $\wp : \mathrm{SL}(2,\mathbb{C}) \to Mob$ given by $\wp : A \mapsto T_A$ is onto but not injective. In fact, it is a (two fold) covering homomorphism. When do two elements of $\mathrm{SL}(2,\mathbb{C})$ map to the same element of Mob?*

5.4 Problem Set

1. Show that $\mathrm{SL}(2,\mathbb{C})$ is simply connected (so that $\wp : \mathrm{SL}(2,\mathbb{C}) \to Mob$ is a covering homomorphism)

2. Show that if we consider $\mathrm{SL}(2,\mathbb{R})$ as a subset of $\mathrm{SL}(2,\mathbb{C})$ in the obvious way then $\mathrm{SL}(2,\mathbb{R})$ is a Lie subgroup of $\mathrm{SL}(2,\mathbb{C})$ and $\wp(\mathrm{SL}(2,\mathbb{R}))$ is a Lie subgroup of Mob. Show that if $T \in \wp(\mathrm{SL}(2,\mathbb{R}))$ then T maps the upper half plane of $\mathbb{C}$ onto itself (bijectively).

3. Determine explicitly the map $T_I \operatorname{inv} : T_I\mathrm{GL}(n,\mathbb{R}) \to T_I\mathrm{GL}(n,\mathbb{R})$ where $\operatorname{inv} : A \mapsto A^{-1}$.

Chapter 6

Vector Bundles

6.1 Definitions

The tangent and cotangent bundles are examples a general object called a vector bundle. Roughly speaking, a vector bundle is a parameterized family of vector spaces called fibers. We shall need both complex vector bundles and real vector bundles and so to facilitate definitions we let $\mathbb{F}$ denote either $\mathbb{R}$ or $\mathbb{C}$. The simplest examples of vector bundles over a manifold M are the product bundles which consist of a Cartesian product $M \times \mathrm{V}$ for some fixed (finite dimensional) vector space V over $\mathbb{F}$ together with the projection onto the first factor $pr_1 : M \times \mathrm{V} \to M$. Each set of the form $\{x\} \times \mathrm{V} \subset M \times \mathrm{V}$ inherits an $\mathbb{F}$-vector space structure from that of V in the obvious way. We think of $M \times \mathrm{V}$ as having copies of V parameterized by $x \in M$.

Definition 6.1 *A C^k* ***vector bundle*** *(with typical fiber V) is a surjective C^k map $\pi : E \to M$ such that for each $x \in M$ the set $E_x := \pi^{-1}\{x\}$ has the structure of a vector space over the field $\mathbb{F}$ isomorphic to the fixed vector space V and such that the following local triviality condition is met: Every $p \in M$ has a neighborhood U such that there is a C^k diffeomorphism $\phi : \pi^{-1}U \to U \times \mathrm{V}$ which makes the following diagram commute;*

$$\begin{tikzcd} \pi^{-1}U \arrow[rr, "\phi"] \arrow[dr, "\pi"'] & & U \times \mathrm{V} \arrow[dl, "pr_1"] \\ & U & \end{tikzcd}$$

and such that for each $x \in U$ the map $\phi|_{E_x} : E_x \to \{x\} \times \mathrm{V}$ is a vector space isomorphism.

In the definition above, the space E is called the **total space**, M is called the **base space**, and $E_x := \pi^{-1}\{x\}$ is called the **fiber** at x (fiber over x). A map $\pi^{-1}U \xrightarrow{\phi} U \times \mathrm{V}$ of the sort described in the definition is called a **local trivialization** and the pair (U, ϕ) is called a **vector bundle chart** (VB-chart).

A family $\{(U_\alpha, \phi_\alpha)\}$ of vector bundle charts such that $\{U_\alpha\}$ is an open cover of M is called a **vector bundle atlas** for $\pi : E \to M$. The definition guarantees that such an atlas exist. The dimension of the typical fiber is called the **rank** of the vector bundle. Note that if we have a surjection $\pi : E \to M$ and an atlas then we certainly have a vector bundle so in practice, if one is trying to show that a surjection is a vector bundle then one just exhibits an atlas. One could have defined vector bundle charts and atlases and then used those notions to define vector bundle. For yet another slightly different approach that starts with just the space E and special notion of atlas see section 3.4 of [A,B,R]. There are obvious notions of equivalent vector bundle atlases and maximal vector bundle atlases.

Remark 6.1 *In the definition of a vector bundle, the $r-$dimensional vector space* V *is usually just presented as $\mathbb{F}^r$ but sometimes it is better to be more specific. For example,* V *might be the vector space of $n \times m$ matrices ($r = mn$), the Lie algebra of a Lie group, or perhaps a tensor space like $T^k_l(\mathbb{R}^n)$.*

If we look at the set of vector bundles over a fixed base space M then we have a category whose morphisms are called strong vector bundle homomorphisms that are defined as follows: Given two vector bundles $\pi_1 : E_1 \to M$ and $\pi_2 : E_2 \to M$ a strong vector bundle homomorphism from π_1 to π_2 is a map $\Phi : E_1 \to E_2$ which restricts on each fiber to a linear map $E_{1p} \to E_{2\Phi(p)}$ and such that the following diagram commutes

$$\begin{array}{ccc} E_1 & \overset{\Phi}{\to} & E_2 \\ \pi_1 \searrow & & \swarrow \pi_2 \\ & M & \end{array} .$$

If Φ is a diffeomorphism it is called a (strong) **vector bundle isomorphism**. If two different base spaces are involved we have a slightly different notion of vector bundle morphism. Let $\pi_1 : E_1 \to M_1$ and $\pi_2 : E_2 \to M_2$ be vector bundles. A (weak) vector bundle homomorphism is a pair of maps (Φ, ϕ) where $\Phi : E_1 \to E_2$ is linear on each fiber and such that the following diagram commutes:

$$\begin{array}{ccc} E_1 & \overset{\Phi}{\to} & E_2 \\ \pi_1 \downarrow & & \downarrow \pi_2 \\ M & \overset{\phi}{\to} & M \end{array} .$$

For any two vector bundle charts (U_α, ϕ_α) and (U_β, ϕ_β) such that $U_\alpha \cap U_\beta \neq \emptyset$ the composition $\phi_\alpha \circ \phi_\beta^{-1} : (U_\alpha \cap U_\beta) \times \mathrm{V} \to (U_\alpha \cap U_\beta) \times \mathrm{V}$ is called an overlap map. It is not hard to see that $\phi_\alpha \circ \phi_\beta^{-1}$ must be of the form $\phi_\alpha \circ \phi_\beta^{-1}(x, v) = (x, g_{\alpha\beta}(x)v)$ for some smooth map $\phi_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(\mathrm{V})$ which is called a transition map. For a given vector bundle atlas we get a family of transition maps which must always satisfy the following properties:

1. $g_{\alpha\alpha}(x) = id_{\mathrm{V}}$ for all $x \in U_\alpha$ and all α;

2. $g_{\alpha\beta}(x) \circ g_{\beta\alpha}(x) = id_{\mathrm{V}}$ for all $x \in U_\alpha \cap U_\beta$

3. $g_{\alpha\gamma}(x) \circ g_{\gamma\beta}(x) \circ g_{\beta\alpha}(x) = id_{\mathrm{V}}$ for all $x \in U_\alpha \cap U_\beta \cap U_\gamma$

A family of maps $\{g_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(\mathrm{V})$ which satisfy the above three conditions for some cover of a manifold is called a $\mathrm{GL}(\mathrm{V})-$**cocycle**.

Theorem 6.1 *Given a cover $\{U_\alpha\}_{\alpha\in A}$ of M and* $\mathrm{GL}(\mathrm{V})-$*cocycle $\{g_{\alpha\beta}\}$ for the cover there exists a vector bundle with an atlas $\{(U_\alpha, \phi_\alpha)\}$ satisfying $\phi_\alpha \circ \phi_\beta^{-1}(x, y) = (x, g_{\alpha\beta}(x) \cdot y))$ on nonempty overlaps $U_\alpha \cap U_\beta$.*

Proof. On the union $\Sigma := \bigcup_\alpha \{\alpha\} \times U_\alpha \times \mathrm{V}$ define an equivalence relation such that

$$(\alpha, u, v) \in \{\alpha\} \times U_\alpha \times \mathrm{V}$$

is equivalent to $(\beta, x, y) \in \{\beta\} \times U_\beta \times \mathrm{V}$ if and only if $u = x$ and $v = g_{\alpha\beta}(x) \cdot y$.

The total space of our bundle is then $E := \Sigma/\sim$. The set Σ is essentially the disjoint union of the product spaces $U_\alpha \times \mathrm{V}$ and so has an obvious topology. We then give $E := \Sigma/\sim$ the quotient topology. The bundle projection π_E is induced by $(\alpha, u, v) \mapsto u$. To get our trivializations we define

$$\phi_\alpha(e) := (u, v) \text{ for } e \in \pi_E^{-1}(U_\alpha)$$

where (u, v) is the unique member of $U_\alpha \times \mathrm{V}$ such that $(\alpha, u, v) \in e$. The point here is that $(\alpha, u_1, v_1) \sim (\alpha, u_2, v_2)$ only if $(u_1, v_1) = (u_2, v_2)$. Now suppose $U_\alpha \cap U_\beta \neq \emptyset$. Then for $x \in U_\alpha \cap U_\beta$ the element $\phi_\beta^{-1}(x, y)$ is in $\pi_E^{-1}(U_\alpha \cap U_\beta) = \pi_E^{-1}(U_\alpha) \cap \pi_E^{-1}(U_\beta)$ and so $\phi_\beta^{-1}(x, y) = [(\beta, x, y)] = [(\alpha, u, v)]$. This means that $x = u$ and $v = g_{\alpha\beta}(x) \cdot y$. From this it is not hard to see that

$$\phi_\alpha \circ \phi_\beta^{-1}(x, y) = (x, g_{\alpha\beta}(x) \cdot y)).$$

We leave the routine verification of the regularity of these maps and the existence of the C^r structure to the reader. ∎

We already know what it means for two vector bundles to be isomorphic and of course any two vector bundles that are naturally isomorphic will be thought of as the same. Since we can and often do construct our bundles according to the above recipe it will pay to know something about when two bundles over M are isomorphic based on their respective transition function.

Proposition 6.1 *Two bundles $\pi : E \to M$ and $\pi' : E' \to M$ with transition functions $\{g_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(\mathrm{V})\}$ and $\{g'_{\alpha\beta} : U_{\alpha\beta} \to \mathrm{GL}(\mathrm{V})\}$ over the same cover $\{U_\alpha\}$ are (strongly) isomorphic if and only if there are* $\mathrm{GL}(\mathrm{V})$ *valued functions f_α defined on each U_a such that*

$$g'_{\alpha\beta}(x) = f_\alpha(x) g_{\alpha\beta}(x) f_\alpha^{-1}(x) \textit{ for } x \in U_\alpha \cap U_\beta \tag{6.1}$$

Proof. Given a vector bundle isomorphism $f : E \to E'$ over M let $f_\alpha := \phi'_\alpha f \phi_\alpha^{-1}$. Check that this works.

Conversely, given functions f_α satisfying 6.1 define $\widetilde{f}_\alpha : U_\alpha \times \mathrm{V} \to E|_{U_a}$ by $(x, v) \mapsto f_\alpha(x)v$. Now we define $f : E \to E'$ by

$$f(e) := \left((\phi'_\alpha)^{-1} \circ \widetilde{f}_\alpha \circ \phi_\alpha \right)(e) \text{ for } e \in E|_{U_a}$$

The conditions 6.1 insure that f is well defined on the overlaps $E|_{U_a} \cap E|_{U_\beta} = E|_{U_a \cap U_\beta}$. One easily checks that this is a vector bundle isomorphism. ■

We can use this construction to arrive at several often used vector bundles.

Example 6.1 *Given an atlas $\{(U_\alpha, \mathtt{x}_\alpha)\}$ for a smooth manifold M we can let $g_{\alpha\beta}(x) = T_x\mathtt{x}_\alpha \circ T_x\mathtt{x}_\beta^{-1}$ for all $x \in U_\alpha \cap U_\beta$. The bundle constructed according to the recipe of theorem 6.1 is a vector bundle (isomorphic to) the tangent bundle TM. If we let $g^*_{\alpha\beta}(x) = (T_x\mathtt{x}_\beta \circ T_x\mathtt{x}_\alpha^{-1})^t$ then we arrive at the cotangent bundle T^*M.*

Example 6.2 *Recall the space $T^k_l(\mathbb{R}^n)$ consisting of functions $[n] \times ... \times [n] \times .. \times [n] \to \mathbb{R}$ of the form $t : (i_1, ..., i_k, j_1, ..., j_l) \mapsto t^{\ i_1 \ldots i_k}\big|_{j_1 \ldots j_l}$. For $g_{\alpha\beta}(x)$ as in the last example we let $T^k_l g_{\alpha\beta}(x)$ be the element of $\mathrm{GL}(T^k_l(\mathbb{R}^n))$ defined by*

$$T^k_l g_{\alpha\beta}(x) : (t')^{\ i_1 \ldots i_k}\big|_{j_1 \ldots j_l} \longmapsto g^{i_1}_{\ r_1} \cdots g^{i_k}_{\ r_l} t^{\ r_1 \ldots r_k}\big|_{s_1 \ldots s_l} \left(g^{-1}\right)^{s_1}_{\ \ j_1} \cdots \left(g^{-1}\right)^{s_k}_{\ \ j_l} \tag{6.2}$$

where $(g^i_j) := g_{\alpha\beta}(x) \in \mathrm{GL}(\mathbb{R}^n)$. Using the $\bar{g}_{\alpha\beta} := T^k_l g_{\alpha\beta}$ as the cocycles we arrive, by the above construction, at a vector bundle (isomorphic to) $T^k_l(TM)$. This is called a tensor bundle and we give a slightly different construction of this bundle below when we study tensor fields.

Example 6.3 *Suppose we have two vector bundles $\pi_1 : E_1 \to M$ and $\pi_2 : E_2 \to M$. We give two constructions of the **Whitney sum** $\pi_1 \oplus \pi_2 : E_1 \oplus E_2 \to M$. This is a globalization of the direct sum construction of vector spaces. In fact, the first construction simply takes $E_1 \oplus E_2 = \bigsqcup_{p \in M} E_{1p} \oplus E_{2p}$. Now we have a vector bundle atlas $\{(\phi_\alpha, U_\alpha)\}$ for π_1 and a vector bundle atlas $\{(\psi_\alpha, U_\alpha)\}$ for π_2 We have assumed that both atlases have the same family of open sets (we can arrange this by taking a common refinement). Now let $\phi_\alpha \oplus \psi_\alpha : (v_p, w_p) \mapsto (p, \phi_\alpha v_p, \psi_\alpha w_p)$ for all $(v_p, w_p) \in (E_1 \oplus E_2)|_{U_\alpha}$. Then $\{(\phi_\alpha \oplus \psi_\alpha, U_\alpha)\}$ is an atlas for $\pi_1 \oplus \pi_2 : E_1 \oplus E_2 \to M$.*

Another method of constructing is to take the cocycle $\{g_{\alpha\beta}\}$ for π_1 and the cocycle $\{h_{\alpha\beta}\}$ for π_2 and then let $g_{\alpha\beta} \oplus h_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(\mathbb{F}^{r_1} \times \mathbb{F}^{r_2})$ be defined by $(g_{\alpha\beta} \oplus h_{\alpha\beta})(x) = g_{\alpha\beta}(x) \oplus h_{\alpha\beta}(x) : (v, w) \longmapsto (g_{\alpha\beta}(x)v, h_{\alpha\beta}(x)w)$. The maps $g_{\alpha\beta} \oplus h_{\alpha\beta}$ form a cocycle which determines a bundle by the construction of proposition 6.1 which is (isomorphic to) the $\pi_1 \oplus \pi_2 : E_1 \oplus E_2 \to M$.

From now on we shall often omit explicit mention of the degree of regularity of a vector bundle: By "vector bundle" we shall mean a C^r vector bundle for some fixed r and all the maps are assumed to be C^r maps. We now introduce the important notion of a section of a vector bundle.

Let $\pi : E \to M$ be a vector bundle. A (global) **section** of the bundle π is a map $\sigma : M \to E$ such that $\pi \circ \sigma = id_M$. The set of all sections of π is denoted by $\Gamma(\pi)$ or by $\Gamma(E)$. A **local section** over $U \subset M$ is a map $\sigma : U \to E$ such that $\pi \circ \sigma = id_U$. The set of local sections over U is denoted $\Gamma(U, E)$. Every vector bundle has sections an obvious one being the **zero section** which maps each $x \in M$ to the zero element 0_x of the fiber E_x.

Exercise 6.1 *Show that the range of the zero section of a vector bundle $E \to M$ is a submanifold of E that is diffeomorphic to M.*

For every open set $U \subset M$ the set $E|_U := \pi^{-1}(U)$ is the total space of an obvious vector bundle $\pi|_U : E|_U \to U$ called the restriction to U. A local section of π over U is just a (global) section of the restricted bundle $\pi|_U : E|_U \to U$.

Definition 6.2 *Let $\pi : E \to M$ be a (C^r) rank k vector bundle. An k-tuple $\sigma = (e_1, ..., e_k)$ of (C^r) sections of E over an open set U is called a (local)* ***moving frame*** *or* ***frame field*** *over U if for all $p \in U$, $\{e_1(p), ..., e_k(p)\}$ is a basis for E_p.*

A choice of a local frame field over an open set $U \subset M$ is equivalent to a local trivialization (a vector bundle chart). Namely, if ϕ is such a trivialization over U then defining $e_i(x) = \phi^{-1}(x, \mathsf{e}_i)$ where $\{\mathsf{e}_i\}$ is a fixed basis of V we have that $\sigma = (e_1, ..., e_k)$ is a moving frame over U. Conversely, if $\sigma = (e_1, ..., e_k)$ is a moving frame over U then every $v \in \pi^{-1}(U)$ has the form $v = \sum v^i(x)e_i(x)$ for a unique x and unique numbers $v^i(x)$. Then the map $v \mapsto (x, v^1(x), ..., v^1(x))$ is a vector bundle chart over U. If there is a global frame field then the vector bundle is **trivial** meaning that it is vector bundle isomorphic to $pr_1 : M \times \mathrm{V} \to M$.

Definition 6.3 *A manifold M is said to be parallelizable if TM is trivial; i.e., if TM has a global frame field.*

Proposition 6.2 *Let $\pi : E \to M$ be vector bundle with typical fiber* V *and transition functions $\{\phi_{\alpha\beta}\}$ with respect to an open cover $\{U_\alpha\}$. Let $s_\alpha : U_\alpha \to$* V *be a collection of functions such that whenever $U_\alpha \cap U_\beta \neq \emptyset$, $s_\alpha(x) = \phi_{\alpha\beta}(x)s_\beta(x)$ for all $x \in U_\alpha \cap U_\beta$. Then there is a global section s such that $s|_{U_\alpha} = s_\alpha$ for all α.*

Proof. Let $\phi_\alpha : E|_{U_\alpha} \to U_\alpha \times \mathrm{V}$ be the trivializations that give rise to the cocycle $\{\phi_{\alpha\beta}\}$. Let $\gamma_\alpha(x) := (x, s_\alpha(x))$ for $x \in U_\alpha$ and then let $s|_{U_\alpha} := \phi_\alpha^{-1} \circ \gamma_\alpha$. This gives a well defined section s because for $x \in U_\alpha \cap U_\beta$ we have

$$\begin{aligned}
\phi_\alpha^{-1} \circ \gamma_\alpha(x) &= \phi_\alpha^{-1}(x, s_\alpha(x)) \\
&= \phi_\alpha^{-1}(x, \phi_{\alpha\beta}(x)s_\beta(x)) \\
&= \phi_\alpha^{-1} \circ \phi_\alpha \circ \phi_\beta^{-1}(x, s_\beta(x)) \\
&= \phi_\beta^{-1}(x, s_\beta(x)) = \phi_\beta^{-1} \circ \gamma_\beta(x)
\end{aligned}$$

∎

6.1.1 Pullback bundles and sections along maps.

Definition 6.4 *Let $\pi : E \to M$ be a vector bundle and $f : N \to M$ be a map. A **section of** π **along** f is a map $\sigma : N \to E$ such that $\pi \circ \sigma = f$.*

A section of the tangent bundle along a map is called a vector field along the map. Sections along maps appear quite often and one can work directly with this definition but one can also give in to the compulsion to make everything in sight a section of a bundle. In order to interpret sections along maps as sections of bundles we need a new vector bundle construction.

Definition 6.5 *Let $\pi : E \to M$ be a vector bundle and $f : N \to M$ be a map. The **pullback** vector bundle $f^*\pi : f^*E \to M$ (or **induced** vector bundle) defined as follows: The total space f^*E is the set $\{(x, e) \in N \times E$ such that $f(x) = \pi(e)$.*

Now one can easily verify that a section of $f^*\pi : f^*E \to M$ is always given by a map of the form $s : x \to (x, \sigma(x))$ where σ is a section along f. We will work directly with the map σ whenever convenient but the bundle viewpoint should be kept in mind.

Example 6.4 *Let $\pi_1 : E_1 \to M$ and $\pi_2 : E_2 \to M$ be vector bundles and let $\triangle : M \to M \times M$ be the diagonal map given by $x \mapsto (x, x)$. From π_1 and π_2 one can construct a bundle $\pi_{E_1 \times E_2} : E_1 \times E_2 \to M \times M$ by $\pi_{E_1 \times E_2}(e_1, e_2)$. The Whitney sum bundle defined previously may also be defined as $\triangle^* \pi_{E_1 \times E_2} : \triangle^*(E_1 \times E_2) \to M$. Of course one would write $E_1 \oplus E_2$ for $\triangle^*(E_1 \times E_2)$.*

6.2 Examples

Example 6.5 (Canonical line bundle) *Recall that $P^n(\mathbb{R})$ is the set of all lines through the origin in $\mathbb{R}^{n+1}$. Define the subset $\mathbb{L}(P^n(\mathbb{R}))$ of $P^n(\mathbb{R}) \times \mathbb{R}^{n+1}$ consisting of all pairs (l, v) such that $v \in l$ (think about this). This set together with the map $\pi_{P^n(\mathbb{R})} : (l, v) \mapsto l$ is a rank one vector bundle. The bundle charts are of the form $\pi^{-1}_{P^n(\mathbb{R})}(U_i), \widetilde{\psi}_i$ where $\widetilde{\psi}_i : (l, v) \mapsto (\psi_i(l), pr_i(v)) \in \mathbb{R}^n \times \mathbb{R}$.*

Example 6.6 (Tautological Bundle) *Let $G(n, k)$ denote the Grassmann manifold of k-planes in $\mathbb{R}^n$. Let $\gamma_{n,k}$ be the subset of $G(n, k) \times \mathbb{R}^n$ consisting of pairs (P, v) where P is a k-plane (k-dimensional subspace) and v is a vector in the plane P. The projection of the bundle is simply $(P, v) \mapsto P$. We leave it to the reader to discover an appropriate VB-atlas.*

Note well that these vector bundles are not just trivial bundles and in fact their topology or twistedness (for large n) is of the utmost importance for the topology of other vector bundles. One may take the inclusions $...\mathbb{R}^n \subset \mathbb{R}^{n+1} \subset ... \subset \mathbb{R}^\infty$ to construct inclusions $...G(n, k) \subset G(n + 1, k)$...and $...\gamma_{n,k} \subset \gamma_{n+1,k}$. from which a "universal bundle" $\gamma_n \to G(n)$ is constructed with the property that every rank k vector bundle E over X is the pull-back by some map $f : X \to G(n)$:

$$\begin{array}{cccc} E \cong & f^*\gamma_n & \to & \gamma_n \\ & \downarrow & & \downarrow \\ & X & \xrightarrow{f} & G(n) \end{array}$$

`To Reviewer: Give detail of the universal bundle construction here, in an appendix of later in the text?`

Exercise 6.2 *To each point on a sphere attach the space of all vectors normal to the sphere at that point. Show that this* normal bundle *is in fact a (smooth) vector bundle. Also, in anticipation of the next section, do the same for the union of tangent planes to the sphere.*

Exercise 6.3 *Let* $Y = \mathbb{R} \times \mathbb{R}$ *and define let* $(x_1, y_1) \sim (x_2, y_2)$ *if and only if* $x_1 = x_2 + jk$ *and* $y_1 = (-1)^{jk} y_2$ *for some integer* k. *Show that* $E := Y/\sim$ *is a vector bundle of rank 1 that is trivial if and only if* j *is even. Convince yourself that this is the Mobius band when* j *is odd.*

6.3 Modules

A module is an algebraic object that shows up quite a bit in differential geometry and analysis. A module is a generalization of a vector space where the field is replace by a ring or algebra over a field. The modules that occur are almost always finitely generated projective modules over the algebra of C^r functions and these correspond to the spaces of C^r sections of vector bundles. We give the abstract definitions but we ask the reader to keep in mind two cases. The first is just the vectors spaces which are the fibers of vector bundles. In this case the ring in the definition below is the field $\mathbb{F}$ (the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$) and the module is just a vector space. The second case, already mentioned, is where the ring is the algebra $C^r(M)$ and the module is the set of C^r sections of a vector bundle.

As we have indicated, a module is similar to a vector space with the differences stemming from the use of elements of a ring R as the scalars rather than the field of complex $\mathbb{C}$ or real numbers $\mathbb{R}$. For a module, one still has $1w = w$, $0w = 0$ and $-1w = -w$. Of course, every vector space is also a module since the latter is a generalization of the notion of vector space. We also have maps between modules, the module homomorphisms (see definition 6.7 below), which make the class of modules and module homomorphism into a category.

Definition 6.6 *Let* R *be a ring. A left* R-***module*** *(or a left* ***module*** *over* R*) is an abelian group* $W, +$ *together with an operation* $\mathsf{R} \times W \to W$ *written* $(a, w) \mapsto aw$ *such that*

1) $(a + b)w = aw + bw$ *for all* $a, b \in \mathsf{R}$ *and all* $w \in W$,

2) $a(w_1 + w_2) = aw_1 + aw_2$ *for all* $a \in \mathsf{R}$ *and all* $w_2, w_1 \in W$.

A right R-***module*** *is defined similarly with the multiplication of the right so that*
1) $w(a+b) = wa + wb$ *for all* $a, b \in \mathsf{R}$ *and all* $w \in W$,
2) $(w_1 + w_2)a = w_1 a + w_2 a$ *for all* $a \in \mathsf{R}$ *and all* $w_2, w_1 \in W$. **.**

If the ring is commutative (the usual case for us) then we may write $aw = wa$ and consider any right module as a left module and visa versa. Even if the ring is not commutative we will usually stick to left modules and so we drop the reference to "left" and refer to such as R-modules.

Remark 6.2 *We shall often refer to the elements of* R *as* ***scalars****.*

Example 6.7 *An abelian group* $A, +$ *is a* $\mathbb{Z}$ *module and a* $\mathbb{Z}$*-module is none other than an abelian group. Here we take the product of* $n \in \mathbb{Z}$ *with* $x \in A$ *to be* $nx := x + \cdots + x$ *if* $n \geq 0$ *and* $nx := -(x + \cdots + x)$ *if* $n < 0$ *(in either case we are adding* $|n|$ *terms).*

Example 6.8 *The set of all* $m \times n$ *matrices with entries being elements of a commutative ring* R *(for example real polynomials) is an* R*-module.*

Definition 6.7 *Let* W_1 *and* W_2 *be modules over a ring* R*. A map* $L : W_1 \to W_2$ *is called* ***module homomorphism*** *if*

$$L(aw_1 + bw_2) = aL(w_1) + bL(w_2).$$

By analogy with the case of vector spaces, which module theory includes, we often characterize a module homomorphism L *by saying that* L *is* ***linear over*** R.

Example 6.9 *The set of all module homomorphisms of a module* W *(lets say over a commutative ring* R*) onto another module* M *is also a module in its own right and is denoted* $\mathrm{Hom}_{\mathsf{R}}(\mathrm{W}, \mathrm{M})$ *or* $L_{\mathsf{R}}(\mathrm{W}, \mathrm{M})$*.*

Example 6.10 *Let* V *be a vector space and* $\ell : \mathrm{V} \to \mathrm{V}$ *a linear operator. Using this one operator we may consider* V *as a module over the ring of polynomials* $\mathbb{R}[t]$ *by defining the "scalar" multiplication by*

$$p(t)v := p(\ell)v$$

for $p \in \mathbb{R}[t]$.

Since the ring is usually fixed we often omit mention of the ring. In particular, we often abbreviate $L_{\mathsf{R}}(\mathrm{V}, \mathrm{W})$ to $L(\mathrm{V}, \mathrm{W})$. Similar omissions will be made without further mention. Also, since every real (resp. complex) Banach space E is a vector space and hence a module over $\mathbb{R}$ (resp. $\mathbb{C}$) we must distinguish between the bounded linear maps which we have denoted up until now as $L(\mathsf{E}; \mathsf{F})$ and the linear maps that would be denoted the same way in the context of modules.

Definition 6.8 ((convention)) *When discussing modules the notation $L(\mathrm{V}, \mathrm{W})$ will mean linear maps rather than bounded linear maps, even if* V *and* W *happen to be vector spaces.*

A **submodule** is defined in the obvious way as a subset $S \subset \mathrm{W}$ that is closed under the operations inherited from W so that S itself is a module. The intersection of all submodules containing a subset $A \subset \mathrm{W}$ is called the **submodule generated by** A and is denoted $\langle A \rangle$. A is called a **generating set**. If $\langle A \rangle = \mathrm{W}$ for a *finite* set A, then we say that W is **finitely generated**.

Let S be a submodule of W and consider the quotient abelian group W/S consisting of cosets, that is sets of the form $[v] := v + S = \{v + x : x \in S\}$ with addition given by $[v] + [w] = [v + w]$. We define a scalar multiplication by elements of the ring R by $a[v] := [av]$ respectively. In this way, W/S is a module called a **quotient module**.

A real (resp. complex) vector space is none other than a module over the field of real numbers $\mathbb{R}$ (resp. complex numbers $\mathbb{C}$). Many of the operations that exist for vector spaces have analogues in the module category. For example, the direct sum of modules is defined in the obvious way. Also, for any module homomorphism $L : \mathrm{W}_1 \to \mathrm{W}_2$ we have the usual notions of kernel and image:

$$\begin{aligned} \ker L &= \{v \in \mathrm{W}_1 : L(v) = 0\} \\ \operatorname{img}(L) = L(\mathrm{W}_1) &= \{w \in \mathrm{W}_2 : w = Lv \text{ for some } v \in \mathrm{W}_1\}. \end{aligned}$$

These are submodules of W_1 and W_2 respectively.

On the other hand, modules are generally not as simple to study as vector spaces. For example, there are several notions of dimension. The following notions for a vector space all lead to the same notion of dimension. For a completely general module these are all potentially different notions:

1. Length n of the longest chain of submodules

$$0 = \mathrm{W}_n \subsetneq \cdots \subsetneq \mathrm{W}_1 \subsetneq \mathrm{W}$$

2. The cardinality of the largest linearly independent set (see below).

3. The cardinality of a basis (see below).

For simplicity in our study of dimension, let us now assume that R is commutative.

Definition 6.9 *A set of elements $w_1, ..., w_k$ of a module are said to be **linearly dependent** if there exist ring elements $r_1, ..., r_k \in \mathsf{R}$ not all zero, such that $r_1 w_1 + \cdots + r_k w_k = 0$. Otherwise, they are said to be **linearly independent**. We also speak of the set $\{w_1, ..., w_k\}$ as being a **linearly independent set**.*

So far so good but it is important to realize that just because $w_1, ..., w_k$ are linearly independent doesn't mean that we may write each of these w_i as a linear combination of the others. It may even be that some single element w forms a linearly dependent set since there may be a nonzero r such that $rw = 0$ (such a w is said to have **torsion**).

If a linearly independent set $\{w_1, ..., w_k\}$ is maximal in size then we say that the module has **rank** k. Another strange possibility is that a maximal linearly independent set may not be a generating set for the module and hence may not be a basis in the sense to be defined below. The point is that although for an arbitrary $w \in \mathrm{W}$ we must have that $\{w_1, ..., w_k\} \cup \{w\}$ is linearly dependent and hence there must be a nontrivial expression $rw + r_1 w_1 + \cdots + r_k w_k = 0$, it does **not** follow that we may solve for w since r may not be an invertible element of the ring (i.e. it may not be a **unit**).

Definition 6.10 *If B is a generating set for a module* W *such that every element of* W *has a* unique *expression as a* finite R*-linear combination of elements of B then we say that B is a **basis** for* W.

Definition 6.11 *If an* R*-module has a basis then it is referred to as a **free module**.*

It turns out that just as for vector spaces the cardinality of a basis for a free module W is the same as that of every other basis for W. If a module over a (commutative) ring R has a basis then the number of elements in the basis is called the **dimension** and must in this case be the same as the rank (the size of a maximal linearly independent set). If a module W is free with basis $w_1, ..., w_n$ then we have an isomorphism $\mathsf{R}^n \cong \mathrm{W}$ given by

$$(r_1, ..., r_n) \mapsto r_1 w_1 + \cdots + r_n w_n.$$

Exercise 6.4 *Show that every finitely generated* R*-module is the homomorphic image of a free module.*

If R is a field then every module is free and is a vector space by definition. In this case, the current definitions of dimension and basis coincide with the usual ones.

The ring R is itself a free R-module with standard basis given by $\{1\}$. Also, $\mathsf{R}^n := \mathsf{R} \times \cdots \times \mathsf{R}$ is a free module with standard basis $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$ where, as usual $\mathbf{e}_i := (0, ..., 1, ..., 0)$; the only nonzero entry being in the i-th position. Up to isomorphism, these account for all free modules.

As we mentioned, $C^r(M)-$modules are of particular interest. But the ring $C^r(M)$ is also a vector space over $\mathbb{R}$ and this means we have one more layer of structure:

Definition 6.12 *Let* k *be a commutative ring, for example a field such as* $\mathbb{R}$ *or* $\mathbb{C}$*. A ring* A*is called a* k***-algebra*** *if there is a ring homomorphism* $\mu : \mathsf{k} \to \mathsf{A}$ *such that the image* $\mu(\mathsf{k})$ *consists of elements that commute with everything in* A*. In particular,* A *is a module over* k*.*

Often we have $\mathsf{k} \subset \mathsf{A}$ and μ is inclusion.

Example 6.11 *The ring $C^r(M)$ is an $\mathbb{R}$-algebra.*

Because of this example we shall consider A-modules where A is an algebra over some k. In this context the elements of A are still called scalars but the elements of k will be referred to as **constants**.

Example 6.12 *The set $\mathfrak{X}_M(U)$ of vector fields defined on an open set $U$is a vector space over $\mathbb{R}$ but it is also a module over the $\mathbb{R}$-algebra $C^\infty(U)$. So for all $X, Y \in \mathfrak{X}_M(U)$ and all $f, g \in C^\infty(U)$ we have*

1. $f(X + Y) = fX + fY$
2. $(f + g)X = fX + gX$
3. $f(gX) = (fg)X$

Similarly, $\mathfrak{X}^*_M(U) = \Gamma(U, T^*M)$ is also a module over $C^\infty(U)$ that is naturally identified with the module dual $\mathfrak{X}_M(U)^*$ by the pairing $(\theta, X) \mapsto \theta(X)$. Here $\theta(X) \in C^\infty(U)$ and is defined by $p \mapsto \theta_p(X_p)$. The set of all vector fields $\mathcal{Z} \subset \mathfrak{X}(U)$ that are zero at a fixed point $p \in U$ is a submodule in $\mathfrak{X}(U)$. If $U, (x^1, ..., x^n)$ is a coordinate chart then the set of vector fields

$$\frac{\partial}{\partial x^1}, ..., \frac{\partial}{\partial x^n}$$

is a basis (over $C^\infty(U)$) for the module $\mathfrak{X}(U)$. Similarly,

$$dx^1, ..., dx^n$$

is a basis for $\mathfrak{X}^*_M(U)$. It is important to realize that if U is not the domain of a coordinate then it may be that $\mathfrak{X}_M(U)$ and $\mathfrak{X}_M(U)^*$ have no basis. In particular, we should not expect $\mathfrak{X}(M)$ to have a basis in the general case.

The sections of any vector bundle over a manifold M form a $C^\infty(M)$-module denoted $\Gamma(E)$. Let $E \to M$ be a trivial vector bundle of finite rank n. Then there exists a basis of vector fields $e_1, ..., e_n$ for the module $\Gamma(E)$. Thus for any section X there exist unique functions f^i such that

$$X = \sum f^i e_i.$$

In fact, since E is trivial, we may as well assume that $E = M \times \mathbb{R}^n \overset{pr_1}{\to} M$. Then for any basis $u_1, ..., u_n$ for $\mathbb{R}^n$ we may take

$$e_i(x) := (x, u_i).$$

(The e_i form a "local frame field").

Definition 6.13 *Let V_i, $i = 1, ..., k$ and W be modules over a ring R. A map $\mu : \mathrm{V}_1 \times \cdots \times \mathrm{V}_k \to \mathrm{W}$ is called* ***multilinear*** *(k-multilinear) if for each i,*

$1 \leq i \leq k$ *and each fixed* $(w_1, ..., \widehat{w_i}, ..., w_k) \in \mathrm{V}_1 \times \cdots \times \widehat{\mathrm{V}_1} \times \cdots \times \mathrm{V}_k$ *we have that the map*

$$v \mapsto \mu(w_1, ..., \underset{i-th}{v}, ..., w_k),$$

obtained by fixing all but the i-th variable, is a module homomorphism. In other words, we require that μ *be* R- *linear in each slot separately. The set of all multilinear maps* $\mathrm{V}_1 \times \cdots \times \mathrm{V}_k \to \mathrm{W}$ *is denoted* $L_{\mathsf{R}}(\mathrm{V}_1, ..., \mathrm{V}_k; \mathrm{W})$*. If* $\mathrm{V}_1 = \cdots = \mathrm{V}_k = \mathrm{V}$ *then we abbreviate this to* $L^k_{\mathsf{R}}(\mathrm{V}; \mathrm{W})$*.*

If R is commutative then the space of multilinear maps $L_{\mathsf{R}}(\mathrm{V}_1, ..., \mathrm{V}_k; \mathrm{W})$ is itself an R-module in a fairly obvious way. For $a, b \in \mathsf{R}$ and $\mu_1, \mu_2 \in L_{\mathsf{R}}(\mathrm{V}_1, ..., \mathrm{V}_k; \mathrm{W})$ then $a\mu_1 + b\mu_2$ is defined in the usual way.

Let us agree to use the following abbreviation: $\mathrm{V}^k = \mathrm{V} \times \cdots \times \mathrm{V}$ ($k-$fold Cartesian product).

Definition 6.14 *The dual of an* R$-$*module* W *is the module* $\mathrm{W}^* := \mathrm{Hom}_{\mathsf{R}}(\mathrm{W},\mathsf{R})$ *of all* R*-linear functionals on* W*.*

Any element of W can be thought of as an element of $\mathrm{W}^{**} := \mathrm{Hom}_{\mathsf{R}}(\mathrm{W}^*,\mathsf{R})$ according to $w(\alpha) := \alpha(w)$. This provides a map $\mathrm{W} \hookrightarrow \mathrm{W}^{**}$ and if this map is an isomorphism then we say that W is **reflexive** .

If W is reflexive then we are free to identify W with W^{**}.

Exercise 6.5 *Show that if* W *is a free module with finite dimension then* W *is reflexive.*

We sometimes write $w\lrcorner(\alpha) = \langle w, \alpha \rangle = \langle \alpha, w \rangle$.

6.3.1 Tensor Product

Suppose now that we have two R-modules V_1 and V_2. Consider the category $\mathcal{C}_{\mathrm{V}_1 \times \mathrm{V}_2}$ whose objects are bilinear maps $\mathrm{V}_1 \times \mathrm{V}_2 \to \mathrm{W}$ where W varies over all R-modules but V_1 and V_2 are fixed. A morphism from, say $\mu_1 : \mathrm{V}_1 \times \mathrm{V}_2 \to \mathrm{W}_1$ to $\mu_2 : \mathrm{V}_1 \times \mathrm{V}_2 \to \mathrm{W}_2$ is defined to be a map $\ell : \mathrm{W}_1 \to \mathrm{W}_2$ such that the diagram

$$\begin{tikzcd} & \mathrm{W}_1 \arrow[dd, "\ell"] \\ \mathrm{V}_1 \times \mathrm{V}_2 \arrow[ur, "\mu_1"] \arrow[dr, "\mu_2"'] & \\ & \mathrm{W}_2 \end{tikzcd}$$

commutes.

Now we come to a main point: Suppose that there is an R-module $\mathrm{T}_{\mathrm{V}_1,\mathrm{V}_2}$ together with a bilinear map $\otimes : \mathrm{V}_1 \times \mathrm{V}_2 \to \mathrm{T}_{\mathrm{V}_1,\mathrm{V}_2}$ that has the following universal property for this category: For every bilinear map $\mu : \mathrm{V}_1 \times \mathrm{V}_2 \to \mathrm{W}$

there is a unique linear map $\widetilde{\mu} : \mathrm{T}_{\mathrm{V}_1,\mathrm{V}_2} \to \mathrm{W}$ such that the following diagram commutes:

$$\begin{array}{ccc} \mathrm{V}_1 \times \mathrm{V}_2 & \overset{\mu}{\to} & \mathrm{W} \\ \otimes \downarrow & \nearrow_{\widetilde{\mu}} & \\ \mathrm{T}_{\mathrm{V}_1,\mathrm{V}_2} & & \end{array}$$

If such a pair $\mathrm{T}_{\mathrm{V}_1,\mathrm{V}_2}, \otimes$ exists with this property then it is unique up to isomorphism in $\mathcal{C}_{\mathrm{V}_1 \times \mathrm{V}_2}$. In other words, if $\widehat{\otimes} : \mathrm{V}_1 \times \mathrm{V}_2 \to \widehat{\mathrm{T}}_{\mathrm{V}_1,\mathrm{V}_2}$ is another object with this universal property then there is a module isomorphism $\mathrm{T}_{\mathrm{V}_1,\mathrm{V}_2} \cong \widehat{\mathrm{T}}_{\mathrm{V}_1,\mathrm{V}_2}$ such that the following diagram commutes:

$$\begin{array}{ccc} & & \mathrm{T}_{\mathrm{V}_1,\mathrm{V}_2} \\ & \nearrow^{\otimes} & \\ \mathrm{V}_1 \times \mathrm{V}_2 & & \cong\downarrow \\ & \searrow^{\widehat{\otimes}} & \\ & & \widehat{\mathrm{T}}_{\mathrm{V}_1,\mathrm{V}_2} \end{array}$$

We refer to any such universal object as a tensor product of V_1 and V_2. We will indicate the construction of a specific tensor product that we denote by $\mathrm{V}_1 \otimes \mathrm{V}_2$ with corresponding map $\otimes : \mathrm{V}_1 \times \mathrm{V}_2 \to \mathrm{V}_1 \otimes \mathrm{V}_2$. This will be *the* tensor product. The idea is simple: We let $\mathrm{V}_1 \otimes \mathrm{V}_2$ be the set of all linear combinations of symbols of the form $v_1 \otimes v_2$ for $v_1 \in \mathrm{V}_1$ and $v_2 \in \mathrm{V}_2$, subject to the relations

$$\begin{aligned} (v_1 + v_2) \otimes w &= v_1 \otimes w + v_2 \otimes w \\ w \otimes (v_1 + v_2) &= w \otimes v_1 + w \otimes v_2 \\ r\,(v \otimes w) &= rv \otimes w = v \otimes rw \end{aligned}$$

The map $\otimes$ is then simply $\otimes : (v_1, v_2) \to v_1 \otimes v_2$. More generally, we seek a universal object for k-multilinear maps $\mu : \mathrm{V}_1 \times \cdots \times \mathrm{V}_k \to \mathrm{W}$.

Definition 6.15 *A module* $\mathrm{T} = \mathrm{T}_{\mathrm{V}_1,\cdots,\mathrm{V}_k}$ *together with a multilinear map* $\otimes : \mathrm{V}_1 \times \cdots \times \mathrm{V}_k \to \mathrm{T}$ *is called* ***universal for k-multilinear maps on*** $\mathrm{V}_1 \times \cdots \times \mathrm{V}_k$ *if for every multilinear map* $\mu : \mathrm{V}_1 \times \cdots \times \mathrm{V}_k \to \mathrm{W}$ *there is a unique linear map* $\widetilde{\mu} : \mathrm{T} \to \mathrm{W}$ *such that the following diagram commutes:*

$$\begin{array}{ccc} \mathrm{V}_1 \times \cdots \times \mathrm{V}_k & \overset{\mu}{\to} & \mathrm{W} \\ \otimes \downarrow & \nearrow_{\widetilde{\mu}} & \\ \mathrm{T} & & \end{array} ,$$

i.e. we must have $\mu = \widetilde{\mu} \circ \otimes$. *If such a universal object exists it will be called a* ***tensor product*** *of* $\mathrm{V}_1, \ldots, \mathrm{V}_k$ *and the module itself* $\mathrm{T} = \mathrm{T}_{\mathrm{V}_1,\cdots,\mathrm{V}_k}$ *is also referred to as a tensor product of the modules* $\mathrm{V}_1, \ldots, \mathrm{V}_k$.

The tensor product is again unique up to isomorphism and the usual specific realization of the tensor product of modules $\mathrm{V}_1, \ldots, \mathrm{V}_k$ is the set of all linear combinations of symbols of the form $v_1 \otimes \cdots \otimes v_k$ subject to the obvious multilinear relations:

$$(\mathrm{v}_1 \otimes \cdots \otimes a\mathrm{v}_i \otimes \cdots \otimes \mathrm{v}_k = a(\mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_i \otimes \cdots \otimes \mathrm{v}_k)$$

$$\begin{aligned}&(\mathrm{v}_1 \otimes \cdots \otimes (\mathrm{v}_i + \mathrm{w}_i) \otimes \cdots \otimes \mathrm{v}_k)\\ &= \mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_i \otimes \cdots \otimes \mathrm{v}_k + \mathrm{v}_1 \otimes \cdots \otimes \mathrm{w}_i \otimes \cdots \otimes \mathrm{v}_k\end{aligned}$$

This space is denoted $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$ which we also write as $\bigotimes_{i=1}^k \mathrm{V}_i$. Also, we will use $\mathrm{V}^{\otimes k}$ to denote $\mathrm{V} \otimes \cdots \otimes \mathrm{V}$ ($k-$fold tensor product of V).

Proposition 6.3 *If $f : \mathrm{V}_1 \to \mathrm{W}_1$ and $f : \mathrm{V}_1 \to \mathrm{W}_1$ are module homomorphisms then their is a unique homomorphism, the tensor product $f \otimes g : \mathrm{V}_1 \otimes \mathrm{V}_2 \to \mathrm{W}_1 \otimes \mathrm{W}_2$ which has the characterizing properties that $f \otimes g$ be linear and that $f \otimes g(v_1 \otimes v_2) = (fv_1) \otimes (gv_2)$ for all $v_1 \in \mathrm{V}_1, v_2 \in \mathrm{V}_2$. Similarly, if $f_i : \mathrm{V}_i \to \mathrm{W}_i$ we may obtain $\otimes_i f_i : \bigotimes_{i=1}^k \mathrm{V}_i \to \bigotimes_{i=1}^k \mathrm{W}_i$.*

Proof. exercise. ■

Definition 6.16 *Elements of $\bigotimes_{i=1}^k \mathrm{V}_i$ that may be written as $\mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_k$ for some v_i, are called* ***decomposable*** .

Exercise 6.6 *Not all elements are decomposable but the decomposable elements generate $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$.*

It may be that the V_i may be modules over more that one ring. For example, any complex vector space is a module over both $\mathbb{R}$ and $\mathbb{C}$. Also, the module of smooth vector fields $\mathfrak{X}_M(U)$ is a module over $C^\infty(U)$ and a module (actually a vector space) over $\mathbb{R}$. Thus it is sometimes important to indicate the ring involved and so we write the tensor product of two R-modules V and W as $\mathrm{V} \otimes_{\mathsf{R}} \mathrm{W}$. For instance, there is a big difference between $\mathfrak{X}_M(U) \otimes_{C^\infty(U)} \mathfrak{X}_M(U)$ and $\mathfrak{X}_M(U) \otimes_{\mathbb{R}} \mathfrak{X}_M(U)$.

Lemma 6.1 *There are the following natural isomorphisms:*
1) $(\mathrm{V} \otimes \mathrm{W}) \otimes \mathrm{U} \cong \mathrm{V} \otimes (\mathrm{W} \otimes \mathrm{U}) \cong \mathrm{V} \otimes \mathrm{W} \otimes \mathrm{U}$ *and under these isomorphisms* $(\mathrm{v} \otimes \mathrm{w}) \otimes \mathrm{u} \longleftrightarrow \mathrm{v} \otimes (\mathrm{w} \otimes \mathrm{u}) \longleftrightarrow \mathrm{v} \otimes \mathrm{w} \otimes \mathrm{u}$.
2) $\mathrm{V} \otimes \mathrm{W} \cong \mathrm{W} \otimes \mathrm{V}$ *and under this isomorphism* $\mathrm{v} \otimes \mathrm{w} \longleftrightarrow \mathrm{w} \otimes \mathrm{v}$.

Proof. We prove (1) and leave (2) as an exercise.

Elements of the form $(\mathrm{v} \otimes \mathrm{w}) \otimes \mathrm{u}$ generate $(\mathrm{V} \otimes \mathrm{W}) \otimes \mathrm{U}$ so any map that sends $(\mathrm{v} \otimes \mathrm{w}) \otimes \mathrm{u}$ to $\mathrm{v} \otimes (\mathrm{w} \otimes \mathrm{u})$ for all $\mathrm{v}, \mathrm{w}, \mathrm{u}$ must be unique. Now we have compositions

$$(\mathrm{V} \times \mathrm{W}) \times \mathrm{U} \overset{\otimes \times \mathrm{id}_\mathrm{U}}{\to} (\mathrm{V} \otimes \mathrm{W}) \times \mathrm{U} \overset{\otimes}{\to} (\mathrm{V} \otimes \mathrm{W}) \otimes \mathrm{U}$$

and

$$\mathrm{V} \times (\mathrm{W} \times \mathrm{U}) \overset{\mathrm{id}_\mathrm{U} \times \otimes}{\to} (\mathrm{V} \times \mathrm{W}) \otimes \mathrm{U} \overset{\otimes}{\to} \mathrm{V} \otimes (\mathrm{W} \otimes \mathrm{U}).$$

It is a simple matter to check that these composite maps have the same universal property as the map $\mathrm{V} \times \mathrm{W} \times \mathrm{U} \stackrel{\otimes}{\to} \mathrm{V} \otimes \mathrm{W} \otimes \mathrm{U}$. The result now follows from the existence and essential uniqueness results proven so far (D.1 and D.1). ■

We shall use the first isomorphism and the obvious generalizations to identify $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$ with all legal parenthetical constructions such as $(((\mathrm{V}_1 \otimes \mathrm{V}_2) \otimes \cdots \otimes \mathrm{V}_j) \otimes \cdots) \otimes \mathrm{V}_k$ and so forth. In short, we may construct $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$ by tensoring spaces two at a time. In particular we assume the isomorphisms (as identifications)

$$(\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k) \otimes (\mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_k) \cong \mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k \otimes \mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_k$$

which map $(\mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_k) \otimes (\mathrm{w}_1 \otimes \cdots \otimes \mathrm{w}_k)$ to $\mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_k \otimes \mathrm{w}_1 \otimes \cdots \otimes \mathrm{w}_k$.

The following proposition gives a basic and often used isomorphism:

Proposition 6.4 *For* $\mathsf{R}-$*modules* $\mathrm{W}, \mathrm{V}, \mathrm{U}$ *we have*

$$\mathrm{Hom}_{\mathsf{R}}(\mathrm{W} \otimes \mathrm{V}, \mathrm{U}) \cong L(\mathrm{W}, \mathrm{V}; \mathrm{U})$$

More generally,

$$\mathrm{Hom}_{\mathsf{R}}(\mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_k, \mathrm{U}) \cong L(\mathrm{W}_1, ..., \mathrm{W}_k; \mathrm{U})$$

Proof. This is more or less just a restatement of the universal property of $\mathrm{W} \otimes \mathrm{V}$. One should check that this association is indeed an isomorphism. ■

Exercise 6.7 *Show that if* W *is free with basis* $(f_1, ..., f_n)$ *then* W^* *is also free and has a dual basis* $f^1, ..., f^n$*, that is,* $f^i(f_j) = \delta^i_j$.

Theorem 6.2 *If* $\mathrm{V}_1, ..., \mathrm{V}_k$ *are free* $\mathsf{R}-$*modules and if* $(v^j_1, ..., v^j_{n_j})$ *is a basis for* V_j *then set of all decomposable elements of the form* $v^1_{i_1} \otimes \cdots \otimes v^k_{i_k}$ *form a basis for* $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$.

Proposition 6.5 *There is a unique* R*-module map* $\iota : \mathrm{W}_1^* \otimes \cdots \otimes \mathrm{W}_k^* \to (\mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_k)^*$ *such that if* $\alpha_1 \otimes \cdots \otimes \alpha_k$ *is a (decomposable) element of* $\mathrm{W}_1^* \otimes \cdots \otimes \mathrm{W}_k^*$ *then*

$$\iota(\alpha_1 \otimes \cdots \otimes \alpha_k)(w_1 \otimes \cdots \otimes w_k) = \alpha_1(w_1) \cdots \alpha_k(w_k).$$

If the modules are all free then this is an isomorphism.

Proof. If such a map exists, it must be unique since the decomposable elements span $\mathrm{W}_1^* \otimes \cdots \otimes \mathrm{W}_k^*$. To show existence we define a multilinear map

$$\vartheta : \mathrm{W}_1^* \times \cdots \times \mathrm{W}_k^* \times \mathrm{W}_1 \times \cdots \times \mathrm{W}_k \to \mathsf{R}$$

by the recipe

$$(\alpha_1, ..., \alpha_k, w_1, ..., w_k) \mapsto \alpha_1(w_1) \cdots \alpha_k(w_k).$$

By the universal property there must be a linear map

$$\widetilde{\vartheta} : \mathrm{W}_1^* \otimes \cdots \otimes \mathrm{W}_k^* \otimes \mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_k \to \mathsf{R}$$

such that $\widetilde{\vartheta} \circ \otimes = \vartheta$ where $\otimes$ is the universal map. Now define

$$\begin{aligned}&\iota(\alpha_1 \otimes \cdots \otimes \alpha_k)(w_1 \otimes \cdots \otimes w_k)\\&:= \widetilde{\vartheta}(\alpha_1 \otimes \cdots \otimes \alpha_k \otimes w_1 \otimes \cdots \otimes w_k).\end{aligned}$$

The fact, that ι is an isomorphism in case the W_i are all free follows easily from exercise **??** and theorem D.2. Once we view an element of W_i as a functional from $\mathrm{W}_i^{**} = L(\mathrm{W}_i^*; \mathsf{R})$ we see that the effect of this isomorphism is to change the interpretation of the tensor product to the "map" tensor product in $(\mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_k)^* = L(\mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_k; \mathsf{R})$. Thus the basis elements match up under ι. ∎

6.4 Tensoriality

In this and the following few sections we introduce some formalism that will not only provide some convenient language but also provides conceptual tools. The study of the objects we introduce here is a whole area of mathematics and may be put to uses far more sophisticated than we do here.

There is an interplay in geometry and topology between local and global data. To set up our discussion, suppose that σ is a section of a smooth vector bundle $\pi : E \to M$. Let us focus our attention near a point p which is contained in an open set U over which the bundle is trivialized. The trivialization provides a local frame field $\{e_1, ..., e_k\}$ on U. A section σ has a local expression $\sigma = \sum s^i e_i$ for some smooth functions s^i. Now the component functions s^i together give a smooth map $(s_i) : U \to \mathbb{R}^k$. We may assume without loss of generality that U is the domain of a chart $\mathtt{x}$ for the manifold M. The map $(s_i) \circ \mathtt{x}^{-1} : \mathtt{x}(U) \to \mathbb{R}^k$ has a Taylor expansion centered at $\mathtt{x}(p)$. It will be harmless to refer to this as the Taylor expansion around p. Now, we say that two sections σ_1 and σ_2 have the "same k-jet at p" if in any chart these two sections have Taylor expansions which agree up to and including terms of order k. This puts an equivalence relation on sections defined near p.

Consider two vector bundles $E_1 \to M$ and $E_2 \to M$. Suppose we have a map $F : \Gamma(M, E_1) \to \Gamma(M, E_2)$ that is not necessarily linear over $C^\infty(M)$ or even $\mathbb{R}$. We say that F is **local** if the support of $F(\sigma)$ is contained in the support of σ. We can ask for more refined kinds of locality. For any $\sigma \in \Gamma(M, E_1)$ we have a section $F(\sigma)$ and its value $F(\sigma)(p) \in E_2$ at some point $p \in M$. What determines the value $F(\sigma)(p)$? Let us consider in turn the following situations:

1. It just might be the case that whenever two sections σ_1 and σ_2 agree on some neighborhood of p then $F(\sigma_1)(p) = F(\sigma_2)(p)$. So all that matters for determining $F(\sigma)(p)$ is the behavior of σ in any arbitrarily small open set containing p. To describe this we say that $F(\sigma)(p)$ only depends on the "germ" of σ at p.

2. Certainly if σ_1 and σ_2 agree on some neighborhood of p then they both have the same Taylor expansion at p (as seen in any local trivializations).

The reverse is not true however. Suppose that whenever two section σ_1 and σ_2 have Taylor series that agree up to and including terms of order k then $\digamma(\sigma_1)(p) = \digamma(\sigma_2)(p)$. Then we say that $(\digamma(\sigma))(p)$ depends only on the $k-$jet of σ at p.

3. Finally, it might be the case that $\digamma(\sigma_1)(p) = \digamma(\sigma_2)(p)$ exactly when $\sigma_1(p) = \sigma_2(p)$.

Of course it is also possible that none of the above hold at any point. Notice that as we go down the list we are saying that the information needed to determine $\digamma(\sigma)(p)$ is becoming more and more local; even to the point of being infinitesimal and then even determine by what happen just at a point. A vector field can be viewed as an $\mathbb{R}$-linear map $C^\infty(M) \to C^\infty(M)$ and since $Xf(p) = Xg(p)$ exactly when $df(p) = dg(p)$ we see that $Xf(p)$ depends only on the $1-$jet of f at p. But this cannot be the whole story since two functions might have the same differential without sharing the same 1-jet since they might not agree at the 0-th jet level (they may differ by a constant near p).

We now restrict our attention to (local) $\mathbb{R}$-linear maps $L : \Gamma(M, E_1) \to \Gamma(M, E_2)$. We start at the bottom, so to speak, with $0-$th order operators. One way that $0-$th order operators arise is from bundle maps. If $\tau : E_1 \to E_2$ is a bundle map (over the identity $M \to M$) then we get an induced map $\Gamma\tau : \Gamma(M, E_1) \to \Gamma(M, E_2)$ on the level of sections: If $\sigma \in \Gamma(M, E_1)$ then

$$\Gamma\tau(\sigma) := \sigma \circ \tau$$

Notice the important property that $\Gamma\tau(f\sigma) = f\Gamma\tau(\sigma)$ and so $\Gamma\tau$ is $C^\infty(M)$ linear (a module homomorphism). Conversely, if $L : \Gamma(M, E_1) \to \Gamma(M, E_2)$ is $C^\infty(M)$ linear then $(L\sigma)(p)$ depends only on the value of $\sigma(p)$ and as we see presently this means that L determines a bundle τ map such that $\Gamma\tau = L$. We shall prove a bit more general result which extends to multi-linear maps. We must assume that M supports cut-off functions 4.6.

Proposition 6.6 *Let $p \in M$ and $\tau : \Gamma(M, E_1) \times \cdots \times \Gamma(M, E_N) \to \Gamma(M, E)$ be a $C^\infty(M)-$multilinear map. Let $\sigma_i, \bar{\sigma}_i \in \Gamma(M, E_i)$ smooth sections such that $\sigma_i(p) = \bar{\sigma}_i(p)$ for $1 \le i \le N$; Then we have that*

$$\tau(\sigma_1, ..., \sigma_N)(p) = \tau(\bar{\sigma}_1, ..., \bar{\sigma}_N)(p)$$

Proof. The proof will follow easily if we can show that $\tau(\sigma_1, ..., \sigma_N)(p) = 0$ whenever one of $\sigma_i(p)$ is zero. We shall assume for simplicity of notation that $N = 3$. Now suppose that $\sigma_1(p) = 0$. If $e_1, ..., e_N$ is a frame field over $U \subset M$ (with $p \in U$) then $\sigma_1|_U = \sum s^i e_i$ for some smooth functions $s^i \in C^\infty(U)$. Let β be a bump function with support in U. Then $\beta\, \sigma_1|_U$ and $\beta^2\, \sigma_1|_U$ extend by zero to elements of $\Gamma(M, E_1)$which we shall denote by $\beta\sigma_1$ and $\beta^2\sigma_1$. Similarly, βe_i and $\beta^2 e_i$ are globally defined sections, βs^i is a global function and $\beta\sigma_1 =$

$\sum \beta s^i \beta e_i$. Thus

$$\begin{aligned}\beta^2 \tau(\sigma_1, \sigma_2, \sigma_3) &= \tau(\beta^2 \sigma_1, \sigma_2, \sigma_3) \\ &= \tau(\sum \beta s^i \beta e_i, \sigma_2, \sigma_3) \\ &= \sum \beta \tau(s^i \beta e_i, \sigma_2, \sigma_3).\end{aligned}$$

Now since $\sigma_1(p) = 0$ we must have $s^i(p) = 0$. Also recall that $\beta(p) = 1$. Plugging p into the formula above we obtain $\tau(\sigma_1, \sigma_2, \sigma_3)(p) = 0$.

A similar argument holds when $\sigma_2(p) = 0$ or $\sigma_3(p) = 0$.

Assume that $\sigma_i(p) = \bar{\sigma}_i(p)$ for $1 \leq i \leq 3$. Then we have

$$\begin{aligned}&\tau(\bar{\sigma}_1, \bar{\sigma}_2, \bar{\sigma}_3) - \tau(\sigma_1, \sigma_2, \sigma_3) \\ &= \tau(\bar{\sigma}_1 - \sigma_1, \bar{\sigma}_2, \bar{\sigma}_3) + \tau(\sigma_1, \sigma_2, \sigma_3) + \tau(\sigma_1, \bar{\sigma}_2 - \sigma_2, \sigma_3) \\ &+ \tau(\bar{\sigma}_1, \sigma_2, \bar{\sigma}_3) + \tau(\bar{\sigma}_1, \sigma_2, \bar{\sigma}_3 - \sigma_3)\end{aligned}$$

Since $\bar{\sigma}_1 - \sigma_1, \bar{\sigma}_2 - \sigma_2$ and $\bar{\sigma}_3 - \sigma_3$ are all zero at p we obtain the result. ■

By the above, linearity over $C^\infty(M)$ on the level of sections corresponds to bundle maps on the vector bundle level. Thus whenever we have a $C^\infty(M)-$multilinear map $\tau : \Gamma(M, E_1) \times \cdots \times \Gamma(M, E_N) \to \Gamma(M, E)$ we also have an $\mathbb{R}-$multilinear map $E_{1p} \times \cdots \times E_{Np} \to E_p$ (which we shall often denote by the symbol τ_p):

$$\tau_p(v_1, ..., v_N) := \tau(\sigma_1, ..., \sigma_N)(p) \text{ for any sections } \sigma_i \text{ with } \sigma_i(p) = v_i$$

The individual maps $E_{1p} \times \cdots \times E_{Np} \to E_p$ combine to give a vector bundle morphism

$$E_1 \oplus \cdots \oplus E_N \to E$$

If an $\mathbb{R}$-multilinear map $\tau : \Gamma(M, E_1) \times \cdots \times \Gamma(M, E_N) \to \Gamma(M, E)$ is actually $C^\infty(M)-$linear in one or more variable then we say that τ is **tensorial** in those variables. If τ is tensorial in all variables say that τ is **tensorial**.

It is worth our while to look more closely at the case $N = 1$ above. Suppose that $\tau : \Gamma(M, E) \to \Gamma(M, F)$ is tensorial ($C^\infty-$linear). Then for each p we get a linear map $\tau_p : E_p \to F_p$ and so a bundle morphism $E \to F$ but also we may consider the assignment $p \mapsto \tau_p$ as a section of the bundle whose fiber at p is $\hom(E_p, F_p)$. This bundle is denoted $\mathrm{Hom}(E, F)$ and is isomorphic to the bundle $F \otimes E^*$.

6.5 Jets and Jet bundles

A map $f :: (\mathsf{E}, x) \to (\mathsf{F}, y)$ is said to have k-th order contact at x with a map $g :: (\mathsf{E}, x) \to (\mathsf{F}, y)$ if f and g have the same Taylor polynomial of order k at x. This notion defines an equivalence class on the set of (germs) of functions $\mathsf{E}, x \to \mathsf{F}, y$ and the equivalence classes are called k-jets. The equivalence class of a function f is denoted by $j^k_x f$ and the space of k-jets of maps $(\mathsf{E}, x) \to (\mathsf{F}, y)$

is denoted $J^k_x(\mathsf{E},\mathsf{F})_y$. We have the following disjoint unions:

$$J^k_x(\mathsf{E},\mathsf{F}) = \bigcup_{y\in\mathsf{F}} J^k_x(\mathsf{E},\mathsf{F})_y$$
$$J^k(\mathsf{E},\mathsf{F})_y = \bigcup_{x\in\mathsf{E}} J^k_x(\mathsf{E},\mathsf{F})_y$$
$$J^k(\mathsf{E},\mathsf{F}) = \bigcup_{x\in\mathsf{E},y\in\mathsf{F}} J^k_x(\mathsf{E},\mathsf{F})_y$$

Now for a (germ of a) smooth map $f :: (M,x) \to (N,y)$ we say that f has k-th order contact at x with a another map $g :: (M,x) \to (N,y)$ if $\phi \circ f \circ \psi$ has k-th order contact at $\psi(x)$ with $\phi\circ\ g \circ \psi$ for some (and hence all) charts ϕ and ψ defined in a neighborhood of $x \in M$ and $y \in N$ respectively. We can then define sets $J^k_x(M,N)$, $J^k(M,N)_y$, and $J^k(M,N)$. The space $J^k(M,N)$ is called the space of jets of maps from M to N and turns out to be a smooth vector bundle. In fact, a pair of charts as above determines a chart for $J^k(M,N)$ defined by

$$J^k(\psi,\phi) : j^k_x f \mapsto (\phi(x), \psi(f(x)), D(\phi\circ f\circ\psi), ..., D^k(\phi\circ f\circ\psi))$$

where $D^j(\phi\circ f\circ\psi) \in L^j_{sym}(\mathsf{M},\mathsf{N})$. In finite dimensions, the chart looks like

$$J^k(\psi,\phi) : j^k_x f \mapsto (x^1, ..., x^n, y^1, ..., y^n, \frac{\partial f^i}{\partial x^j}(x), ..., \frac{\partial^\alpha f^i}{\partial x^\alpha}(x))$$

where $|\alpha| = k$. Notice that the chart has domain $U_\psi \times U_\phi$.

Exercise 6.8 *What vector space is naturally the typical fiber of this vector bundle?*

Definition 6.17 *The k-**jet extension** of a map $f : M \to N$ is defined by $j^k f : x \mapsto j^k_x f \in J^k(M,N)$. There is a strong transversality theorem that uses the concept of a jet space:*

Theorem 6.3 (Strong Transversality) *Let S be a closed submanifold of $J^k(M,N)$. Then the set of maps such that the k-jet extensions $j^k f$ are transverse to S is an open everywhere dense subset in $C^\infty(M,N)$.*

6.6 Sheaves

Now we have seen that the section $\Gamma(M,E)$ of a vector bundle form a module over the smooth functions $C^\infty(M)$. It is important to realize that having a vector bundle at hand not only provide a module but a family of them parameterized by the open subsets of M. How are these modules related to each other?

Consider a local section $\sigma : M \to E$ of a vector bundle E. Given any open set $U \subset M$ we may always produce the restricted section $\sigma|_U : U \to E$. This gives us a family of sections; one for each open set U. To reverse the situation, suppose that we have a family of sections $\sigma_U : U \to E$ where U varies over the open sets (or just a cover of M). When is it the case that such a family is just the family of restrictions of some (global) section $\sigma : M \to E$? To help with these kinds of questions and to provide a language that will occasionally be convenient we will introduce another formalism. This is the formalism of sheaves and presheaves. The formalism of sheaf theory is convenient for conveying certain concepts concerning the interplay between the local and the global. This will be our main use. More serious use of sheaf theory is found in cohomology theory and is especially useful in complex geometry. Sheaf theory also provides a very good framework within which to develop the foundations of supergeometry which is an extension of differential geometry that incorporates the important notion of "fermionic variables". A deep understanding of sheaf theory is not necessary for most of what we do here and it would be enough to acquire a basic familiarity with the definitions.

Definition 6.18 *A* ***presheaf of abelian groups (resp. rings etc.)*** *on a manifold (or more generally a topological space) M is an assignment $U \to \mathcal{M}(U)$ to each open set $U \subset M$ together with a family of abelian group homomorphisms (resp. ring homomorphisms etc.) $r_V^U : \mathcal{M}(U) \to \mathcal{M}(V)$ for each nested pair $V \subset U$ of open sets and such that*

Presheaf 1 $r_W^V \circ r_V^U = r_W^U$ whenever $W \subset V \subset U$.

Presheaf 2 $r_V^V = \mathrm{id}_V$ for all open $V \subset M$.

Definition 6.19 *Let $\mathcal{M}$ be a presheaf and $\mathcal{R}$ a presheaf of rings. If for each open $U \subset M$ we have that $\mathcal{M}(U)$ is a module over the ring $\mathcal{R}(U)$ and if the multiplication map $\mathcal{R}(U) \times \mathcal{M}(U) \to \mathcal{M}(U)$ commutes with the restriction maps r_W^U then we say that $\mathcal{M}$ is a presheaf of modules over $\mathcal{R}$.*

Definition 6.20 *A presheaf homomorphism $h : \mathcal{M}_1 \to \mathcal{M}_2$ is an is an assignment to each open set $U \subset M$ an abelian group (resp. ring, module, etc.) morphism $h_U : \mathcal{M}_1(U) \to \mathcal{M}_2(U)$ such that whenever $V \subset U$ then the following diagram commutes:*

$$\begin{array}{ccc} \mathcal{M}_1(U) & \overset{h_U}{\to} & \mathcal{M}_2(U) \\ r_V^U \downarrow & & r_V^U \downarrow \\ \mathcal{M}_1(V) & \overset{h_V}{\to} & \mathcal{M}_2(V) \end{array} .$$

Note we have used the same notation for the restriction maps of both presheaves.

Definition 6.21 *We will call a presheaf a* ***sheaf*** *if the following properties hold whenever $U = \bigcup_{U_\alpha \in \mathcal{U}} U_\alpha$ for some collection of open sets $\mathcal{U}$.*

Sheaf 1 If $s_1, s_2 \in \mathcal{M}(U)$ and $r_{U_\alpha}^U s_1 = r_{U_\alpha}^U s_2$ for all $U_\alpha \in \mathcal{U}$ then $s_1 = s_2$.

Sheaf 2 If $s_\alpha \in \mathcal{M}(U_\alpha)$ and whenever $U_\alpha \cap U_\beta \neq \emptyset$ we have

$$r^U_{U_\alpha \cap U_\beta} s_\alpha = r^U_{U_\alpha \cap U_\beta} s_\beta$$

then there exists $s \in \mathcal{M}(U)$ such that $r^U_{U_\alpha} s = s_\alpha$.

If we need to indicate the space M involved we will write $\mathcal{M}_M$ instead of $\mathcal{M}$.

Definition 6.22 *A morphism of sheaves is a morphism of the underlying presheaf.*

The assignment $C^\infty(.,M) : U \mapsto C^\infty(U)$ is a presheaf of rings. This sheaf will also be denoted by $\mathcal{C}^\infty_M$. The best and most important example of a sheaf of modules over $C^\infty(.,M)$ is the assignment $\Gamma(E,.) : U \mapsto \Gamma(E,U)$ for some vector bundle $E \to M$ and where by definition $r^U_V(s) = s|_V$ for $s \in \Gamma(E,U)$. In other words r^U_V is just the restriction map. Let us denote this (pre)sheaf by $\Gamma_E : U \mapsto \Gamma_E(U) := \Gamma(E,U)$.

Notation 6.1 *Many if not most of the constructions operations we introduce for sections of vector bundles are really also operations appropriate to the (pre)sheaf category. Naturality with respect to restrictions is one of the features that is often not even mentioned (sometime this precisely because it is obvious). This is the inspiration for a slight twist on our notation.*

	Global	local	Sheaf notation
functions on M	$C^\infty(M)$	$C^\infty(U)$	C^∞_M
Vector fields on M	$\mathfrak{X}(M)$	$\mathfrak{X}(U)$	$\mathfrak{X}_M$
Sections of E	$\Gamma(E)$	$\Gamma(U,E)$	$-_E$

Notation 6.2 *where $C^\infty_M : U \mapsto C^\infty_M(U) := C^\infty(U), \mathfrak{X}_M : U \mapsto \mathfrak{X}_M(U) := \mathfrak{X}(U)$ and so on.*

For example, when we say that $D : C^\infty_M \to C^\infty_M$ is a derivation we mean that D is actually a family of algebra derivations $D_U : C^\infty_M(U) \to C^\infty_M(U)$ indexed by open sets U such that we have naturality with respect to restrictions. I.e. the diagrams of the form below for $V \subset U$ commute:

$$\begin{array}{ccc} C^\infty_M(U) & \stackrel{D_U}{\to} & C^\infty_M(U) \\ r^U_V \downarrow & & r^U_V \downarrow \\ C^\infty_M(V) & \stackrel{D_V}{\to} & C^\infty_M(V) \end{array}.$$

It is easy to see that all of the following examples are sheaves. In each case the maps $r^U_{U_\alpha}$ are just the restriction maps.

Example 6.13 (Sheaf of holomorphic functions) *Sheaf theory really shows its strength in complex analysis. This example is one of the most studied. However, we have not studied notion of a complex manifold and so this example is* *for those readers with some exposure to complex manifolds. Let M be a complex manifold and let $\mathcal{O}_M(U)$ be the algebra of holomorphic functions defined on U. Here too, $\mathcal{O}_M$ is a sheaf of modules over itself. Where as the sheaf $\mathcal{C}_M^\infty$ always has global sections, the same is not true for $\mathcal{O}_M$. The sheaf theoretic approach to the study of obstructions to the existence of global sections has been very successful.*

Recall that $s_1 \in \Gamma_E(U)$ and $s_2 \in \Gamma_E(V)$ determine the same germ of sections at p if there is an open set $W \subset U \cap V$ such that $r_W^U s_1 = r_W^V s_2$. Now on the union

$$\bigcup_{p \in U} \Gamma_E(U)$$

we impose the equivalence relation $s_1 \sim s_2$ if and only if s_1 and s_2 determine the same germ of sections at p. The set of equivalence classes (called germs of section at p) is an abelian group in the obvious way and is denoted Γ_p^E. If we are dealing with a sheaf of rings then Γ_p^E is a ring. The set $\Gamma_E((U)) = \bigcup_{p \in U} \Gamma_p^E$ is called the sheaf of germs and can be given a topology so that the projection $pr : \Gamma_E((U)) \to M$ defined by the requirement that $pr([s]) = p$ if and only if $[s] \in \mathcal{S}_p^E$ is a local homeomorphism.

More generally, let $\mathcal{M}$ be a presheaf of abelian groups on M. For each $p \in M$ we define the direct limit group

$$\mathcal{M}_p = \varinjlim_{p \in U} \mathcal{M}(U)$$

with respect to the restriction maps r_V^U.

Definition 6.23 *$\mathcal{M}_p$ is a set of equivalence classes called germs at p. Here $s_1 \in \mathcal{M}(U)$ and $s_2 \in \mathcal{M}(V)$ determine the same germ of sections at p if there is an open set $W \subset U \cap V$ containing p such that $r_W^U s_1 = r_W^V s_2$. The germ of $s \in \mathcal{M}(U)$ at p is denoted s_p.*

Now we take the union $\widetilde{\mathcal{M}} = \bigcup_{p \in M} \mathcal{M}_p$ and define a surjection $\pi : \widetilde{\mathcal{M}} \to M$ by the requirement that $\pi(s_p) = p$ for $s_p \in \mathcal{M}_p$. The space $\widetilde{\mathcal{M}}$ is called the **sheaf of germs** generated by $\mathcal{M}$. We want to topologize $\widetilde{\mathcal{M}}$ so that π is continuous and a local homeomorphism but first a definition.

Definition 6.24 (étalé space) *A topological space Y together with a continuous surjection $\pi : Y \to M$ that is a local homeomorphism is called an* ***étalé space****. A local section of an étalé space over an open subset $U \subset M$ is a map $s_U : U \to Y$ such that $\pi \circ s_U = \mathrm{id}_U$. The set of all such sections over U is denoted $\Gamma(U, Y)$.*

Definition 6.25 *For each $s \in \mathcal{M}(U)$ we can define a map (of sets) $\widetilde{s} : U \to \widetilde{\mathcal{M}}$ by*

$$\widetilde{s}(x) = s_x$$

and we give $\widetilde{\mathcal{M}}$ the smallest topology such that the images $\widetilde{s}(U)$ for all possible U and s are open subsets of $\widetilde{\mathcal{M}}$.

With the above topology $\widetilde{\mathcal{M}}$ becomes an étalé space and all the sections $\widetilde{s}$ are continuous open maps. Now if we let $\widetilde{\mathcal{M}}(U)$ denote the sections over U for this étalé space, then the assignment $U \to \widetilde{\mathcal{M}}(U)$ is a presheaf that is always a sheaf.

Proposition 6.7 *If $\mathcal{M}$ was a sheaf then $\widetilde{\mathcal{M}}$ is isomorphic as a sheaf to $\mathcal{M}$.*

The notion of a germ not only makes sense of sections of a bundle but we also have the following definition:

Definition 6.26 *Let $F_p(M, N)$ be the set of all smooth maps $f :: M, p \to N$ that are locally defined near p. We will say that two such functions f_1 and f_2 have the same **germ** at p if they are equal on some open subset of the intersection of their domains that also contains the point p. In other words they agree in some neighborhood of p. This defines an equivalence relation on $F_p(M, N)$ and the equivalence classes are called germs of maps. The set of all such germs of maps from M to N is denoted by $\mathcal{F}_p(M, N)$.*

It is easy to see that for $[f_1], [f_2] \in \mathcal{F}_p(M, \mathbb{F})$ where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, the product $[f_1][f_2] = [f_1 f_2]$ is well defined as are the linear operations $a[f_1] + b[f_2] = [af_1 + bf_2]$. One can easily see that $\mathcal{F}_p(M, \mathbb{F})$ is a commutative ring. In fact, it is an algebra over $\mathbb{F}$. Notice also that in any case we have a well defined evaluation $[f](p) = f(p)$.

Exercise 6.9 *Reformulate theorem 1.9 in terms of germs of maps.*

6.7 Problem Set

1. Show that $S^n \times \mathbb{R}$ is parallelizable.

2. Let $\pi : E \to M$ be a smooth vector bundle and let 0_p denote the zero element of the fiber E_p. The map $\mathbf{0} : p \to 0_p$ is a section of $\pi : E \to M$. Show that $\mathbf{0}$ is an embedding of M into E. Thus we sometimes identify the image $\mathbf{0}(M) \subset E$ with M itself. Both the map $\mathbf{0}$ and its image are called the **zero section** of $\pi : E \to M$.

Chapter 7

Lie Bracket and Lie Derivative

Lemma 7.1 *Let $U \subset M$ be an open set. If $\mathcal{L}_X\, f = 0$ for all $f \in C^\infty(U)$ then $X|_U = 0$.*

Proof. Working locally in a chart $\mathtt{x}, U$, let X_U be the local representative of X (defined in section 4.7). Suppose $X_U(p) \neq 0$ and that $\ell : \mathsf{E} \to \mathbb{R}$ is a continuous linear map such that $\ell(X_U(p)) \neq 0$ (exists by Hahn-Banach theorem). Let $f := \ell \circ \mathtt{x}$. Then the local representative of $\mathcal{L}_X\, f(p)$ is $d(\ell \circ \mathtt{x})(X(p)) = D\ell|_p \cdot X_U = \ell(X_U(p)) \neq 0$ i.e $\mathcal{L}_X\, f(p) \neq 0$. We have used the fact that $D\ell|_p\, \ell = \ell$ (since ℓ is linear). ■

In the infinite dimensional case we do not know that all derivations are of the form $\mathcal{L}_X$ for some vector field X.

Theorem 7.1 *For every pair of vector fields $X, Y \in \mathfrak{X}(M)$ there is a unique vector field $[X, Y]$ such that for any open set $U \subset M$ and $f \in C^\infty(U)$ we have $[X, Y]_p f = X_p(Yf) - Y_p(Xf)$ and such that in a local chart $(U, \mathtt{x})$ the vector field $[X, Y]$ has as local expression given by*

$$DY_U \cdot X_U - DX_U \cdot Y_U.$$

Expressed more fully, if $f \in C^\infty(U)$, then letting $f_U := f \circ \mathtt{x}^{-1}$ be the local representative we have

$$([X_U, Y_U]f_U)(x) = Df_U \cdot (DY_U(x) \cdot X_U(x) - DX_U(x) \cdot Y_U(x))$$

where $x \sim p$ and $[X_U, Y_U]f_U$ is the local representative of $[X, Y]f$.

Proof. We sketch the main points and let the reader fill in the details. Check that $\mathcal{L}_X \circ \mathcal{L}_Y - \mathcal{L}_Y \circ \mathcal{L}_X$ defines a derivation. One can prove that the formula $f \mapsto Df_U \cdot (DY_U \cdot X_U - DX_U \cdot Y_U)$ defines a derivation locally on U and one may check that this is none other than the local representation with respect to $(U, \mathtt{x})$ of the derivation $\mathcal{L}_X \circ \mathcal{L}_Y - \mathcal{L}_Y \circ \mathcal{L}_X$. Furthermore, this derivation is that given by the vector field on U whose local representative is $DY_U \cdot X_U - DX_U \cdot Y_U$. Now these local vector fields on charts may be directly checked to be related by the correct transformation laws so that we have a well defined global vector field. Thus by pulling back via charts we get a vector field $[X, Y]$ on each chart domain. The global derivation $\mathcal{L}_X \circ \mathcal{L}_Y - \mathcal{L}_Y \circ \mathcal{L}_X$ agrees with $\mathcal{L}_{[X,Y]}$ and so by **??** $[X, Y]$ is uniquely determined.

In the finite dimensional case we can see the result more simply. Namely, every derivation of $C^\infty(M)$ is $\mathcal{L}_X$ for some unique vector field X so we can define $[X, Y]$ as the unique vector field such that $\mathcal{L}_{[X,Y]} = \mathcal{L}_X \circ \mathcal{L}_Y - \mathcal{L}_Y \circ \mathcal{L}_X$. ■

We ought to see what the local formula for the Lie derivative looks like in the finite dimensional case where we may employ classical notation. Suppose we have $X = \sum_{i=1}^m X^i \frac{\partial}{\partial x^i}$ and $Y = \sum_{i=1}^m Y^i \frac{\partial}{\partial x^i}$. Then $[X, Y] = \sum_i \left(\sum_j \frac{\partial Y^i}{\partial x^j} X^j - \frac{\partial X^i}{\partial x^j} Y^j \right) \frac{\partial}{\partial x^i}$

Exercise 7.1 *Check this.*

Definition 7.1 *The vector field $[X, Y]$ from the previous theorem is called the **Lie bracket** of X and Y.*

The following properties of the Lie Bracket are checked by direct calculation. For any $X, Y, Z \in \mathfrak{X}(M)$,

1. $[X, Y] = -[Y, X]$
2. $[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0$

The map $(X, Y) \mapsto [X, Y]$ is linear over $\mathbb{R}$ but not over $C^\infty(M)$. The $\mathbb{R}-$vector space $\mathfrak{X}(M)$ together with the $\mathbb{R}-$bilinear map $(X, Y) \mapsto [X, Y]$ is an example of an extremely important abstract algebraic structure:

Definition 7.2 (Lie Algebra) *A vector space $\mathfrak{a}$ (over a field $\mathbb{F}$) is called a **Lie algebra** if it is equipped with a bilinear map $\mathfrak{a} \times \mathfrak{a} \to \mathfrak{a}$ (a multiplication) denoted $(v, w) \mapsto [v, w]$ such that*

$$[v, w] = -[w, v]$$

and such that we have the Jacobi identity

$$[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0$$

for all $x, y, z \in \mathfrak{a}$.

We shall take a look at the structure theory of Lie algebras in chapter 15.

7.1 Action by pull-back and push forward

Given a diffeomorphism $\phi : M \to N$ we define the **pull-back** $\phi^* Y \in \mathfrak{X}(M)$ for $Y \in \mathfrak{X}(N)$ and the **push-forward** $\phi_* X \in \mathfrak{X}(N)$ of $X \in \mathfrak{X}(M)$ via ϕ by defining

$$\begin{aligned}\phi^* Y &= T\phi^{-1} \circ Y \circ \phi \text{ and} \\ \phi_* X &= T\phi \circ X \circ \phi^{-1}.\end{aligned}$$

In other words, $(\phi^* Y)(p) = T\phi^{-1} \cdot Y_{\phi(p)}$ and $(\phi_* X)(p) = T\phi \cdot X_{\phi^{-1}(p)}$. Notice that $\phi^* Y$ and $\phi_* X$ are both smooth vector fields. Let $\phi : M \to N$ be a smooth map of manifolds. The following commutative diagrams summarize some of the concepts:

Pointwise:

$$\begin{array}{ccc} T_p M & \stackrel{T_p\phi}{\to} & T_{\phi p} N \\ \downarrow & & \downarrow \\ (M,p) & \stackrel{\phi}{\to} & (N,\phi(p)) \end{array}$$

Global:

$$\begin{array}{ccc} TM & \stackrel{T\phi}{\to} & TN \\ \downarrow & & \downarrow \\ M & \stackrel{\phi}{\to} & N \end{array}$$

If ϕ is a diffeomorphism then

$$\begin{array}{ccc} \mathfrak{X}(M) & \stackrel{\phi_*}{\to} & \mathfrak{X}(N) \\ \downarrow & & \downarrow \\ M & \stackrel{\phi}{\to} & N \end{array}$$

and also

$$\begin{array}{ccc} \mathfrak{X}(M) & \stackrel{\phi^*}{\leftarrow} & \mathfrak{X}(N) \\ \downarrow & & \downarrow \\ M & \stackrel{\phi}{\to} & N \end{array} .$$

Notice the arrow reversal. The association $\phi \mapsto \phi^*$ is a contravariant functor as opposed to $\phi \mapsto \phi_*$ which is a covariant functor (see appendix B). In fact, we have the following concerning a composition of smooth diffeomorphisms $M \stackrel{\phi}{\to} N \stackrel{f}{\to} P$:

$$\begin{aligned} (\phi \circ f)_* &= \phi_* \circ f_* \qquad \text{covariant} \\ (\phi \circ f)^* &= f^* \circ \phi^* \qquad \text{contravariant} \end{aligned}$$

If $M = N$, this gives a right and left pair of actions[1] of the diffeomorphism group $\mathrm{Diff}(M)$ on the space of vector fields: $\mathfrak{X}(M) = \Gamma(M, TM)$.

$$\begin{aligned}\mathrm{Diff}(M) \times \mathfrak{X}(M) &\to \mathfrak{X}(M) \\ (\phi_*, X) &\mapsto \phi_* X\end{aligned}$$

[1] For the definition of group action see the sections on Lie groups below.

and

$$\begin{aligned}\mathfrak{X}(M)\times \mathrm{Diff}(M) &\to \mathfrak{X}(M)\\ (X,\phi) &\mapsto \phi^* X\end{aligned}$$

Now even if $\phi : M \to N$ is not a diffeomorphism it still may be that there is a vector field $Y \in \mathfrak{X}(N)$ such that

$$T\phi \circ X = Y \circ \phi.$$

Or on other words, $T\phi \cdot X_p = Y_{\phi(p)}$ for all p in M. In this case we say that Y is **ϕ-related** to X and write $X \sim_\phi Y$.

Theorem 7.2 *The Lie derivative on functions is natural with respect to pull-back and push-forward by diffeomorphisms. In other words, if $\phi : M \to N$ is a diffeomorphism and $f \in C^\infty(M)$, $g \in C^\infty(N)$, $X \in \mathfrak{X}(M)$ and $Y \in \mathfrak{X}(N)$ then*

$$\mathcal{L}_{\phi^* Y}\phi^* g = \phi^* \mathcal{L}_Y g$$

and

$$\mathcal{L}_{\phi_* X}\phi_* f = \phi_* \mathcal{L}_X f$$

Proof.

$$\begin{aligned}(\mathcal{L}_{\phi^* Y}\phi^* g)(p) &= d(\phi^* g)(\phi^* Y)(p)\\ &= (\phi^* dg)(T\phi^{-1} Y(\phi p)) = dg(T\phi T\phi^{-1} Y(\phi p))\\ &= dg(Y(\phi p)) = (\phi^* \mathcal{L}_Y g)(p)\end{aligned}$$

The second statement follows from the first since $(\phi^{-1})^* = \phi_*$. ∎

In case the map $\phi : M \to N$ is not a diffeomorphism we still have a result when two vector fields are **ϕ-related**.

Theorem 7.3 *Let $\phi : M \to N$ be a smooth map and suppose that $X \sim_\phi Y$. Then we have for any $g \in C^\infty(N)$ $\mathcal{L}_X \phi^* g = \phi^* \mathcal{L}_Y g$.*

The proof is similar to the previous theorem and is left to the reader.

7.2 Flows and Vector Fields

All flows of vector fields near points where the field doesn't vanish look the same.

A family of diffeomorphisms $\Phi_t : M \to M$ is called a **(global) flow** if $t \mapsto \Phi_t$ is a group homomorphism from the additive group $\mathbb{R}$ to the diffeomorphism group of M and such that $\Phi_t(x) = \Phi(t, x)$ gives a smooth map $\mathbb{R}\times M \to M$. A **local flow** is defined similarly except that $\Phi(t, x)$ may not be defined on all of $\mathbb{R}\times M$ but rather on some open set in $\mathbb{R}\times M$ and so we explicitly require that

1. $\Phi_t \circ \Phi_s = \Phi_{t+s}$ and
2. $\Phi_t^{-1} = \Phi_{-t}$

for all t and s such that both sides of these equations are defined.
Using a smooth local flow we can define a vector field X^Φ by

$$X^\Phi(p) = \left.\frac{d}{dt}\right|_0 \Phi(t,p) \in T_pM.$$

If one computes the velocity vector $\dot{c}(0)$ of the curve $c : t \mapsto \Phi(t,x)$ one gets $X^\Phi(x)$. On the other hand, if we are given a smooth vector field X in open set $U \subset M$ then we say that $c : (a,b) \to M$ is an **integral curve** for X if $\dot{c}(t) = X(c(t))$ for $t \in (a,b)$.

Our study begins with a quick recounting of a basic existence and uniqueness theorem for differential equations stated here in the setting of Banach spaces. The proof may be found in Appendix H.

Theorem 7.4 *Let E be a Banach space and let $X : U \subset E \to E$ be a smooth map. Given any $x_0 \in U$ there is a smooth curve $c : (-\epsilon,\epsilon) \to U$ with $c(0) = x_0$ such that $c'(t) = X(c(t))$ for all $t \in (-\epsilon,\epsilon)$. If $c_1 : (-\epsilon_1,\epsilon_1) \to U$ is another such curve with $c_1(0) = x_0$ and $c_1'(t) = X(c(t))$ for all $t \in (-\epsilon_1,\epsilon_1)$ then $c = c_1$ on the intersection $(-\epsilon_1,\epsilon_1) \cap (-\epsilon,\epsilon)$. Furthermore, there is an open set V with $x_0 \in V \subset U$ and a smooth map $\Phi : V \times (-a,a) \to U$ such that $t \mapsto c_x(t) := \Phi(x,t)$ is a curve satisfying $c'(t) = X(c(t))$ for all $t \in (-a,a)$.*

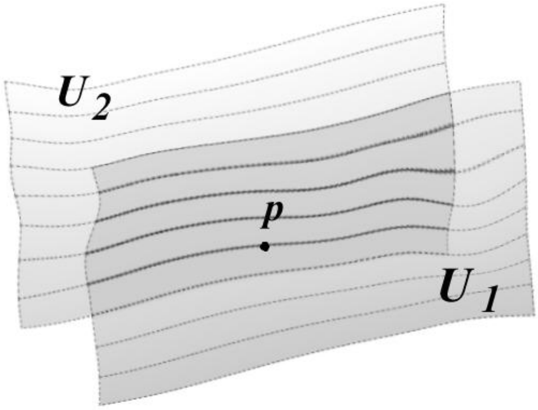
We will now use this theorem to obtain similar but more global results on smooth manifolds. First of all we can get a more global version of uniqueness:

Lemma 7.2 *If $c_1 : (-\epsilon_1,\epsilon_1) \to M$ and $c_2 : (-\epsilon_2,\epsilon_2) \to M$ are integral curves of a vector field X with $c_1(0) = c_2(0)$ then $c_1 = c_2$ on the intersection of their domains.*

Proof. Let $K = \{t \in (-\epsilon_1,\epsilon_1) \cap (-\epsilon_2,\epsilon_2) : c_1(t) = c_2(t)\}$. The set K is closed since M is Hausdorff. If follows from the local theorem 7.4 that K contains a (small) open interval $(-\epsilon,\epsilon)$. Now let t_0 be any point in K and consider the translated curves $c_1^{t_0}(t) = c_1(t_0+t)$ and $c_2^{t_0}(t) = c_2(t_0+t)$. These are also integral curves of X and agree at $t = 0$ and by 7.4 again we see that $c_1^{t_0} = c_2^{t_0}$ on some open neighborhood of 0. But this means that c_1 and c_2 agree in this neighborhood so in fact this neighborhood is contained in K implying K is also open since t_0 was an arbitrary point in K. Thus, since $I = (-\epsilon_1,\epsilon_1) \cap (-\epsilon_2,\epsilon_2)$ is connected, it must be that $I = K$ and so c_1 and c_2 agree on $I = (-\epsilon_1,\epsilon_1) \cap (-\epsilon_2,\epsilon_2)$. ∎

Flow box and Straightening

Let X be a C^r vector field on M with $r \geq 1$. A **flow box** for X at a point $p_0 \in M$ is a triple (U, a, Fl^X) where



1. U is an open set in M containing p.
2. $\mathrm{Fl}^X : U \times (-a, a) \to M$ is a C^r map and $0 < a \leq \infty$.
3. For each $p \in M$ the curve $t \mapsto c_p(t) = \mathrm{Fl}^X(p, t)$ is an integral curve of X with $c_p(0) = p$.
4. The map $\mathrm{Fl}_t^X : U \to M$ given by $\mathrm{Fl}_t^X(p) = \mathrm{Fl}^X(p, t)$ is a diffeomorphism onto its image for all $t \in (-a, a)$.

Now before we prove that flow boxes actually exist, we make the following observation: If we have a triple that satisfies 1-3 above then both $c_1 : t \mapsto \mathrm{Fl}_{t+s}^X(p)$ and $c_2 : t \mapsto \mathrm{Fl}_t^X(\mathrm{Fl}_s^X(p))$ are integral curves of X with $c_1(0) = c_2(0) = \mathrm{Fl}_s^X(p)$ so by uniqueness (Lemma 7.2) we conclude that $\mathrm{Fl}_t^X(\mathrm{Fl}_s^X(p)) = \mathrm{Fl}_{t+s}^X(p)$ as long as they are defined. This also shows that

$$\mathrm{Fl}_s^X \circ \mathrm{Fl}_t^X = \mathrm{Fl}_{t+s}^X = \mathrm{Fl}_t^X \circ \mathrm{Fl}_s^X$$

whenever defined. This is the local group property, so called because if Fl_t^X were defined for all $t \in \mathbb{R}$ (and X a global vector field) then $t \mapsto \mathrm{Fl}_t^X$ would be a group homomorphism from $\mathbb{R}$ into $\mathrm{Diff}(M)$. Whenever this happens we say that X is a **complete vector field**. The group property also implies that $\mathrm{Fl}_t^X \circ \mathrm{Fl}_{-t}^X = \mathrm{id}$ and so Fl_t^X must at least be a locally defined diffeomorphism with inverse Fl_{-t}^X.

Exercise 7.2 *Show that on $\mathbb{R}^2$ the vector fields $y^2 \frac{\partial}{\partial x}$ and $x^2 \frac{\partial}{\partial y}$ are complete but $y^2 \frac{\partial}{\partial x} + x^2 \frac{\partial}{\partial y}$ is not complete.*

Theorem 7.5 (Flow Box) *Let X be a C^r vector field on M with $r \geq 1$. Then for every point $p_0 \in M$ there exists a flow box for X at p_0. If $(U_1, a_1, \mathrm{Fl}_1^X)$ and $(U_2, a_2, \mathrm{Fl}_2^X)$ are two flow boxes for X at p_0,then $\mathrm{Fl}_1^X = \mathrm{Fl}_2^X$ on $(-a_1, a_1) \cap (-a_2, a_2) \times U_1 \cap U_2$.*

Proof. First of all notice that the U in the triple (U, a, Fl^X) does not have to be contained in a chart or even homeomorphic to an open set in the model space. However, to prove that there are flow boxes at any point we can work in the domain of a chart U_α, ψ_α and so we might as well assume that the vector field is defined on an open set in the model space as in 7.4. Of course, we may have to choose a to be smaller so that the flow stays within the range of the chart map ψ_α. Now a vector field in this setting can be taken to be a map $U \to E$ so the theorem 7.4 provides us with the flow box data (V, a, Φ) where we have taken $a > 0$ small enough that $V_t = \Phi(t, V) \subset U_\alpha$ for all $t \in (-a, a)$. Now the flow box is transferred back to the manifold via ψ_α

$$U = \psi_\alpha^{-1}(V)$$
$$\mathrm{Fl}^X(t, p) = \Phi(t, \psi_\alpha(p)).$$

Now if we have two such flow boxes $(U_1, a_1, \mathrm{Fl}_1^X)$ and $(U_2, a_2, \mathrm{Fl}_2^X)$ then by lemma 7.2 we have for any $x \in U_1 \cap U_2$ we must have $\mathrm{Fl}_1^X(t, x) = \mathrm{Fl}_2^X(t, x)$ for all $t \in (-a_1, a_1) \cap (-a_2, a_2)$.

Finally, since both $\mathrm{Fl}_t^X = \mathrm{Fl}^X(t, .)$ and $\mathrm{Fl}_{-t}^X = \mathrm{Fl}^X(-t, .)$ are both smooth and inverse of each other we see that Fl_t^X is a diffeomorphism onto its image $U_t = \psi_\alpha^{-1}(V_t)$. ∎

Definition 7.3 *Let X be a C^r vector field on M with $r \geq 1$. For any given $p \in M$ let $(T_{p,X}^-, T_{p,X}^+) \subset \mathbb{R}$ be the largest interval such that there is an integral curve $c : (T_{p,X}^-, T_{p,X}^+) \to M$ of X with $c(0) = p$. The **maximal flow** Fl^X is defined on the open set (called the maximal flow domain)*

$$\mathcal{FD}_X = \bigcup (T_{p,X}^-, T_{p,X}^+) \times \{p\}.$$

Remark 7.1 *Up until now we have used the notation Fl^X ambiguously to refer to any (local or global) flow of X and now we have used the same notation for the* unique *maximal flow defined on $\mathcal{FD}_X$. We could have introduced notation such as $\mathrm{Fl}_{\max}^X$ but prefer not to clutter up the notation to that extent unless necessary. We hope that the reader will be able to tell from context what we are referring to when we write Fl^X.*

Exercise 7.3 *Show that what we have proved so far implies that the maximal interval $(T_{p,X}^-, T_{p,X}^+)$ exists for all $p \in M$ and prove that $\mathcal{FD}_X$ is an open subset of $\mathbb{R} \times M$.*

Definition 7.4 *We say that X is a **complete vector field** if and only if $\mathcal{FD}_X = \mathbb{R} \times M$.*

Notice that we always have $\mathrm{Fl}_s^X \circ \mathrm{Fl}_t^X = \mathrm{Fl}_{t+s}^X = \mathrm{Fl}_t^X \circ \mathrm{Fl}_s^X$ whenever s, t are such that all these maps are defined but if X is a complete vector field then this equation is true for all s,$t \in \mathbb{R}$.

Definition 7.5 *The support of a vector field X is the closure of the set $\{p : X(p) \neq 0\}$ and is denoted* $\operatorname{supp}(X)$*.*

Lemma 7.3 *Every vector field that has compact support is a complete vector field. In particular if M is compact then every vector field is complete.*

Proof. Let c_p^X be the maximal integral curve through p and $(T_{p,X}^-, T_{p,X}^+)$ its domain. It is clear that for any $t \in (T_{p,X}^-, T_{p,X}^+)$ the image point $c_p^X(t)$ must always lie in the support of X. But we show that if $T_{p,X}^+ < \infty$ then given any compact set $K \subset M$, for example the support of X, there is an $\epsilon > 0$ such that for all $t \in (T_{p,X}^+ - \epsilon, T_{p,X}^+)$ the image $c_p^X(t)$ is outside K. If not then we may take a sequence t_i converging to $T_{p,X}^+$ such that $c_p^X(t_i) \in K$. But then going to a subsequence if necessary we have $x_i := c_p^X(t_i) \to x \in K$. Now there must be a flow box (U, a, x) so for large enough k, we have that t_k is within a of $T_{p,X}^+$ and $x_i = c_p^X(t_i)$ is inside U. We then a guaranteed to have an integral curve $c_{x_i}^X(t)$ of X that continues beyond $T_{p,X}^+$ and thus can be used to extend c_p^X a contradiction of the maximality of $T_{p,X}^+$. Hence we must have $T_{p,X}^+ = \infty$. A similar argument give the result that $T_{p,X}^- = -\infty$. ■

Exercise 7.4 *Let $a > 0$ be any positive real number. Show that if for a given vector field X the flow Fl^X is defined on $(-a, a) \times M$ then in fact the (maximal) flow is defined on $\mathbb{R} \times M$ and so X is a complete vector field.*

7.3 Lie Derivative

Let X be a vector field on M and let $\mathrm{Fl}^X(p, t) = \mathrm{Fl}_p^X(t) = \mathrm{Fl}_t^X(p)$ be the flow so that

$$\left.\frac{d}{dt}\right|_0 \mathrm{Fl}_p^X(t) = T_0\mathrm{Fl}_p^X \left.\frac{\partial}{\partial t}\right|_0 = X_p$$

Recall our definition of the Lie derivative of a function (4.21). The following is an alternative definition.

Definition 7.6 *For a smooth function $f : M \to \mathbb{R}$ and a smooth vector field $X \in \mathfrak{X}(M)$ define the Lie derivative $\mathcal{L}_X$ of f with respect to X by*

$$\begin{aligned} \mathcal{L}_X f(p) &= \left.\frac{d}{dt}\right|_0 f \circ \mathrm{Fl}^X(p, t) \\ &= X_p f \end{aligned}$$

Exercise 7.5 *Show that this definition is compatible with definition 4.21.*

Definition 7.7 *Let X and Y be smooth vector fields on M. Let $\varphi_t = Fl_t^X$. Then the* ***Lie derivative*** *$L_X Y$ defined by*

$$L_X Y(p) = \lim_{t \to 0} \frac{\left(T_{\varphi_t(p)}\varphi_{-t}(p)\right) \cdot Y_{\varphi_t(p)} - Y}{t} \tag{7.1}$$

for any p in the domain of X. and where we .

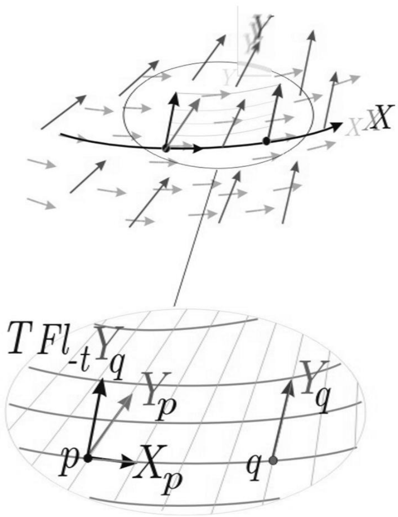
Discussion: Notice that if X is a complete vector field then for each $t \in \mathbb{R}$ the map Fl_t^X is a diffeomorphism $M \to M$ and we may define $(\mathrm{Fl}_t^{X*}Y)(p) := (T\mathrm{Fl}_t^X)^{-1}Y(\mathrm{Fl}_t(p))$ or

$$\mathrm{Fl}_t^{X*}Y = (T\mathrm{Fl}_t^X)^{-1} \circ Y \circ \mathrm{Fl}_t. \tag{7.2}$$

Here we should think of $t \mapsto \mathrm{Fl}_t^X \in \mathrm{Diff}(M)$ as a one-parameter subgroup of the "Lie group" $\mathrm{Diff}(M)$ and one may write

$$\mathcal{L}_X Y = \left.\frac{d}{dt}\right|_0 (\mathrm{Fl}_t^{X*}Y)$$

On the other hand, if X is not complete then there exist no t such that Fl_t^X is a diffeomorphism of M since for any specific t there might be points of M for which Fl_t^X is not even defined! For X which is not necessarily complete it is best to consider the map $\mathrm{Fl}^X : (t,x) \longmapsto \mathrm{Fl}^X(t,x)$ which is defined on some open neighborhood of $\{0\} \times M$ in $\mathbb{R} \times M$ which just doesn't need to contain any interval of form $\{\epsilon\} \times M$ unless $\epsilon = 0$. In fact, suppose that the domain of Fl^X contained such an interval with $\epsilon > 0$. It follow that for all $0 \le t \le \epsilon$ the map Fl_t^X is defined on all of M and $\mathrm{Fl}_t^X(p)$ exists for $0 \le t \le \epsilon$ independent of p. But now a standard argument shows that $t \mapsto \mathrm{Fl}_t^X(p)$ is defined for *all* t which means that X is a complete vector field. If X is not complete we really have no business writing $\mathcal{L}_X Y = \left.\frac{d}{dt}\right|_0 (\mathrm{Fl}_t^{X*}Y)$. Despite this it has become common to write this expression anyway especially when we are taking a derivative with respect to t. Whether or not this is just a mistake or liberal use of notation is not clear. Here is what we can say. Given any relatively compact open set $U \subset M$ the map Fl_t^X will be defined at least for all $t \in (-\varepsilon, \varepsilon)$ for some ε depending only on X and the choice of U. Because of this, the expression $\mathrm{Fl}_t^{X*}Y = (T_p\mathrm{Fl}_t^X)^{-1} \circ Y \circ \mathrm{Fl}_t$ is a well defined map on U for all $t \in (-\varepsilon, \varepsilon)$. Now if our manifold has a cover by relatively compact open sets $M = \bigcup U_i$ then we can make sense of $\mathrm{Fl}_t^{X*}Y$ on as large of a relatively compact set we like as long as t is small enough. Furthermore, if $\left.\mathrm{Fl}_t^{X*}Y\right|_{U_i}$ and $\left.\mathrm{Fl}_t^{X*}Y\right|_{U_j}$ are both defined for the same t then they both restrict to $\left.\mathrm{Fl}_t^{X*}Y\right|_{U_i \cap U_j}$. So $\mathrm{Fl}_t^{X*}Y$ makes sense point by point for small enough t. It is just that "small enough" may never be uniformly small enough over M so literally speaking Fl_t^X is just not a map on M since t has to be *some* number and for a vector field that is not complete, no t will be small enough (see the exercise 7.4 above). At any rate $t \mapsto (\mathrm{Fl}_t^{X*}Y)(p)$ has a well defined germ at $t = 0$. With this in mind the following definition makes sense even for vector fields that are not complete as long as we take a loose interpretation of $\mathrm{Fl}_t^{X*}Y$.



The following slightly different but equivalent definition is perfectly precise in any case.

Definition 7.8 (version2) *Let X and Y be smooth vector fields on M or some open subset of M. Then the **Lie derivative** $L_X Y$ defined by*

$$L_X Y(p) = \lim_{t \to 0} \frac{T\mathrm{Fl}_{-t}^X \cdot Y_{\mathrm{q}} - Y}{t} \tag{7.3}$$

for any p in the domain of X and where $q = \mathrm{Fl}_t^X(p)$.

In the sequel we will sometimes need to keep in mind our comments above in order to make sense of certain expressions.

If $X, Y \in \mathfrak{X}(M)$ where M is modelled on a C^∞ Banach space then we have defined $[X, Y]$ via the correspondence between vector fields and derivations on $C^\infty(M) = \mathfrak{F}(M)$. Even in the general case $[X, Y]$ makes sense as a derivation and if we already know that there is a vector field that acts as a derivation in the same way as $[X, Y]$ then it is unique by 4.4.

Theorem 7.6 *If $X, Y \in \mathfrak{X}(M)$ and $f \in C^\infty(M)$ then $(L_X Y)f = [X, Y]f$. That is $L_X Y$ and $[X, Y]$ are equal as derivations. Thus $L_X Y = [X, Y]$ as vector fields. In other words, $L_X Y$ is the unique vector field such that*

$$(L_X Y)(p) = X_p(Yf) - Y_p(Xf)$$

Proof. We shall show that both sides of the equation act in the same way as derivations. Consider the map $\alpha : I \times I \to \mathbb{R}$ given by $\alpha(s, t) = Y(\mathrm{Fl}^X(p, s))(f \circ \mathrm{Fl}_t^{\mathbf{X}})$. Notice that $\alpha(s, 0) = Y(\mathrm{Fl}^X(p, s))(f)$ and $\alpha(0, t) = Y(p)(f \circ \mathrm{Fl}_t^X)$ so that we have

$$\frac{\partial}{\partial s}\alpha(0,0) = \left.\frac{\partial}{\partial s}\right|_0 Y_{\mathrm{Fl}^{\mathbf{X}}(p,s)} f = \left.\frac{\partial}{\partial s}\right|_0 Yf \circ Fl^{\mathbf{X}}(p, s) = X_p Y f$$

and similarly

$$\frac{\partial}{\partial t}\alpha(0,0) = Y_p X f.$$

Subtracting we get $[X,Y](p)$. On the other hand we also have that

$$\frac{\partial}{\partial s}\alpha(0,0) - \frac{\partial}{\partial t}\alpha(0,0) = \left.\frac{d}{dr}\right|_0 \alpha(r,-r) = (L_X Y)_p$$

so we are done. To prove the last statement we just use lemma 4.4. ■

Theorem 7.7 *Let $X, Y \in \mathfrak{X}(M)$*

$$\frac{d}{dt}\mathrm{Fl}_{\mathbf{t}}^{\mathbf{X}*}Y = \mathrm{Fl}_t^{X*}(L_X Y)$$

Proof.

$$\begin{aligned}
\left.\frac{d}{dt}\right|_t \mathrm{Fl}_t^{X*}Y &= \left.\frac{d}{ds}\right|_0 \mathrm{Fl}_{t+s}^{X*}Y \\
&= \left.\frac{d}{ds}\right|_0 \mathrm{Fl}_t^{X*}(\mathrm{Fl}_s^{X*}Y) \\
&= \mathrm{Fl}_t^{X*} \left.\frac{d}{ds}\right|_0 (\mathrm{Fl}_s^{X*}Y) \\
&= \mathrm{Fl}_t^{X*} L_X Y
\end{aligned}$$

Now we can see that the infinitesimal version of the action

$$\begin{aligned}
\Gamma(M,TM) \times \mathrm{Diff}(M) &\to \Gamma(M,TM) \\
(X,\Phi) &\mapsto \Phi^* X
\end{aligned}$$

is just the Lie derivative. As for the left action of $\mathrm{Diff}(M)$ on $\mathfrak{X}(M) = \Gamma(M,TM)$ we have for $X, Y \in \mathfrak{X}(M)$

$$\begin{aligned}
\left.\frac{d}{dt}\right|_0 (\mathrm{Fl}_{t*}^X Y)(p) &= \left.\frac{d}{dt}\right|_0 T\mathrm{Fl}_t^X(Y(\mathrm{Fl}_t^{-1}(p))) \\
&= -\left.\frac{d}{dt}\right|_0 (T\mathrm{Fl}_{-t}^X)^{-1} Y(\mathrm{Fl}_{-t}(p)) \\
&= -(L_X Y) = -[X,Y]
\end{aligned}$$

It is easy to see that the Lie derivative is linear in both variables and over the real numbers. ■

Proposition 7.1 *Let $X \in \mathfrak{X}(M)$ and $Y \in \mathfrak{X}(N)$ be ϕ-related vector fields for a smooth map $\phi : M \to N$. Then*

$$\phi \circ \mathrm{Fl}_t^X = \mathrm{Fl}_t^Y \circ \phi$$

whenever both sides are defined. Suppose that $\phi : M \to M$ is a diffeomorphism and $X \in \mathfrak{X}(M)$. Then the flow of $\phi_* X = (\phi^{-1})^* X$ is $\phi \circ \mathrm{Fl}_t^X \circ \phi^{-1}$ and the flow of $\phi^* X$ is $\phi^{-1} \circ \mathrm{Fl}_t^X \circ \phi$.

Proof. Differentiating we have $\frac{d}{dt}(\phi \circ \mathrm{Fl}_t^X) = T\phi \circ \frac{d}{dt}\mathrm{Fl}_t^X = T\phi \circ X \circ \mathrm{Fl}_t^X = Y \circ \phi \circ \mathrm{Fl}_t^X$. But $\phi \circ \mathrm{Fl}_0^X(x) = \phi(x)$ and so $t \mapsto \phi \circ \mathrm{Fl}_t^X(x)$ is an integral curve of Y starting at $\phi(x)$. By uniqueness we have $\phi \circ \mathrm{Fl}_t^X(x) = \mathrm{Fl}_t^Y(\phi(x))$. ■

Lemma 7.4 *Suppose that $X \in \mathfrak{X}(M)$ and $\widetilde{X} \in \mathfrak{X}(N)$ and that $\phi : M \to N$ is a smooth map. Then $X \sim_\phi \widetilde{X}$ if and only if*

$$\widetilde{X} f \circ \phi = X(f \circ \phi)$$

for all $f \in C^\infty(U)$ and all open sets $U \subset N$.

Proof. We have $(\widetilde{X} f \circ \phi)(p) = df_{\phi(p)}\widetilde{X}(\phi(p))$. Using the chain rule we have

$$\begin{aligned} X(f \circ \phi)\,(p) &= d(f \circ \phi)(p)X_p \\ &= df_{\phi(p)} T_p\phi \cdot X_p \end{aligned}$$

and so if $X \sim_\phi \widetilde{X}$ then $T\phi \circ X = \widetilde{X} \circ \phi$ and so we get $\widetilde{X} f \circ \phi = X(f \circ \phi)$. On the other hand, if $\widetilde{X} f \circ \phi = X(f \circ \phi)$ for all $f \in C^\infty(U)$ and all open sets $U \subset N$ then we can pick U to be a chart domain and f the pull-back of a linear functional on the model space M. So assuming that we are actually on an open set $U \subset \mathsf{M}$ and $f = \alpha$ is any functional from M^* we would have $\widetilde{X}\alpha \circ \phi(p) = X(\alpha \circ \phi)$ or $d\alpha(\widetilde{X}_{\phi p}) = d(\alpha \circ \phi)X_p$ or again

$$\begin{aligned} d\alpha_{\phi p}(\widetilde{X}_{\phi p}) &= d\alpha_{\phi p}(T\phi \cdot X_p) \\ \alpha_{\phi p}(\widetilde{X}_{\phi p}) &= \alpha_{\phi p}(T\phi \cdot X_p) \end{aligned}$$

so by the Hahn-Banach theorem $\widetilde{X}_{\phi p} = T\phi \cdot X_p$. Thus since p was arbitrary $X \sim_\phi \widetilde{X}$. ■

Theorem 7.8 *Let $\phi : M \to N$ be a smooth map, $X, Y \in \mathfrak{X}(M)$, $\widetilde{X}, \widetilde{Y} \in \mathfrak{X}(N)$ and suppose that $X \sim_\phi \widetilde{X}$ and $Y \sim_\phi \widetilde{Y}$. Then*

$$[X, Y] \sim_\phi [\widetilde{X}, \widetilde{Y}]$$

In particular, if ϕ is a diffeomorphism then $[\phi_* X, \phi_* Y] = \phi_*[X, Y]$.

Proof. By lemma 7.4 we just need to show that for any open set $U \subset N$ and any $f \in C^\infty(U)$ we have $([\widetilde{X}, \widetilde{Y}]f) \circ \phi = [X, Y](f \circ \phi)$. We calculate using 7.4:

$$\begin{aligned} ([\widetilde{X}, \widetilde{Y}]f) \circ \phi &= \widetilde{X}(\widetilde{Y} f) \circ \phi - \widetilde{Y}(\widetilde{X} f) \circ \phi \\ &= X((\widetilde{Y} f) \circ \phi) - Y((\widetilde{X} f) \circ \phi) \\ &= X(Y(f \circ \phi)) - Y(X(f \circ \phi)) = [X, Y] \circ \phi. \end{aligned}$$

■

Theorem 7.9 *For $X, Y \in \mathfrak{X}(M)$ each of the following are equivalent:*

1. $\mathcal{L}_X Y = 0$
2. $(\mathrm{Fl}_t^X)^* Y = Y$
3. *The flows of X and Y commute:*

$$\mathrm{Fl}_t^X \circ \mathrm{Fl}_s^Y = \mathrm{Fl}_s^Y \circ \mathrm{Fl}_t^X \text{ whenever defined.}$$

Proof. The equivalence of 1 and 2 is follows easily from the proceeding results and is left to the reader. The equivalence of 2 and 3 can be seen by noticing that $\mathrm{Fl}_t^X \circ \mathrm{Fl}_s^Y = \mathrm{Fl}_s^Y \circ \mathrm{Fl}_t^X$ is true and defined exactly when $\mathrm{Fl}_s^Y = \mathrm{Fl}_t^X \circ \mathrm{Fl}_s^Y \circ \mathrm{Fl}_t^X$ which happens exactly when

$$\mathrm{Fl}_s^Y = \mathrm{Fl}_s^{(\mathrm{Fl}_t^X)^* Y}$$

and in turn exactly when $Y = (\mathrm{Fl}_t^X)^* Y$. ∎

Proposition 7.2 $[fX,gY] = fg[X,Y] + f(Xg)Y - g(Yf)X$.

This is a straightforward calculation.

The Lie derivative and the Lie bracket are essentially the same object and are defined for local sections $X \in \mathfrak{X}_M(U)$ as well as global sections. As is so often the case for operators in differential geometry, the Lie derivative is natural with respect to restriction so we have the commutative diagram

$$\begin{array}{ccc} \mathfrak{X}_M(U) & \stackrel{L_{X_U}}{\to} & \mathfrak{X}_M(U) \\ r_V^U \downarrow & & \downarrow r_V^U \\ \mathfrak{X}_M(V) & \stackrel{L_{X_V}}{\to} & \mathfrak{X}_M(V) \end{array}$$

where $X_U = X|_U$ denotes the restriction of $X \in \mathfrak{X}_M$ to the open set U and r_V^U is the map that restricts from U to $V \subset U$.

7.4 Time Dependent Fields

Consider a small charged particle pushed along by the field set up by a large stationary charged body. The particle will follow integral curves of the field. What if while the particle is being pushed along, the larger charged body responsible for the field is put into motion? The particle must still have a definite trajectory but now the field is *time dependent*. To see what difference this makes, consider a time dependent vector field $X(t,.)$ on a manifold M. For each $x \in M$ let $\phi_t(x)$ be the point at which a particle that was at x at time 0,ends up after time t. Will it be true that $\phi_s \circ \phi_t(x) = \phi_{s+t}(x)$? The answer is the in general this equality does not hold. The flow of a time dependent vector field is not a

1-parameter group. On the other hand, if we define $\Phi_{s,t}(x)$ to be the location of the particle that was at x at time s at the later time t then we expect

$$\Phi_{s,r} \circ \Phi_{r,t} = \Phi_{s,t}$$

which is called the Chapman-Kolmogorov law. If in a special case $\Phi_{r,t}$ depends only on $s - t$ then setting $\phi_t := \Phi_{0,t}$ we recover a flow corresponding to a time-independent vector field. The formal definitions are as follows:

Definition 7.9 *A C^r* ***time dependent vector field*** *on M is a C^r map $X : (a, b) \times M \to TM$ such that for each fixed $t \in (a, b) \subset \mathbb{R}$ the map $X_t : M \to TM$ given by $X_t(x) := X(t, x)$ is a C^r vector field.*

Definition 7.10 *Let X be a time dependent vector field. A curve $c : (a, b) \to M$ is called an* ***integral curve*** *of X if and only if*

$$\dot{c}(t) = X(t, c(t)) \text{ for all } t \in (a, b).$$

The ***evolution operator*** *$\Phi^X_{t,s}$ for X is defined by the requirement that*

$$\frac{d}{dt}\Phi^X_{t,s}(x) = X(t, \Phi^X_{t,s}(x)) \text{ and } \Phi^X_{s,s}(x) = x.$$

In other words, $t \mapsto \Phi^X_{t,s}(x)$ is the integral curve that goes through x at time s.

We have chosen to use the term "evolution operator" as opposed to "flow" in order to emphasize that the local group property does not hold in general. Instead we have the following

Theorem 7.10 *Let X be a time dependent vector field. Suppose that $X_t \in \mathfrak{X}^r(M)$ for each t and that $X : (a, b) \times M \to TM$ is continuous. Then $\Phi^X_{t,s}$ is C^r and we have $\Phi^X_{s,a} \circ \Phi^X_{a,t} = \Phi^X_{s,t}$ whenever defined.*

Theorem 7.11 *Let X and Y be smooth time dependent vector fields and let $f : \mathbb{R} \times M \to \mathbb{R}$ be smooth. We have the following formulas:*

$$\frac{d}{dt}(\Phi^X_{t,s})^* f_t = (\Phi^X_{t,s})^*(\frac{\partial f}{\partial t} + X_t f_t)$$

where $f_t() := f(t, .)$, and

$$\frac{d}{dt}(\Phi^X_{t,s})^* Y_t = (\Phi^X_{t,s})^*(\frac{\partial Y}{\partial t} + [X_t, Y_t]).$$

7.5 Problem Set

1. Find the integral curves of the vector field on $\mathbb{R}^2$ given by $X(x, y) := x^2 \frac{\partial}{\partial x} + xy\frac{\partial}{\partial y}$

2. Find the integral curves (and hence the flow) for the vector field on $\mathbb{R}^2$ given by $X(x,y) := -y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}$

3. Find an vector field on $S^2 - \{N\}$ which is not complete.

4. Using the usual spherical coordinates (ϕ,θ) on S^n calculate the bracket $[\phi\frac{\partial}{\partial\theta}, \theta\frac{\partial}{\partial\phi}]$.

5. Show that if X and Y are (time independent) vector fields that have flows Fl_t^X and Fl_t^Y then if $[X,Y]=0$, the flow of $X+Y$ is $Fl_t^X \circ Fl_t^Y$.

6. Recall that the tangent bundle of the open set $\mathrm{GL}(n,\mathbb{R})$ in $\mathbb{M}_{n\times n}(\mathbb{R})$ is identified with $\mathrm{GL}(n,\mathbb{R})\times\mathbb{M}_{n\times n}(\mathbb{R})$. Consider the vector field on $\mathrm{GL}(n,\mathbb{R})$ given by $X : g \mapsto (g, g^2)$. Find the flow of X.

7. Let $Q_t = \begin{pmatrix} \cos t & -\sin t & 0 \\ \sin t & \cos t & 0 \\ 0 & 0 & 1 \end{pmatrix}$ for $t \in (0, 2\pi]$. Let $\phi(t,P) := Q_t P$ where P is a plane in $\mathbb{R}^3$. Show that this defines a flow on $Gr(3,2)$. Find the local expression in some coordinate system of the vector field X^Q that gives this flow. Do the same thing for the flow induced by the matrices $R_t = \begin{pmatrix} \cos t & 0 & -\sin t \\ 0 & 1 & 0 \\ \sin t & 0 & \cos t \end{pmatrix}$ $t \in (0,2\pi]$, finding the vector field X^R. Find the bracket $[X^R, X^Q]$.

Chapter 8

Tensors and Differential Forms

Tensor fields (usually referred to simply as tensors) can be introduce in rough and ready way by describing their local expressions in charts and then going on to explain how such expressions are transformed under a change of coordinates. With this approach one can gain proficiency with tensor calculations in short order and this is usually the way physicists and engineers are introduced to tensors. The cost is that this approach hides much of the underlying algebraic and geometric structure. Let us start out with a provisional quick look at tensor fields from this rough and ready point of view. Provisionally, a tensor is defined obliquely by characterizing its "description" from within a coordinate chart. A tensor is described in a given coordinate chart, say $(\{x^i\}, U)$ by an indexed array of functions. By an array we refer to the obvious concept which generalizes both vectors and matrices and includes what could be pictured as 3-dimensional blocks of functions. Tensors are also distinguished according to "variance". For example, a contravariant vector field is given by an $n-$tuple of functions $(X^1, ..., X^n)$ and a covector field (1-form) is also given by an $n-$tuple of functions $(f_1, ..., f_n)$. We already know the transformation laws for these and we already know that the position of the index is significant. The first is an example of a rank 1 contravariant tensor field and the second is a rank 1 covariant tensor field. For rank 2 tensors we are assigning a matrix of functions to each coordinate system in such a way that they are related in a certain way under change of coordinates. Actually, there are several cases, (τ_{ij}), (τ^i_j), or (τ^{ij}), since the position of the index is significant here also since it is used to implicitly specify the variance of the tensor. The indices run over $\{1, ...n\}$ where n is the dimension of the manifold. The tensor (τ_{ij}) is referred to as a 2-covariant tensor while (τ^i_j) is a tensors that is 1-contravariant 1-covariant tensor. (τ^{ij}) is a 2-contravariant tensor. All three are said to be rank 2 tensors but this terminology is unfortunate since this not a reference to the rank of the underlying matrices.

Let (x^i) and $(\bar{x}^i)$ be two sets of coordinates. Let (τ_{ij}), (τ^i_j), and (τ^{ij}) refer to three generic tensors as given in the (x^i) coordinates and let $(\bar{\tau}_{ij})$, $(\bar{\tau}^i_j)$, and $(\bar{\tau}^{ij})$ refer to the same tensor as expressed in the $(\bar{x}^i)$ coordinates. The variance of a tensor is encoded in the placement of the indices and characterizes the form of the transformation law. The transformation laws for tensors of these three types are

$$\begin{aligned}
\bar{\tau}_{ij} &= \tau_{kl}\frac{\partial x^k}{\partial \bar{x}^i}\frac{\partial x^l}{\partial \bar{x}^j} \text{ (summation implied)} \\
\bar{\tau}^i_j &= \tau^k_l \frac{\partial \bar{x}^i}{\partial x^k}\frac{\partial x^l}{\partial \bar{x}^j} \\
\bar{\tau}^{ij} &= \tau^{kl}\frac{\partial \bar{x}^i}{\partial x^k}\frac{\partial \bar{x}^j}{\partial x^l}
\end{aligned}$$

Take a close look at the index positions and the location of the bars. Once one sees all the relevant notational patterns it becomes almost impossible to get these transformation laws wrong. Similarly, the index grammar helps a great deal with tensor calculations of all sorts. The exhibited transformation laws above are for rank two tensors and so can be expressed as matrix equations (not generally possible for higher rank tensors). For example, the first one becomes the matrix equation $\left(\bar{\tau}^{ij}\right) = \left(\frac{\partial x^k}{\partial \bar{x}^i}\right)(\tau_{kl})\left(\frac{\partial x^l}{\partial \bar{x}^j}\right)$. Here, we have $\left(\frac{\partial \bar{x}^i}{\partial x^k}\right) = \left(\frac{\partial x^i}{\partial \bar{x}^k}\right)^{-1}$.

A rank 2 tensor is symmetric if the underlying matrix of functions is symmetric. This definition of symmetry will be superseded once we give a more sophisticated definition of tensor.

Remark 8.1 *As with vector fields, and functions on a manifold, the components of a tensor as given in a coordinate chart* $(\mathtt{x}, U)$ *can be thought of as being functions of* $(x^1, ..., x^n) \in \mathtt{x}(U) \subset \mathbb{R}^n$ *or as functions of points* $p \in U \subset M$. *In many common situations, this distinction is minor pedantry since, after all, the role of a coordinate chart is to refer to points in* $U \subset M$ *by their images under the chart map* $\mathtt{x}$.

Example 8.1 *A thin flat slab of material has associated with it a rank 2 tensor* T *called the* ***stress tensor***. *This tensor is symmetric. If rectangular coordinates* (x, y) *are imposed and if the material is homogeneous then the tensor has constant component functions* (a_{ij}), $a_{ij} = -a_{ji}$. *In polar coordinates* $(r, \theta) = ()$ *the same tensor has components*

$$(\bar{T}_{ij}(r,\theta)) = \begin{pmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{pmatrix}\begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}\begin{pmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{pmatrix}$$

Exercise 8.1 *If a thin and flat material object (a "thin film") is both homogeneous and isotropic then in rectangular coordinates we have* $(T_{ij}) = (c\delta_{ij})$ *for some constant* c. *Show that in this case the matrix expression for the stress tensor in polar coordinates is* $\begin{pmatrix} \cos^2\theta - \left(\sin^2\theta\right) r & \cos\theta\sin\theta + (\sin\theta)\, r\cos\theta \\ -(\sin\theta)\, r\cos\theta - r^2\cos\theta\sin\theta & -\left(\sin^2\theta\right) r + r^2\cos^2\theta \end{pmatrix}$
.

Example 8.2 *If M is a surface in $\mathbb{R}^3$ then there is a tensor called the metric tensor. With respect to some coordinates (u^1, u^2) on the surface this tensor is usually denoted by (g_{ij}). If (v^1, v^2) are the components (with respect to coordinates $((u^1, u^2))$) of a vector v tangent to the surface at some point p in the domain of the given chart then the length of v is given by $g_{ij}v^i v^j$.*

Exercise 8.2 *If $(x(u^1, u^2), y(u^1, u^2), z(u^1, u^2)) = \mathbf{r}(u^1, u^2)$ is a map $V \subset \mathbb{R}^2 \to \mathbb{R}^3$ is which parameterizes a patch V of a surface M then with certain obvious assumptions about $\mathbf{r}$ under certain the inverse $\mathbf{r}^{-1} : U \to \mathbb{R}^2$ is a coordinate chart on M. Show that with these coordinates the metric tensor is given by*

$$g_{ij}(u^1, u^2) = \frac{\partial \mathbf{r}}{\partial u^i} \cdot \frac{\partial \mathbf{r}}{\partial u^j}$$

Geometrically, a tensor field consists of an assignment of a "multilinear" linear object in each tangent space. For example, $2-$covariant tensor is a smooth assignment of a bilinear form on each tangent space. Building on this idea, we now enter our more modern sophisticated explanation of tensor fields. We will need a bit more module theory. Recall the k-th tensor power of a module W is defined to be $\mathrm{W}^{\otimes k} := \mathrm{W} \otimes \cdots \otimes \mathrm{W}$. We also define the space of algebraic $({}^r{}_s)$-**tensors on** W :

$$\bigotimes{}^r{}_s(\mathrm{W}) := \mathrm{W}^{\otimes r} \otimes (\mathrm{W}^*)^{\otimes s}.$$

Similarly, $\bigotimes{}_s{}^r(\mathrm{W}) := (\mathrm{W}^*)^{\otimes s} \otimes \mathrm{W}^{\otimes r}$ is the space of $({}_s{}^r)$-**tensors on** W.

Let us be a bit more general in order to include spaces like

$$\bigotimes{}^{r_1}{}_s{}^{r_2}(\mathrm{W}) := \mathrm{W}^{\otimes r_1} \otimes (\mathrm{W}^*)^{\otimes s} \otimes \mathrm{W}^{\otimes r_2}.$$

It is extremely cumbersome to come up with a general notation to denote the spaces we have in mind but one more example should make the general pattern clear:

$$\bigotimes{}^{r_1}{}_{s_1}{}^{r_2}{}_{s_2}(\mathrm{W}) := \mathrm{W}^{\otimes r_1} \otimes (\mathrm{W}^*)^{\otimes s_1} \otimes \mathrm{W}^{\otimes r_2} \otimes (\mathrm{W}^*)^{\otimes s_2}$$

One can form a product among such spaces. For example,

$$\begin{aligned} T &= v^1 \otimes v_1 \in \bigotimes{}_1{}^1(\mathrm{W}) \\ S &= w_1 \otimes \alpha^1 \otimes \alpha^2 \in \bigotimes{}^1{}_2(\mathrm{W}) \\ T \otimes_{abs} S &= v^1 \otimes v_1 \otimes v_2 \otimes \alpha^1 \otimes \alpha^2 \in \bigotimes{}_1{}^2{}_2(\mathrm{W}) \end{aligned}$$

The altitude (up or down) of the indices may seems strange but this is just a convention we have chosen which has an aesthetic payoff later on. Notice that order of the factors is preserved. This we could call the **absolute tensor product**. For many purposes, in fact, for most purposes, it is not necessary to preserve the interlacing of the W factors and the W* factors. This leads us to define we call the consolidated tensor product. First notice that we have a canonical map cons : $\bigotimes^{r_1}{}_s{}^{r_2}(\mathrm{W}) \to \bigotimes^{r_1+r_2}{}_s(\mathrm{W})$ given on decomposable elements by simply moving all of the factors from W^* to the left while preserving all other order relations. We call this map the (left) consolidation map. For example $\mathrm{cons}(\upsilon^1 \otimes v_1 \otimes v_2 \otimes \alpha^1 \otimes \alpha^2) = \upsilon^1 \otimes \alpha^1 \otimes \alpha^2 \otimes v_1 \otimes v_2$. The consolidated tensor product is simply

$$T \otimes S = \mathrm{cons}(T \otimes_{abs} S).$$

Taking the direct sum of all possible spaces $\bigotimes^r{}_s(\mathrm{W})$ gives a space $\bigotimes^*{}_*\mathrm{W}$ which is an algebra under the consolidated tensor product (here after referred to simply as the **tensor product**). We call this the algebra of tensors of mixed type.

If $(v_1, ..., v_n)$ is a basis for a module V and $(\upsilon^1, ..., \upsilon^n)$ the dual basis for V^* then a basis for $\bigotimes^r{}_s(\mathrm{V})$ is given by

$$\{v_{i_1} \otimes \cdots \otimes v_{i_r} \otimes \upsilon^{j_1} \otimes \cdots \otimes \upsilon^{j_s}\}$$

where the index set is the set $\mathcal{I}(r, s; n)$ defined by

$$\mathcal{I}(r, s; n) := \{(i_1, ..., i_r; j_1, ..., j_s) : 1 \le i_k \le n \text{ and } 1 \le j_k \le n\}.$$

Thus $\bigotimes^r{}_s(\mathrm{V})$ has dimension $n^r n^s$ (where $n = \dim(\mathrm{V})$).

In differential geometry what is of most significance is multilinear maps rather than the abstract tensor spaces that we have just defined. The main objects are now defined.

Definition 8.1 *Let* V *and* W *be* R*-modules. A* W*-valued* $({}^r{}_s)$*-**tensor map** on* V *is a multilinear mapping of the form*

$$\Lambda : \underbrace{\mathrm{V}^* \times \mathrm{V}^* \cdots \times \mathrm{V}^*}_{r-\text{times}} \times \underbrace{\mathrm{V} \times \mathrm{V} \times \cdots \times \mathrm{V}}_{s-\text{times}} \to \mathrm{W}.$$

The set of all tensor maps into W will be denoted $T^r{}_s(\mathrm{V}; \mathrm{W})$. Similarly, a W-valued $({}_s{}^r)$-**tensor map** on V is a multilinear mapping of the form

$$\Lambda : \underbrace{\mathrm{V} \times \mathrm{V} \times \cdots \times \mathrm{V}}_{s-\text{times}} \times \underbrace{\mathrm{V}^* \times \mathrm{V}^* \cdots \times \mathrm{V}^*}_{r-\text{times}} \to \mathrm{W}$$

and the corresponding space of all such is denoted $T_s{}^r(\mathrm{V}; \mathrm{W})$.

There is, of course, a natural isomorphism $T_s{}^r(\mathrm{V}; \mathrm{W}) \cong T^r{}_s(\mathrm{V}; \mathrm{W})$ induced by the map $\mathrm{V}^s \times \mathrm{V}^{*r} \cong \mathrm{V}^{*r} \times \mathrm{V}^s$ given on homogeneous elements by $v \otimes \omega \mapsto \omega \otimes v$.

Warning: In the presence of an inner product there is another possible isomorphism here given by $v \otimes \omega \mapsto \flat v \otimes \sharp\omega$. **This map is a "transpose" map and just as we do not identify a matrix with its transpose we do not generally identify individual elements under this isomorphism. Furthermore, in the context of supergeometry another map is used that sometimes introduces a minus sign.**

Remark 8.2 *The reader may have wondered about the possibility of multilinear maps where the covariant and contravariant variables are interlaced, such as* Υ : $\mathrm{V} \times \mathrm{V}^* {\times} \mathrm{V} \times \mathrm{V}^* {\times} \mathrm{V}^* \to \mathrm{W}$. *Of course, such things exist and this example would be an element of what we might denote by* $T_1{}^1{}_1{}^2(\mathrm{V};\mathrm{W})$. *But we can agree to associate to each such object a unique element of* $T^3{}_2(\mathrm{V};\mathrm{W})$ *by simply keeping the order among the covariant variables and among the contravariant variables but shifting all covariant variables to the left of the contravariant variables. This is again a consolidation map like the consolidation map defined earlier. Some authors have insisted on the need to avoid this consolidation for reasons which we will show to be unnecessary below.*

We can now conveniently restate the universal property in this special case of tensors:

Proposition 8.1 (Universal mapping property) *Given a module (or vector space)* V *over* R, *then* $\bigotimes{}^r{}_s(\mathrm{V})$ *has associated with it, a map*

$$\otimes^r_{.s} : \mathrm{V}^r \times (\mathrm{V}^*)^s \to \bigotimes{}^r{}_s(\mathrm{V})$$

such that for any tensor map $\Lambda \in T_r{}^s(\mathrm{V};\mathsf{R})$;

$$\Lambda : \mathrm{V}^r \times (\mathrm{V}^*)^s \to \mathsf{R}$$

there is a unique linear map $\widetilde{\Lambda} : \bigotimes{}^r{}_s(\mathrm{V}) \to \mathsf{R}$ *such that* $\widetilde{\Lambda} \circ \otimes^r_{.s} = \Lambda$. *Up to isomorphism, the space* $\bigotimes{}^r{}_s(\mathrm{V})$ *is the unique space with this universal mapping property.*

Similar statements apply to $\bigotimes{}_r{}^s(\mathrm{V})$ *and* $T^r{}_s(\mathrm{V};\mathsf{R})$

Corollary 8.1 *There is an isomorphism* $(\bigotimes{}^r{}_s(\mathrm{V}))^* \cong T_r{}^s(\mathrm{V})$ *given by* $\widetilde{\Lambda} \mapsto \widetilde{\Lambda} \circ \otimes^r_s$. *(**Warning**: Notice that the* $T_r{}^s(\mathrm{V})$ *occurring here is **not** the default space* $T^r{}_s(\mathrm{V})$ *that we eventually denote by* $T^r_s(\mathrm{V})$*).*

Similarly, there is an isomorphism $(\bigotimes({}_r{}^s\mathrm{V}))^* \cong T^r{}_s(\mathrm{V})$.

Corollary 8.2 $(\bigotimes{}^r{}_s(\mathrm{V}^*))^* \cong T_r{}^s(\mathrm{V}^*)$ *and* $(\bigotimes{}_r{}^s(\mathrm{V}^*))^* \cong T^r{}_s(\mathrm{V}^*)$.

Now along the lines of the map of proposition 6.5 we have a homomorphism

$$\iota^r_{.s} : \bigotimes{}^r{}_s(\mathrm{V}) \to T^r{}_s(\mathrm{V}) \tag{8.1}$$

given by

$$\begin{aligned}&\iota^r_{.s}\big(\big(\mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_r \otimes \eta^1 \otimes \cdots .. \otimes \eta^s\big)\big)(\theta^1, ..., \theta^r, \mathrm{w}_1, .., \mathrm{w}_s)\\ &= \theta_1(\mathrm{v}_1)\theta_2(\mathrm{v}_2)\cdots\theta_l(\mathrm{v}_l)\eta^1(\mathrm{w}_1)\eta^2(\mathrm{w}_2)\cdots\eta^k(\mathrm{w}_k)\end{aligned}$$

$\theta^1, \theta^2, ..., \theta^r \in \mathrm{V}^*$ and $\mathrm{w}_1, \mathrm{w}_2, .., \mathrm{w}_k \in \mathrm{V}$. If V is a finite dimensional free module then we have $\mathrm{V} = \mathrm{V}^{**}$. This is the reflexive property.

Definition 8.2 *We say that* V *is* ***totally reflexive*** *if the homomorphism 8.1 just given is in fact an isomorphism.*

All finite dimensional free modules are totally reflexive:

Proposition 8.2 *For a finite dimensional free module* V *we have a natural isomorphism* $\bigotimes{}^r{}_s(\mathrm{V}) \cong T^r{}_s(\mathrm{V})$. *The isomorphism is given by the map* $\iota^r_{.s}$ *(see 8.1)*

Proof. Just to get the existence of a natural isomorphism we may observe that

$$\begin{aligned}\bigotimes{}^r{}_s(\mathrm{V}) &= \bigotimes{}^r{}_s(\mathrm{V}^{**}) = (\bigotimes{}^r{}_s(\mathrm{V}^*))^* \\ &= T_r{}^s(\mathrm{V}^*) = L(\mathrm{V}^{*r}, \mathrm{V}^{**s}; \mathsf{R}) \\ &= L(\mathrm{V}^{*r}, \mathrm{V}^s; \mathsf{R}) : = T^r{}_s(\mathrm{V}).\end{aligned}$$

We would like to take a more direct approach. Since V is free we may take a basis $\{f_1, ..., f_n\}$ and a dual basis $\{f^1, ..., f^n\}$ for V^*. It is easy to see that ι^r_s sends the basis elements of $\bigotimes{}^r{}_s(\mathrm{V})$ to basis elements of $T^r{}_s(\mathrm{V}; \mathsf{R})$ as for example

$$\iota^1_1 : f^i \otimes f_j \mapsto f^i \otimes f_j$$

where only the interpretation of the $\otimes$ changes. On the right side $f^i \otimes f_j$ is by definition the multilinear map $f^i \otimes f_j : (\alpha, v) := \alpha\left(f^i\right) f_j(v)$ ■

In the totally reflexive case the map $\iota^r_{.s}$ allows us to interpret elements of $\bigotimes{}^r{}_s(\mathrm{V})$ as multilinear maps by identifying $S \in \bigotimes{}^r{}_s(\mathrm{V})$ with its image under $\iota^r_{.s}$. Also the space $T^r{}_s(\mathrm{V})$ inherits a product from the tensor product on $\bigotimes{}^r{}_s(\mathrm{V})$. If we write out what this tensor product of multilinear maps must be, we arrive at a definition that makes sense even if V is not totally reflexive. It is at this point that we arrive at definitions of tensor and tensor product that are really used in most situations in differential geometry. Using these definitions one can often just forget about the abstract tensor product spaces defined above. There will be exceptions.

Definition 8.3 *A* ***tensor*** *on a module* V *is a tensor map, that is an element of* $T^r{}_s(\mathrm{V})$ *for some* r, s . *For* $S \in T^{r_1}{}_{s_1}(\mathrm{V})$ *and* $T \in T^{r_2}{}_{s_2}(\mathrm{V})$ *we define the* ***tensor product*** $S \otimes T \in T^{r_1+r_2}{}_{s_1+s_2}(\mathrm{V})$ *by*

$$\begin{aligned}&S \otimes T(\theta^1, ..., \theta^{r_1+r_2}, \mathrm{v}_1, .., \mathrm{v}_{s_1+s_2}) \\ &:= S(\theta^1, ..., \theta^{r_1}, \mathrm{v}_1, .., \mathrm{v}_{s_1}) T(\theta^1, ..., \theta^{r_2}, \mathrm{v}_1, .., \mathrm{v}_{s_2})\end{aligned}$$

Thus, if we think of an element of V as an element of V^{**} then we may freely think of a decomposable tensor $\mathrm{v}_1 \otimes ... \otimes \mathrm{v}_r \otimes \eta^1 \otimes ... \otimes \eta^s$ as a *multilinear map* by the formula

$$\begin{aligned}&(\mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_r \otimes \eta^1 \otimes \cdots .. \otimes \eta^s) \cdot (\theta^1, ..., \theta^r, \mathrm{w}_1, .., \mathrm{w}_s) \\ &= \theta_1(\mathrm{v}_1)\theta_2(\mathrm{v}_2) \cdots \theta_l(\mathrm{v}_l)\eta^1(\mathrm{w}_1)\eta^2(\mathrm{w}_2) \cdots \eta^k(\mathrm{w}_k)\end{aligned}$$

Let $T^*{}_*(\mathrm{V})$ denote the direct sum of all spaces of the form $T^{r_1+r_2}{}_{s_1+s_2}(\mathrm{V})$ where we take $T^0{}_0(\mathrm{V}) = \mathsf{R}$. The tensor product gives $T^*{}_*(\mathrm{V})$ the structure of an algebra over R as long as we make the definition that $r \otimes T := rT$ for $r \in \mathsf{R}$.

The maps $\iota^r_{\cdot s} : \bigotimes^r{}_s(\mathrm{V}) \to T^r{}_s(\mathrm{V})$ combine in the obvious way to give an algebra homomorphism $\iota : \bigotimes^*{}_*(\mathrm{V}) \to T^*{}_*(\mathrm{V})$ and we have constructed our definitions such that in the totally reflexive case ι is an isomorphism.

Exercise 8.3 *Let M be finite dimensional. Show that the $C^\infty(M)$ module $\mathfrak{X}(M)$ is totally reflexive. Hint: All free modules are totally reflexive and if $U, \mathtt{x}$ is a chart then $\mathfrak{X}(U)$ is a free module with basis $\{\frac{\partial}{\partial x^1}, ..., \frac{\partial}{\partial x^n}\}$. Now use a partition of unity argument.*

There are two approaches to tensor fields that turn out to be equivalent in the case of finite dimensional manifolds. We start with the bottom up approach where we start with multilinear algebra applied to individual tangents spaces. The second approach directly defines tensors as elements of $T^r{}_s(\mathfrak{X}(M))$.

8.1 Bottom up approach to tensors.

Let $E \to M$ be a rank r, C^∞ vector bundle and let p be a point in M. Since the fiber E_p at p is a vector space, multilinear algebra can be applied.

Definition 8.4 *Let $T^k{}_l(E) = \bigsqcup_{p\in M} T^k{}_l(E_p)$. A map $\Upsilon : M \to T^k{}_l(E)$ such that $\Upsilon(p)$ for all $p \in M$ is an element of $\Gamma(T^k{}_l(E))$ if $p \mapsto \Upsilon(p)(\alpha_1(p), ..., \alpha_k(p), X_1(p), ..., X_l(p))$ is smooth for all smooth sections $p \mapsto \alpha_i(p)$ and $p \mapsto X_i(p)$ of $E^* \to M$ and $E \to M$ respectively.*

Of course, $T^k{}_l(E)$ has the structure of a smooth vector bundle and $\Gamma(T^k{}_l(E))$ is the space of sections of this bundle.

Let $\mathrm{Aut}(\mathrm{V})$ be the set of module automorphisms of a module V. The set $\mathrm{Aut}(\mathrm{V})$ is a group. Now if G is a group and $\rho : G \to \mathrm{Aut}(\mathrm{V})$ is a group homomorphism then we get a map $\rho^{\otimes k} : G \to \mathrm{Aut}(\mathrm{V}^{\otimes k})$ defined by $\rho^{\otimes k}(g)(\mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_k) = \rho(g)\mathrm{v}_1 \otimes \cdots \otimes \rho(g)\mathrm{v}_k$. If $f \in \mathrm{Aut}(\mathrm{V})$ then the transpose $f^* : \mathrm{V}^* \to \mathrm{V}^*$, defined by $f^*(\alpha)(v) := \alpha(fv)$, is a module automorphism of V^*. Let $f_* := f^{*-1}$. Now given a group homomorphism $\rho : G \to \mathrm{Aut}(\mathrm{V})$ as before we define $\rho^{\otimes k} : G \to \mathrm{Aut}(\mathrm{V}^{*\otimes k})$ by $\rho^{\otimes k}(g)(\alpha_1 \otimes \cdots \otimes \alpha_k) = \rho_*(g)\alpha_1 \otimes \cdots \otimes \rho_*(g)\alpha_k$. We can also combine these to get $\rho^{\otimes k} : G \to \mathrm{Aut}(\bigotimes^k{}_l(\mathrm{V}))$ where

$$\begin{aligned}&\rho^{\otimes k,l}(g)(\mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_k \otimes \alpha_1 \otimes \cdots \otimes \alpha_l)\\ &:= (\rho(g)\mathrm{v}_1 \otimes \cdots \otimes \rho(g)\mathrm{v}_k \otimes \rho_*(g)\alpha_1 \otimes \cdots \otimes \rho_*(g)\alpha_l)\end{aligned}$$

If R is a field so that V is a vector space then $\rho : G \to \mathrm{Aut}(\mathrm{V})$ is a group representation by definition and we get an induced representation $\rho^{\otimes k,l}$ on $\bigotimes^k{}_l(\mathrm{V})$.

Suppose now that $\{U_\alpha, \Phi_\alpha\}$ is a cover of M by VB-charts for $E \to M$ and consider the associated cocycle $\{g_{\alpha\beta}\}$. Each $g_{\alpha\beta}$ is a map $g_{\alpha\beta} : U_a \cap U_\beta \to \mathrm{GL}(\mathbb{R}^r) = \mathrm{Aut}(\mathbb{R}^r)$. From what we have just explained we now get maps $g_{\alpha\beta}$:

$U_a \cap U_\beta \to \mathrm{GL}(\mathbb{R}^r)$ which in turn induce maps $g_{\alpha\beta}^{\otimes k,l} : U_a \cap U_\beta \to \mathrm{GL}(\bigotimes{}^k{}_l\, \mathbb{R}^r)$. These also form a cocycle and so we can build a vector bundle we denote by $\bigotimes{}^k{}_l\,(E)$.

Exercise 8.4 *Show that there is a natural bundle isomorphism* $\bigotimes{}^k{}_l\,(E) \cong T^k{}_l\,(E)$.

If $\Upsilon \in \Gamma(T^k{}_l\,(E))$ then we have a map (denoted by the same symbol) $\Upsilon : (\Gamma E^*)^k \times (\Gamma E)^l \to C^\infty(M)$ where $\Upsilon(\alpha_1, ..., \alpha_k, X_1, ..., X_l)$ is defined by

$$\Upsilon(\alpha_1, ..., \alpha_k, X_1, ..., X_l) : p \mapsto \Upsilon_p(\alpha_1(p), ..., \alpha_k(p), X_1(p), ..., X_l(p)).$$

While each Υ_p is multilinear over $\mathbb{R}$ the map Υ is multilinear over $C^\infty(M)$. Actually, $C^\infty(M)$ is not just a ring; it is an $\mathbb{R}-$algebra. The maps from $\Gamma(T_l^k(E))$ respect this structure also. Looking at the level of smooth sections in this way puts us into the ken of module theory.

Like most linear algebraic structures existing at the level of the a single fiber E_p the notion of tensor product is easily extended to the level of sections: For $\tau \in \Gamma(T^{k_1}{}_{l_1}\,(E))$ and $\eta \in \Gamma(T^{k_2}{}_{l_2}\,(E))$we *define* $\tau \otimes \eta \in \Gamma(T^{k_1+k_2}{}_{l_1+l_2}\,(E))$ by

$$\begin{aligned}&(\tau \otimes \eta)\,(p)(\alpha^1, ..., \alpha^{k_1+k_2}, v_1, ..., v_{l_1+l_2}) \\ &= \tau(\alpha^1, ..., \alpha^{k_1}, v_1, ..., v_{l_1})\eta(\alpha^{k_1+1}, ...\alpha^{k_1+k_2}, v_{l_1+1}, ..., v_{l_1+l_2})\end{aligned}$$

for all $\alpha^i \in E_p^*$ and $v_i \in E_p$. In other words, we define $(\tau \otimes \eta)\,(p)$ by using $\tau_p \otimes \eta_p$ which is already defined since E_p is a vector space. Of course, for sections $\omega^i \in \Gamma\,(E^*)$ and $X_i \in \Gamma\,(E)$ we have

$$(\tau \otimes \eta)\,(\omega^1, ..., \omega^{k_1+k_2}, X_1, ..., X_{l_1+l_2}) \in C^\infty(M)$$

Now let us assume that E is a finite rank vector bundle. Let $s_1, ..., s_n$ be a local frame field for E over an open set U and let $\sigma^1, ..., \sigma^n$ be the frame field of the dual bundle $E^* \to M$ so that $\sigma^i(s_j) = \delta^i_j$. If $\Upsilon \in \Gamma(T^k{}_l\,(E))$ then, in an obvious way, we have functions $\Upsilon^{i_1...i_k}{}_{j_1...j_l} \in C^\infty(U)$ defined by $\Upsilon^{i_1...i_k}{}_{j_1...j_l} = \tau(\sigma^{i_1}, ..., \sigma^{i_k}, s_{j_1}, ..., s_{j_l})$ and it is easy to deduce that Υ (restricted to U) has the expansion

$$\Upsilon = \Upsilon^{i_1...i_k}{}_{j_1...j_l}\, \sigma^{i_1} \otimes \cdots \otimes \sigma^{i_k} \otimes s_{j_1} \otimes \cdots \otimes s_{j_l}$$

Exercise 8.5 *Show that* $T^k{}_l\,(E) = \bigsqcup_{p \in M} T^k{}_l\,(E_p)$ *has a natural structure as a vector bundle.*

8.1.1 Tangent Bundle; Tensors

In the case of the tangent bundle TM we have special terminology and notation.

Definition 8.5 *The bundle* $T^k{}_l\,(TM)$ *is called the* $(k,l)-$***tensor bundle***.

Definition 8.6 *A* $(k,l)-$ ***tensor*** τ_p *at* p *is a real valued* $k+l$*-linear map*

$$\tau_p : \left(T_p^* M\right)^k \times \left(T_p M\right)^l \to \mathbb{R}.$$

or in other words an element of $T^k{}_l\,(E_p)$*. The space of sections* $\Gamma(T^k{}_l\,(TM))$ *is denoted by* $\mathfrak{T}^k{}_l\,(M)$ *and its elements are referred to as* $k-$*contravariant* $l-$*covariant* ***tensor fields*** *or also,* ***type*** (k,l)***-tensor fields.***

An important point is that a tangent vector may be considered as a contravariant tensor[1], or a tensor of type $(1,0)$ according to the prescription $X_p(\alpha_p) := \alpha_p(X_p)$ for each $\alpha_p \in T_p^* M$. This type of definition by duality is common in differential geometry and eventually this type of thing is done without comment. That said, we realize that if p is in a coordinate chart $(U, \mathtt{x})$ as before then the vectors $\left.\frac{\partial}{\partial x^1}\right|_p, ..., \left.\frac{\partial}{\partial x^n}\right|_p$ form a basis for $T_p M$ and using the identification $T_p M = (T_p M^*)^*$ just mentioned we may form a basis for $T^k{}_l(T_p M)$ consisting of all tensors of the form

$$\left.\frac{\partial}{\partial x^{i_1}}\right|_p \otimes \cdots \otimes \left.\frac{\partial}{\partial x^{i_k}}\right|_p \otimes \left. dx^{j_1}\right|_p \otimes \cdots \otimes \left. dx^{j_l}\right|_p .$$

For example, a $(1,1)-$tensor A_p at p could be expressed in coordinate form as

$$A_p = A^i{}_j \left.\frac{\partial}{\partial x^i}\right|_p \otimes \left. dx^j\right|_p .$$

Now the notation for coordinate expressions is already quite cluttered to the point of being intimidating and we shall take steps to alleviate this shortly but first behold the coordinate expression for a $(k,l)-$tensor (at p) :

$$\tau_p = \sum \tau^{i_1 \ldots . i_k}{}_{j_1 \ldots . j_l} \left.\frac{\partial}{\partial x^{i_1}}\right|_p \otimes \cdots \otimes \left.\frac{\partial}{\partial x^{i_k}}\right|_p \otimes \left. dx^{j_1}\right|_p \otimes \cdots \otimes \left. dx^{j_l}\right|_p$$

In differential geometry tensor fields are sometimes referred to as a simple 'tensors'. A (k,l)**-tensor field** τ over $U \subset M$ is an assignment of a (k,l)-tensor to each $p \in M$. Put in yet another way, a tensor field is a map $\tau : U \to \cup_{p \in U} T^k{}_l(T_p M)$ given by

$$\tau : p \mapsto \tau(p)$$

where each $\tau(p)$ is a (k,l)-tensor p. A tensor field is represented in a coordinate system as

$$\tau = \sum \tau^{i_1 \ldots . i_k}{}_{j_1 \ldots . j_l} \frac{\partial}{\partial x^{i_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{i_k}} \otimes dx^{j_1} \otimes \cdots \otimes dx^{j_l}$$

[1] The reader is warned that the ability to view tangent vectors as (bounded) linear functionals on $T_p^* M$ is due to the natural isomorphism $\left(T_p^* M\right)^* \cong T_p M$ which is not necessarily true for infinite dimensional spaces.

where ${\tau^{i_1....i_k}}_{j_1....j_l}$ are now functions on U. Actually, it is the restriction of τ to U that can be written in this way. In accordance with our definitions above, a tensor field τ over U is a C^∞ tensor field (over U) if whenever $\alpha_1, ..., \alpha_k$ are C^∞ 1-forms over U and $X_1, ..., X_l$ are C^∞ vector fields over U then the function

$$p \mapsto \tau(p)(\alpha_1(p), ..., \alpha_k(p), X_1(p), ..., X_l(p))$$

is smooth. It is easy to show that such a map is smooth if and only if the components ${\tau^{i_1....i_k}}_{j_1....j_l}$ are $C^\infty(U)$ for every choice of coordinates $(U, \mathtt{x})$.

Example 8.3 *We may identify* ${\mathfrak{T}^1}_0(U)$ *with* $\mathfrak{X}(U)$*.*

Example 8.4 *Let* $[\mathrm{g}_{ij}]$ *be a matrix of smooth functions on* $\mathbb{R}^3$*. Then the map*

$$\mathrm{g} : p \mapsto \sum_{1 \le i,j \le 3} \mathrm{g}_{ij}(p)\, dx^{j_1}\big|_p \otimes dx^{j_1}\big|_p$$

is a tensor field. If the matrix $[\mathrm{g}_{ij}]$ *is positive definite then* g *is a metric tensor.*

Exercise 8.6 *Show that if* X_p *and* Y_p *are vectors based at* p *(so they are in* $T_p\mathbb{R}^3$*) then*

$$\mathrm{g}(p)(X_p, Y_p) = \sum_{1 \le i,j \le 3} \mathrm{g}_{ij}(p)\xi^i\gamma^j$$

where $X_p = \sum_{1 \le i \le 3} \xi^i \frac{\partial}{\partial x^i}\big|_p$ *and* $X_p = \sum_{1 \le j \le 3} \gamma^j \frac{\partial}{\partial x^j}\big|_p$*. Here* $\mathrm{g}_{ij}(p) := \mathrm{g}(p)(\frac{\partial}{\partial x^i}\big|_p, \frac{\partial}{\partial x^j}\big|_p)$*.*

Definition 8.7 *If* $f : M \to N$ *is a smooth map and* τ *is a* $k-$*covariant tensor on* N *then we define the pull-back* $f^*\tau \in {\mathfrak{T}^0}_k(M)$ *by*

$$f^*\tau(v_1, ..., v_k)(p) = \tau(Tf \cdot v_1, ..., Tf \cdot v_k)$$

for all $v_1, ..., v_1 \in T_pM$ *and any* $p \in M$*.*

It is not hard to see that $f^* : {\mathfrak{T}^0}_k(N) \to {\mathfrak{T}^0}_k(M)$ is linear over $\mathbb{R}$ and for any $h \in C^\infty(N)$ and $\tau \in {\mathfrak{T}^0}_k(N)$ we have $f^*(h\tau) = (h \circ f)\, f^*\tau$.

Exercise 8.7 *Let* f *be as above. Show that for* $\tau_1 \in {\mathfrak{T}^0}_{k_1}(M)$ *and* $\tau_2 \in {\mathfrak{T}^0}_{k_2}(M)$ *we have* $f^*(\tau_1 \otimes \tau_2) = f^*\tau_1 \otimes f^*\tau_2$*.*

8.2 Top down approach to tensor fields

There is another way to think about tensors that is sometimes convenient (and makes the notation a bit simpler). The point is this; Suppose that X and Y are actually vector fields defined on U. We have just given a definition of tensor field that we will use to discover an alternate definition. Take a 2-covariant-tensor τ field over U for concreteness, then we can think of τ as a multilinear creature whose sole ambition in life is to "eat" a pair of vector fields like X and Y and spit out a smooth function. Thus

$$X, Y \mapsto \tau(X, Y)$$

where $\tau(X,Y)(p) := \tau_p(X(p), Y(p))$. It is easy to see how to generalize this to view any tensor field $S \in \mathfrak{T}^k{}_l(M)$ as an object that takes as inputs, k elements $\mathfrak{X}(U)$ and l elements of $\mathfrak{X}^*(U)$. This leads to an alternate definition of a tensor field over $U \subset M$. In this "top down" view we *define* a $k-$contravariant $l-$covariant tensor field to be a $C^\infty(M)$ multilinear map

$$\mathfrak{X}(U)^k \times \mathfrak{X}^*(U)^l \to C^\infty(U);$$

that is an element of $T^k{}_l(\mathfrak{X}(M))$ For example, a global covariant 2- tensor field on a manifold M is a map $\tau : \mathfrak{X}(M) \times \mathfrak{X}(M) \to C^\infty(M)$ such that

$$\begin{aligned}
\tau(f_1X_1 + f_2X_2, Y) &= f_1\tau(X_1, Y) + f_2\tau(X_2, Y) \\
\tau(Y, f_1X_1 + f_2X_2) &= f_1\tau(Y, X_1) + f_2\tau(Y, X_2)
\end{aligned}$$

for all $f_1, f_2 \in C^\infty(M)$ and all $X_1, X_2, Y \in \mathfrak{X}(M)$.

8.3 Matching the two approaches to tensor fields.

If we define a tensors field as we first did, that is, as a field of point tensors, then we immediately obtain a tensor as defined in the top down approach. As an example consider τ as a smooth field of covariant 2-tensors. We wish to define $\tau(X,Y)$ for any $X, Y \in \mathfrak{X}(M)$. We want this to be a function and to be bilinear over functions. In fact, it is rather obvious that we should just take $\tau(X,Y)(p)$ to be defined by $\tau(p)(X(p), Y(p))$ or $\tau_p(X_p, Y_p)$. This all works just fine. On the other hand if τ is initially defined as a bilinear map $\tau : \mathfrak{X}(M)\times\mathfrak{X}(M) \to C^\infty(M)$ then how should we obtain point tensors for each $p \in M$? This is exactly where things might not go so well if the manifold is not finite dimensional. What we need is the existence of smooth cut-off functions. Some Banach manifolds support cut-off functions but not all do. Let us assume that M supports cut-off functions. This is always the case if M is finite dimensional.

Proposition 8.3 *Let $p \in M$ and $\tau \in T^k{}_l(\mathfrak{X}(M))$. Let $\theta_1, ..., \theta_k$ and $\bar{\theta}_1, ..., \bar{\theta}_k$ be smooth 1-forms such that $\theta_i(p) = \bar{\theta}_i(p)$ for $1 \le i \le k$; also let $X_1, ..., X_k$ and $\bar{X}_1, ..., \bar{X}_k$ be smooth vector fields such that $X_i(p) = \bar{X}_i(p)$ for $1 \le i \le l$. Then we have that*

$$\tau(\theta_1, ..., \theta_k, X_1, ..., X_k)(p) = \tau(\bar{\theta}_1, ..., \bar{\theta}_k, \bar{X}_1, ..., \bar{X}_k)(p)$$

Proof. The proof will follow easily if we can show that $\tau(\theta_1, ..., \theta_k, X_1, ..., X_k)(p) = 0$ whenever one of $\theta_1(p), ..., \theta_k(p), X_1(p), ..., X_k(p)$ is zero. We shall assume for simplicity of notation that $k = 1$ and $l = 2$. Now suppose that $X_1(p) = 0$. If $U, x^1, ..., x^n$ then $X_1|_U = \sum \xi^i \frac{\partial}{\partial x^i}$ for some smooth functions $\xi^i \in C^\infty(U)$. Let

β be a cut-off function with support in U. Then $\beta\ X_1|_U$ and $\beta^2\ X_1|_U$ extend by zero to elements of $\mathfrak{X}(M)$ which we shall denote by βX_1 and $\beta^2 X_1$. Similarly, $\beta\frac{\partial}{\partial x^i}$ and $\beta^2\frac{\partial}{\partial x^i}$ are globally defined vector fields, $\beta\xi^i$ is a global function and $\beta X_1 = \sum \beta\xi^i\beta\frac{\partial}{\partial x^i}$. Thus

$$\begin{aligned}\beta^2\tau(\theta_1, X_1, X_2) &= \tau(\theta_1, \beta^2 X_1, X_2)\\ &= \tau(\theta_1, \sum \beta\xi^i\beta\frac{\partial}{\partial x^i}, X_2)\\ &= \sum \beta\tau(\theta_1, \xi^i\beta\frac{\partial}{\partial x^i}, X_2).\end{aligned}$$

Now since $X_1(p) = 0$ we must have $\xi^i(p) = 0$. Also recall that $\beta(p) = 1$. Plugging p into the formula above we obtain $\tau(\theta_1, X_1, X_2)(p) = 0$.

A similar argument holds when $X_1(p) = 0$ or $\theta_1(p) = 0$.

Assume that $\bar{\theta}_1(p) = \theta_1(p), \bar{X}_1(p) = \bar{X}_1(p)$ and $X_2(p) = \bar{X}_2(p)$. Then we have

$$\begin{aligned}&\tau(\bar{\theta}_1, \bar{X}_1, \bar{X}_2) - \tau(\theta_1, X_1, X_2)\\ &= \tau(\bar{\theta}_1 - \theta_1, \bar{X}_1, \bar{X}_2) + \tau(\theta_1, \bar{X}_1, \bar{X}_2) + \tau(\theta_1, \bar{X}_1 - X_1, \bar{X}_2)\\ &+ \tau(\bar{\theta}_1, X_1, \bar{X}_2) + \tau(\bar{\theta}_1, X_1, \bar{X}_2 - X_2)\end{aligned}$$

Since $\bar{\theta}_1 - \theta_1$, $\bar{X}_1 - X_1$ and $\bar{X}_2 - X_1$ are all zero at p we obtain the result that $\tau(\bar{\theta}_1, \bar{X}_1, \bar{X}_2)(p) = \tau(\theta_1, X_1, X_2)(p)$. ■

Let $r^U_V : \mathfrak{T}^k_l(U) \to \mathfrak{T}^k_l(V)$ denote the obvious restriction map. The assignment $U \to \mathcal{T}^k_l(U)$ is an example of a sheaf. This means that we have the following easy to verify facts.

1. $r^V_W \circ r^U_V = r^U_W$ whenever $W \subset V \subset U$.
2. $r^V_V = \mathrm{id}_V$ for all open $V \subset M$.
3. Let $U = \bigcup_{U_\alpha \in \mathcal{U}} U_\alpha$ for some collection of open sets $\mathcal{U}$. If $s_1, s_2 \in \mathfrak{T}^k_l(U)$ and $r^U_{U_\alpha} s_1 = r^U_{U_\alpha} s_2$ for all $U_\alpha \in \mathcal{U}$ then $s_1 = s_2$.
4. Let $U = \bigcup_{U_\alpha \in \mathcal{U}} U_\alpha$ for some collection of open sets $\mathcal{U}$. If $s_\alpha \in \mathfrak{T}^k_l(U_\alpha)$ and whenever $U_\alpha \cap U_\beta \neq \emptyset$ we have

$$r^U_{U_\alpha \cap U_\beta} s_\alpha = r^U_{U_\alpha \cap U_\beta} s_\beta$$

then there exists a $s \in \mathfrak{T}^k_l(U)$ such that $r^U_{U_\alpha} s = s_\alpha$.

We shall define several operations on spaces of tensor fields. We would like each of these to be natural with respect to restriction. We already have one such operation; the tensor product. If $\Upsilon_1 \in \mathfrak{T}^{k_1}_{l_1}(U)$ and $\Upsilon_2 \in \mathfrak{T}^{k_2}_{l_2}(U)$ and $V \subset U$ then $r^U_V(\Upsilon_1 \otimes \Upsilon_2) = r^U_V\Upsilon_1 \otimes r^U_V\Upsilon_2$. Because of the naturality of all the operations we introduce, it is generally safe to use the same letter to denote a tensor field and its restriction to an open set such as a coordinate neighborhood.

A $\binom{k}{\ l}$−tensor field τ may generally be expressed in a coordinate neighborhood as

$$\tau = \sum \tau^{i_1,\ldots,i_k}{}_{j_1,\ldots,j_l} \frac{\partial}{\partial x^{i_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{i_k}} \otimes dx^{j_1} \otimes \cdots \otimes dx^{j_l}$$

where the $\tau^{i_1,\ldots,i_k}{}_{j_1,\ldots,j_l}$ are now functions defined on the coordinate neighborhood and are called the (numerical) components of the tensor field. If the components of a tensor field are C^r functions for any choice of coordinates defined on the domain of the tensor field then we say that we have a C^r tensor field. A "smooth" tensor field usually means a C^∞−tensor field.

Definition 8.8 *A differential* 1−*form is a* $\binom{0}{\ 1}$−*tensor field. Every* 1-*form* θ *can be written in a coordinate chart as* $\theta = \sum_{i=1}^{n} \theta_i dx^i$ *for some unique functions* θ_i *defined on the coordinate domain.*

8.4 Index Notation

In a given coordinate chart a $\binom{k}{\ l}$−tensor field is clearly determined by its component functions. The same is true of vector fields and differential 1-forms. If a 1-form has components θ_i in some chart and a vector field has components ξ^j then a $\binom{1}{\ 1}$−tensor field with components $\tau^i{}_j$ acts on the form and field by

$$(\theta_i), (\xi^j) \mapsto \sum \tau^i{}_j \, \xi^j \theta_i.$$

Many geometers and physicists suggest that we keep the indices of the components for vector fields as superscripts and those of a 1-form as subscripts. Further, the indices of tensors should be distinguished in a similar manner. We have been following this advice so far and shall continue. This happens naturally, since

$$\tau^{i_1,\ldots,i_k}{}_{j_1,\ldots,j_l} := \tau(dx^{i_1}, ..., dx^{i_k}, \frac{\partial}{\partial x^{j_1}}, ..., \frac{\partial}{\partial x^{j_k}}).$$

Now an interesting pattern appears that helps in bookkeeping. We only seem to be summing over indices that are repeated-one up and one down. Thus Einstein suggested that we can drop the $\sum$ from the notation and look for the repeated indices to inform us of an implied summation. Thus we may write $\tau^i{}_j \, \xi^j \theta_i$ instead of $\sum_{i,j} \tau^i{}_j \, \xi^j \theta_i$. Some people have jokingly suggested that this was Einstein's greatest contribution to mathematics.

Now there seems to be a problem with this index notation in that it is only valid in the domain of a coordinate chart and also depends on the choice of coordinates. This seems to defy the spirit of modern differential geometry. But this need not be the case. Suppose we interpret τ^a_{bc} so that the "indices" a, b and c do not to refer to actual numerical indices but rather as markers for the slots or inputs of a an abstract tensor. The up and down distinguish between slots for forms and slots for vector fields. Thus we take $\tau^a_{bc}\theta_a X^b Y^c$ as just an alternative notation for $\tau(\theta, X, Y)$. This approach is called the abstract index approach.

Returning to the concrete index point of view, consider tensor field τ and suppose we have two overlapping charts $U, \mathtt{x}$ and $V, \bar{\mathtt{x}}$. How are the components of τ in these two charts related in the overlap $U \cap V$? It is a straight forward exercise to show that for a $({}^1{}_2)-$tensor field τ the relation is

$$\bar{\tau}^i_{jk} = \tau^l_{rs} \frac{\partial x^r}{\partial \bar{x}^j} \frac{\partial x^s}{\partial \bar{x}^k} \frac{\partial \bar{x}^i}{\partial x^l}$$

which has the obvious generalization to more general $({}^k{}_l)-$tensor fields. It is important to realize that $\frac{\partial x^j}{\partial \bar{x}^m}$ means $\frac{\partial x^j}{\partial \bar{x}^m}(\bar{x}^1, ..., \bar{x}^n)$ while $\frac{\partial \bar{x}^i}{\partial x^l}$ means $\frac{\partial \bar{x}^i}{\partial x^l} \circ \mathtt{x} \circ \bar{\mathtt{x}}^{-1}(\bar{x}^1, ..., \bar{x}^n)$. Alternatively, we may think of all of these functions as living on the manifold. In this interpretation we read the above as

$$\bar{\tau}^i_{jk}(p) = \tau^l_{rs}(p) \frac{\partial x^r}{\partial \bar{x}^j}(p) \frac{\partial x^s}{\partial \bar{x}^k}(p) \frac{\partial \bar{x}^i}{\partial x^l}(p) \text{ for each } p \in U \cap V$$

which is fine for theoretical and general calculations but rather useless for concrete coordinate expressions such as polar coordinates. For example, suppose that a tensor τ has components with respect to rectangular coordinates on $\mathbb{R}^2$ given by τ^i_{jk} where for indexing purposes we take $(x, y) = (u^1, u^2)$ and $(r, \theta) = (v^1, v^2)$. Then

$$\bar{\tau}^i_{jk} = \tau^l_{rs} \frac{\partial u^r}{\partial v^j} \frac{\partial u^s}{\partial v^k} \frac{\partial v^i}{\partial u^l}$$

should be read so that $\frac{\partial u^j}{\partial v^j} = \frac{\partial u^j}{\partial v^j}(v^1, v^2)$ while $\frac{\partial v^i}{\partial u^i} = \frac{\partial v^i}{\partial u^i}(u^1(v^1, v^2), u^2(v^1, v^2))$.

Exercise 8.8 *If in rectangular coordinates on $\mathbb{R}^2$ a tensor field with components (arranged as a matrix) given by*

$$\tau^i{}_j = \begin{bmatrix} \tau^1{}_1 = 1 & \tau^2{}_1 = xy \\ \tau^1{}_2 = xy & \tau^2{}_2 = 1 \end{bmatrix}$$

then what are the components of the same tensor in polar coordinates?

Exercise 8.9 *Notice that a $({}^1{}_1)-$ tensor field $\tau^a{}_b$ may be thought of as a linear map (for each p) $\tau(p) : T_pM \mapsto T_pM$ given by $v^b \mapsto \tau^a{}_b v^b$ (summation convention here!).*

8.5 Tensor Derivations

We would like to be able to define derivations of tensor fields . In particular we would like to extend the Lie derivative to tensor fields. For this purpose we introduce the following definition which will be useful not only for extending the Lie derivative but can also be used in several other contexts. Recall the presheaf of tensor fields $U \mapsto \mathfrak{T}^r_s(U)$ on a manifold M.

Definition 8.9 *A differential tensor derivation is a collection of maps $\mathcal{D}^r_s|_U : \mathfrak{T}^r_s(U) \to \mathfrak{T}^r_s(U)$, all denoted by $\mathcal{D}$ for convenience, such that*

1. *$\mathcal{D}$ is a presheaf map for $\mathfrak{T}^r_s$ considered as a presheaf of vector spaces over $\mathbb{R}$. In particular, for all open U and V with $V \subset U$ we have*
$$(\mathcal{D}\Upsilon|_U)_V = \mathcal{D}\,\Upsilon|_V$$
for all $\Upsilon \in \mathfrak{T}^r_s(U)$. I.e., the restriction of $\mathcal{D}\Upsilon|_U$ to V is just $\mathcal{D}\,\Upsilon|_V$.
2. *$\mathcal{D}$ commutes with contractions against simple tensors.*
3. *$\mathcal{D}$ satisfies a derivation law. Specifically, for $\Upsilon_1 \in \mathfrak{T}^r_s(U)$ and $\Upsilon_2 \in \mathfrak{T}^j_k(U)$ we have*
$$\mathcal{D}(\Upsilon_1 \otimes \Upsilon_2) = \mathcal{D}\Upsilon_1 \otimes \Upsilon_2 + \Upsilon_1 \otimes \mathcal{D}\Upsilon_2.$$

For finite dimensional manifolds, the conditions 2 and 3 imply that for $\Upsilon \in \mathfrak{T}^r_s(U)$, $\alpha_1, ..., \alpha_r \in \mathfrak{X}^*(U)$ and $X_1, ..., X_s \in \mathfrak{X}(U)$ we have

$$\begin{aligned}\mathcal{D}(\Upsilon(\alpha_1, ..., \alpha_r, X_1, ..., X_s)) &= \mathcal{D}\Upsilon(\alpha_1, ..., \alpha_r, X_1, ..., X_s) \\ &+ \sum_i \Upsilon(\alpha_1, ..., \mathcal{D}\alpha_i, ..., \alpha_r, X_1, ..., X_s) \\ &+ \sum_i \Upsilon(\alpha_1, ..., \alpha_r, X_1, ..., ..., \mathcal{D}X_i, ..., X_s).\end{aligned} \tag{8.2}$$

This follows by noticing that

$$\Upsilon(\alpha_1, ..., \alpha_r, X_1, ..., X_s) = C(\Upsilon \otimes (\alpha_1 \otimes \cdots \otimes \alpha_r \otimes X_1 \otimes \cdots \otimes X_s))$$

and applying 2 and 3. For the infinite dimensional case we could take equation 8.2 as a criterion to replace 2 and 3.

Proposition 8.4 *Let M be a finite dimensional manifold and suppose we have a map on global tensors $\mathcal{D} : \mathfrak{T}^r_s(M) \to \mathfrak{T}^r_s(M)$ for all r, s nonnegative integers such that 2 and 3 above hold for $U = M$. Then there is a unique induced tensor derivation that agrees with $\mathcal{D}$ on global sections.*

Proof. We need to define $\mathcal{D} : \mathfrak{T}^r_s(U) \to \mathfrak{T}^r_s(U)$ for arbitrary open U as a derivation. Let δ be a function that vanishes on a neighborhood of V of $p \in U$.

Claim 8.1 *We claim that $(\mathcal{D}\delta)(p) = 0$.*

■

Proof. To see this let β be a cut-off function equal to 1 on a neighborhood of p and zero outside of V. Then $\delta = (1 - \beta)\delta$ and so

$$\begin{aligned}\mathcal{D}\delta(p) &= \mathcal{D}((1-\beta)\delta)(p) \\ &= \delta(p)\mathcal{D}(1-\beta)(p) + (1-\beta(p))\mathcal{D}\delta(p) = 0\end{aligned}$$

Given $\tau \in \mathfrak{T}^r_s(U)$ let β be a cut-off function with support in U and equal to 1 on neighborhood of $p \in U$. Then $\beta\tau \in \mathfrak{T}^r_s(M)$ after extending by zero. Now define

$$(\mathcal{D}\tau)(p) = \mathcal{D}(\beta\tau)(p).$$

Now to show this is well defined let β_2 be any other cut-off function with support in U and equal to 1 on neighborhood of $p_0 \in U$. Then we have

$$\begin{aligned}
&\mathcal{D}(\beta\tau)(p_0) - \mathcal{D}(\beta_2\tau)(p_0) \\
&= \mathcal{D}(\beta\tau) - \mathcal{D}(\beta_2\tau))(p_0) = \mathcal{D}((\beta - \beta_2)\tau)(p_0) = 0
\end{aligned}$$

where that last equality follows from our claim above with $\delta = \beta - \beta_2$. Thus $\mathcal{D}$ is well defined on $\mathfrak{T}^r_s(U)$. We now show that $\mathcal{D}\tau$ so defined is an element of $\mathfrak{T}^r_s(U)$. Let ψ_α, U_α be a chart containing p. Let $\psi_\alpha : U_\alpha \to V_\alpha \subset \mathbb{R}^n$. Then we can write $\tau|_{U_\alpha} \in \mathfrak{T}^r_s(U_\alpha)$ as $\tau_{U_\alpha} = \tau^{j_1,\dots,j_r}_{i_1,\dots,i_s} dx^{i_1} \otimes \cdots \otimes dx^{i_s} \otimes \frac{\partial}{\partial x^{j_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{j_r}}$. We can use this to show that $\mathcal{D}\tau$ as defined is equal to a global section in a neighborhood of p and so must be a smooth section itself since the choice of $p \in U$ was arbitrary. To save on notation let us take the case $r = 1$, $s = 1$. Then $\tau_{U_\alpha} = \tau^i_j dx^j \otimes \frac{\partial}{\partial x^i}$. Let β be a cut-off function equal to one in a neighborhood of p and zero outside of $U_\alpha \cap U$. Now extend each of the sections $\beta\tau^i_j \in \mathfrak{F}(U)$, $\beta dx^j \in \mathfrak{T}^0_1(U)$ and $\beta\frac{\partial}{\partial x^i} \in \mathfrak{T}^1_0(U)$ to global sections and apply $\mathcal{D}$ to $\beta^3\tau\big|_{U_\alpha} = \beta\tau^i_j \beta dx^j \otimes \beta\frac{\partial}{\partial x^i}$ to get

$$\begin{aligned}
&= \mathcal{D}(\beta^3\tau) = \mathcal{D}(\beta\tau^i_j\beta dx^j \otimes \beta\frac{\partial}{\partial x^i}) \\
= \mathcal{D}(\beta\tau^i_j)\beta dx^j \otimes \beta\frac{\partial}{\partial x^i} &+ \beta\tau^i_j\mathcal{D}(\beta dx^j) \otimes \beta\frac{\partial}{\partial x^i} \\
&+ \beta\tau^i_j\beta dx^j \otimes \mathcal{D}(\beta\frac{\partial}{\partial x^i})
\end{aligned}$$

Now by assumption $\mathcal{D}$ takes smooth global sections to smooth global sections so both sides of the above equation are smooth. On the other hand, independent of the choice of β we have $\mathcal{D}(\beta^3\tau)(p) = \mathcal{D}(\tau)(p)$ by definition and valid for all p in a neighborhood of p_0. Thus $\mathcal{D}(\tau)$ is smooth and is the restriction of a smooth global section. We leave the proof of the almost obvious fact that this gives a unique derivation $\mathcal{D} : \mathfrak{T}^r_s(U) \to \mathfrak{T}^r_s(U)$ to the reader. ∎

Now one big point that follows from the above considerations is that the action of a tensor derivation on functions, 1-forms and vector fields determines the derivation on the whole tensor algebra. We record this as a theorem.

Theorem 8.1 *Let $\mathcal{D}_1$ and $\mathcal{D}_2$ be two tensor derivations (so satisfying 1,2, and 3 above) that agree on functions, 1-forms and vector fields. Then $\mathcal{D}_1 = \mathcal{D}_2$. Furthermore, if $\mathcal{D}_U$ can be defined on $\mathfrak{F}(U)$ and $\mathfrak{X}(U)$ for each open $U \subset M$ so that*

1. *$\mathcal{D}_U(f \otimes g) = \mathcal{D}_U f \otimes g + f \otimes \mathcal{D}_U g$ for all $f, g \in \mathfrak{F}(U)$,*

2. *for each* $f \in \mathfrak{F}(M)$ *we have*$(\mathcal{D}_M f)|_U = \mathcal{D}_U \, f|_U$,

3. $\mathcal{D}_U(f \otimes X) = \mathcal{D}_U f \otimes X + f \otimes \mathcal{D}_U X$ *for all* $f \in \mathfrak{F}(U)$ *and* $X \in \mathfrak{X}(U)$,

4. *for each* $X \in \mathfrak{X}(M)$ *we have*$(\mathcal{D}_M X)|_U = \mathcal{D}_U \, X|_U$,

then there is a unique tensor derivation D *on the presheaf of all tensor fields that is equal to* $\mathcal{D}_U$ *on* $\mathfrak{F}(U)$ *and* $\mathfrak{X}(U)$ *for all* U.

Sketch of Proof. Define $\mathcal{D}$ on $\mathfrak{X}^*(U)$ by requiring $\mathcal{D}_U(\alpha \otimes X) = \mathcal{D}_U \alpha \otimes X + \alpha \otimes \mathcal{D}_U X$

so that after contraction we see that we must have $(\mathcal{D}_U \alpha)(X) = \mathcal{D}_U(\alpha(X)) - \alpha(\mathcal{D}_U X)$. Now using that $\mathcal{D}$ must satisfy the properties 1,2,3 and 4 and that we wish $\mathcal{D}$ to behave as a derivation we can easily see how $\mathcal{D}$ must act on any simple tensor field and then by linearity on any tensor field. But this prescription can serve as a definition. ∎

Corollary 8.3 *The Lie derivative* $\mathcal{L}_X$ *can be extended to a tensor derivation for any* $X \in \mathfrak{X}(M)$.

We now present a different way of extending the Lie derivative to tensors that is equivalent to what we have just done. First let $\Upsilon \in \mathfrak{T}^r_s(U)$ If $\phi : U \to \phi(U)$ is a diffeomorphism then we can define $\phi^* \Upsilon \in \mathfrak{T}^r_s(U)$ by

$$
\begin{aligned}
&(\phi^*\Upsilon)(p)(\alpha_1, ..., \alpha_r, v^1, ..., v^s) \\
&= \Upsilon(\phi(p))(T^*\phi^{-1} \cdot \alpha_1, ..., T^*\phi^{-1} \cdot \alpha_r, T\phi \cdot v^1, ..., T\phi \cdot v^s).
\end{aligned}
$$

Now if X is a complete vector field on M we can define

$$
\mathcal{L}_X \Upsilon = \left. \frac{d}{dt} \right|_0 (\mathrm{Fl}_t^{X*} \Upsilon)
$$

just as we did for vector fields. Also, just as before this definition will make sense point by point even if X is not complete.

The Lie derivative on tensor fields is natural in the sense that for any diffeomorphism $\phi : M \to N$ and any vector field X we have

$$
\mathcal{L}_{\phi_* X} \phi_* \tau = \phi_* \mathcal{L}_X \Upsilon.
$$

Exercise 8.10 *Show that the Lie derivative is natural by using the fact that it is natural on functions and vector fields.*

8.6 Differential forms

Just as for tensors, differential forms may be introduced in a way that avoids sophistication since in one guise, a differential form is nothing but an antisymmetric tensor field. We can give a rough and ready description of differential forms in $\mathbb{R}^n$. sufficient to a reformulation of vector calculus in terms of differential forms. We will do this now as a warm up.

8.6.1 Differential forms on $\mathbb{R}^n$

First of all, on $\mathbb{R}^n$ will have 0-forms, $1-$forms and so on until we get to $n-$forms. There are no k-forms for $k > n$. Let $(x^1, ..., x^n)$ be standard coordinates on $\mathbb{R}^n$. We already know what $1-$forms are and we know that every one form on $\mathbb{R}^n$ can be written $\alpha = \alpha^i dx^i$. By definition a $0-$form is a smooth function. We now include expressions $dx^i \wedge dx^j$ for $1 \leq i, j \leq n$. These are our basic $2-$forms. A generic $2-$form on $\mathbb{R}^n$ is an expression of the form $\omega = w_{ij} dx^i \wedge dx^j$ (summation) where w_{ij} are smooth functions. Our first rule is that $dx^i \wedge dx^j = -dx^j \wedge dx^i$ which implies that $dx^i \wedge dx^i = 0$. This means that in we may as well assume that $w_{ij} = -w_{ji}$ and also the summation may be taken just over ij such that $i < j$. The $\wedge$ will become a type of product bilinear over functions called the exterior product or the wedge product that we explain only by example at this point.

Example 8.5 $\mathbb{R}^3$

$$
\begin{aligned}
&(xydx + zdy + dz) \wedge (xdy + zdz) \\
&= xydx \wedge xdy + xydx \wedge zdz + zdy \wedge xdy \\
&+ zdy \wedge zdz + dz \wedge xdy + dz \wedge zdz \\
&= x^2ydx \wedge dy + xyzdx \wedge dz + z^2dy \wedge dz \\
&+ xdz \wedge dy \\
&= x^2ydx \wedge dy + xyzdx \wedge dz + (z^2 - x)dy \wedge dz
\end{aligned}
$$

There exist k forms for $k \leq n$.

Example 8.6

$$
\begin{aligned}
&(xyz^2dx \wedge dy + dy \wedge dz) \wedge (dx + xdy + zdz) \\
&= xyz^2dx \wedge dy \wedge dx + dy \wedge dz \wedge dx + xyz^2dx \wedge dy \wedge xdy \\
&+ xyz^2dx \wedge dy \wedge zdz \\
&+ dy \wedge dz \wedge dz + dy \wedge dz \wedge zdz \\
&= dy \wedge dz \wedge dx + xyz^3dx \wedge dy \wedge dz = (xyz^3 + 1)dx \wedge dy \wedge dz
\end{aligned}
$$

where we have used $dy \wedge dz \wedge dx = -dy \wedge dx \wedge dz = dx \wedge dy \wedge dz$

Notice that all $n-$forms in $\mathbb{R}^n$ can be written $dx^1 \wedge dx^2 \wedge \cdots \wedge dx^n$ by using the (anti) commutativity of the wedge product and collecting like terms.

Now we know that the differential of a $0-$form is a 1-forms: $d : f \mapsto \frac{\partial f}{\partial x^i} dx^i$. We can inductively extend the definition of d to an operator that takes $k-$forms to $k+1$ forms. this operator will be introduced again in more generality; it is called **exterior differentiation** operator. Every k-form is a sum of terms that look like. We declare d to be linear over real numbers and then define

$d(fdx^{i_1} \wedge \cdots \wedge dx^{i_1}) = (df \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_1})$. For example, if in $\mathbb{R}^2$ we have a 1-form $\alpha = x^2dx + xydy$ then

$$\begin{aligned} d\alpha &= (x^2dx + xydy) \\ &= d\left(x^2\right) \wedge dx + d\left(xy\right) \wedge dy \\ &= 2xdx \wedge dx + (ydx + xdy) \wedge dy \\ &= ydx \wedge dy \end{aligned}$$

We are surprised that the answer is a multiple of $dx \wedge dy$ since in $\mathbb{R}^2$ all 2−forms have the form $f(x,y)dx \wedge dy$ for some function f.

Exercise 8.11 *Show that for a k form α on $\mathbb{R}^n$ and any other form β on $\mathbb{R}^n$ we have $d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^k d\beta$.*

Exercise 8.12 *Verify that $dd = 0$.*

Since $k-$forms are built up out of 1-forms using the wedge product we just need to know how to write a 1−form in new coordinates; but we already know how to do this.

8.6.2 Vector Analysis on $\mathbb{R}^3$

In $\mathbb{R}^3$ the space of 0−forms on some open domain U is, as we have said, just the smooth functions $C^\infty(U) = \Omega^0(U)$. The one forms may all be written (even globally) in the form $\theta = f_1dx + f_2dy + f_3dz$ for some smooth functions f_1 , f_2 and f_2. As we shall see in more detail later, a 1-form is exactly the kind of thing that can be integrated along an oriented path given by a (piecewise) smooth curve. The curve gives an orientation to its image (the path). A one form "wants to be integrated over a curve. If $c : (a,b) \to \mathbb{R}^3$ is such a curve then the oriented integral of θ along c is

$$\begin{aligned} \int_c \theta &= \int_c f_1dx + f_2dy + f_3dz \\ &= \int_a^b \left(f_1(c(t))\frac{dx}{dt} + f_2(c(t))\frac{dy}{dt} + +f_3(c(t))\frac{dz}{dt} \right) dt \end{aligned}$$

and as we know from calculus this is independent of the parameterization of the path given by the curve as long as the curve orients the path in the same direction.

Now in $\mathbb{R}^3$ all 2−forms β may be written $\beta = g_1dy \wedge dz + g_2dz \wedge dx + g_3dx \wedge dy$. The forms $dy \wedge dz, dz \wedge dx$, $dx \wedge dy$ form a basis (in the module sense) for the space of 2-forms on $\mathbb{R}^3$ just as dx, dy, dz form a basis for the 1−forms. This single form $dx \wedge dy \wedge dz$ a basis for the 3-forms in $\mathbb{R}^3$. Again, there are no higher order forms in $\mathbb{R}^3$. The funny order for the basis 2−forms is purposeful. Now a 2-form is exactly the kind of thing that wants to be integrated over a two dimensional subset of $\mathbb{R}^3$. Suppose that $\mathbf{x}(u,v)$ parameterizes a surface

$S \subset \mathbb{R}^3$ so we have a map $\mathbf{x} : U \to \mathbb{R}^3$. Then the surface is oriented by this parameterization and the oriented integral of β over S is

$$\begin{aligned}\int_S \beta &= \int_S g_1 dy \wedge dz + g_2 dz \wedge dx + g_3 dx \wedge dy \\ &= \int_U \left[g_1(\mathbf{x}(u,v)) \frac{dy \wedge dz}{du \wedge dv} + g_2 \left(\mathbf{x}(u,v)\right) \frac{dz \wedge dx}{du \wedge dv} + g_3 \left(\mathbf{x}(u,v)\right) \frac{dx \wedge dy}{du \wedge dv} \right] dudv\end{aligned}$$

where, for example, $\frac{dy \wedge dz}{du \wedge dv}$ is the determinant of the matrix $\begin{bmatrix} \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} \end{bmatrix}$.

Exercise 8.13 *Find the integral of $\beta = xdy \wedge dz + 1dz \wedge dx + xzdx \wedge dy$ over the sphere oriented by the parameterization given by the usual spherical coordinates ϕ, θ, ρ.*

The only 3-forms in R^3 are all of the form $\omega = hdx \wedge dy \wedge dz$ for some function h. With these we may integrate over any open (say bounded) open subset U which we may take to be given the usual orientation implied by the rectangular coordinates x, y, z. In this case

$$\begin{aligned}\int_U \omega &= \int_U hdx \wedge dy \wedge dz \\ &= \int_U hdxdydz.\end{aligned}$$

It is important to notice that $\int_U hdy \wedge dx \wedge dz = -\int_U hdydxdz$ since $\int_U hdydxdz = \int_U hdxdydz$ but $dy \wedge dx \wedge dz = -dx \wedge dy \wedge dz$.

In order to relate the calculus of differential forms on $\mathbb{R}^3$ we will need some ways to move between different ranks of forms by what is called the star operator and we will need a way to relate $1-$forms and vector fields. To a one form $\theta = f_1 dx + f_2 dy + f_3 dz$ we can obviously associate the vector field $\theta^\natural = f_1 \mathbf{i} + f_2 \mathbf{j} + f_3 \mathbf{k}$ and this works fine but we must be careful. The association depends on the notion of orthonormality provided by the dot product. Differential forms can be expressed in other coordinate systems. If θ is expressed in say spherical coordinates $\theta = \widetilde{f}_1 d\rho + \widetilde{f}_2 d\theta + \widetilde{f}_3 d\phi$ then it is **not** true that $\theta^\natural = \widetilde{f}_1 \mathbf{i} + \widetilde{f}_2 \mathbf{j} + \widetilde{f}_3 \mathbf{k}$ or even the more plausible $\theta^\natural = \widetilde{f}_1 \widehat{\rho} + \widetilde{f}_2 \widehat{\theta} + \widetilde{f}_3 \widehat{\phi}$ where $\widehat{\rho}, \widehat{\theta}, \widehat{\phi}$ are unit vectors fields in the coordinate directions[2]. Rather what is true is

$$\theta^\natural = \widetilde{f}_1 \widehat{\rho} + \widetilde{f}_2 \frac{1}{\rho} \widehat{\theta} + \widetilde{f}_3 \frac{1}{\rho \sin\theta} \widehat{\phi}$$

The sharping operator in $\mathbb{R}^3$ is defined on $1-$forms by

$$\begin{aligned} \natural : \quad dx &\to \mathbf{i} \\ \natural : \quad dy &\to \mathbf{j} \\ \natural : \quad dz &\to \mathbf{k} \end{aligned}$$

[2]Here θ is the polar angle ranging from 0 to π.

and extended bilinearity over functions. In good old fashioned tensor notation **sharping** (also called index raising) is $f_j \mapsto f^i := g^{ij}\, f_j$ where again the Einstein summation convention is applied. The inverse operation taking vector fields to $1-$forms is called **flatting** and is

$$\flat : f^i \frac{\partial}{\partial u^i} \mapsto f_j du^j := g_{ij} f^i du^j$$

but in rectangular coordinates we simply have the expected

$$\begin{array}{lccc} \flat : & \mathbf{i} & \mapsto & dx \\ \flat : & \mathbf{j} & \mapsto & dy \\ \flat : & \mathbf{k} & \mapsto & dz \end{array}.$$

We now come to our first connection with traditional vector calculus. If f is a smooth function then $df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz$ is a 1-form associated vector field is $\frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j} + \frac{\partial f}{\partial z}\mathbf{k}$ which is none other than the gradient $\operatorname{grad} f$. This flatting and sharping looks so easy as to seem silly but the formulas are only guaranteed to work if we use standard coordinates. In curvilinear coordinates (and more generally on a surface or Riemannian manifold) the formula for flatting is bit more complicated. For example, in spherical coordinates we have

$$\begin{array}{lccc} \natural : & d\rho & \mapsto & \widehat{\rho} \\ \natural : & \rho d\theta & \mapsto & \widehat{\theta} \\ \natural : & \rho \sin\theta d\phi & \mapsto & \widehat{\phi} \end{array}.$$

Sharping is always just the inverse of flatting.

A more general formula: Let u^1, u^2, u^3 be general curvilinear coordinate on (some open set in) $\mathbb{R}^3$ and $\frac{\partial}{\partial u^1}, \frac{\partial}{\partial u^2}, \frac{\partial}{\partial u^3}$ are the corresponding basis vector fields then defining then the Euclidean metric components in these coordinates to be $g_{ij} = \langle \frac{\partial}{\partial u^i}, \frac{\partial}{\partial u^j} \rangle$. We also need $(g^{ij}\ )$ the inverse of (g_{ij}).Using the summation convention we have

$$\natural : \theta = f_j du^j \mapsto f^i \frac{\partial}{\partial u^i} := g^{ij}\ f_j \frac{\partial}{\partial u^i}.$$

As an example, we can derive the familiar formula for the gradient in spherical coordinate by first just writing f in the new coordinates $f(\rho, \theta, \phi) := f(x(\rho, \theta, \phi), y(\rho, \theta, \phi), z(\rho, \theta, \phi))$ and then sharping the differential:

$$df = \frac{\partial f}{\partial \rho} d\rho + \frac{\partial f}{\partial \theta} d\theta + \frac{\partial f}{\partial \phi} d\phi$$

to get

$$\operatorname{grad} f = (df)^\natural = \frac{\partial f}{\partial \rho}\frac{\partial}{\partial \rho} + \frac{1}{\rho}\frac{\partial f}{\partial \theta}\frac{\partial}{\partial \theta} + \frac{1}{\rho \sin\theta}\frac{\partial f}{\partial \phi}\frac{\partial}{\partial \phi}$$

where we have used

$$(g^{ij}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \rho & 0 \\ 0 & 0 & \rho\sin\theta \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\rho} & 0 \\ 0 & 0 & \frac{1}{\rho\sin\theta} \end{bmatrix}$$

8.6.3 Star Operator in $\mathbb{R}^3$

In order to proceed to the point of including the curl and divergence we need a way to relate 2-forms with vector fields. This part definitely depends on the fact that we are talking about forms in $\mathbb{R}^3$. We associate to a $2-$form $\eta = g_1 dy \wedge dz + g_2 dz \wedge dx + g_3 dx \wedge dy$ the vector field $X = g_1\mathbf{i} + g_2\mathbf{j} + g_3\mathbf{k}$. This should be thought of as first applying the so called star operator to η to get the one form $g_1 dx + g_2 dy + g_3 dz$ and then applying the sharping operator to get the resulting vector field. Again things are more complicated in curvilinear coordinates because of the role of the hidden role of the dot product and the lack of orthonormality of curvilinear frame fields. The star operator $*$ works on any form and is based on the prescription valid in rectangular coordinates given below:

$$\begin{array}{ccc}
f & \mapsto & f dx \wedge dy \wedge dz \\
f_1 dx + f_2 dy + f_3 dz & \mapsto & f_1 dy \wedge dz + f_2 dz \wedge dx + f_3 dx \wedge dy \\
g_1 dy \wedge dz + g_2 dz \wedge dx + g_3 dx \wedge dy & \mapsto & g_1 dx + g_2 dy + g_3 dz \\
f dx \wedge dy \wedge dz & \mapsto & f
\end{array}$$

So $*$ is a map that takes $k-$forms to $(3-k)-$ forms[3]. It is easy to check that in our current rather simple context we have $*(*\beta) = \beta$ for any form on $U \subset \mathbb{R}^3$. This always true up to a possible sign.

Now we can see how the divergence of a vector field comes about. First flat the vector field, say $X = f_1\mathbf{i}+f_2\mathbf{j}+f_3\mathbf{k}$, to obtain $X^\flat = f_1 dx + f_2 dy + f_3 dz$ and then apply the star operator to obtain $f_1 dy \wedge dz + f_2 dz \wedge dx + f_3 dx \wedge dy$ and then finally apply exterior differentiation! We then obtain

$$\begin{aligned}
&d\left(f_1 dy \wedge dz + f_2 dz \wedge dx + f_3 dx \wedge dy\right) \\
&= df_1 \wedge dy \wedge dz + df_2 \wedge dz \wedge dx + df_3 \wedge dx \wedge dy \\
&= \left(\frac{\partial f_1}{\partial x} dx + \frac{\partial f_1}{\partial y} dy + \frac{\partial f_1}{\partial z} dz\right) \wedge dy \wedge dz + \text{the obvious other two terms} \\
&= \frac{\partial f_1}{\partial x} dx \wedge dy \wedge dz + \frac{\partial f_2}{\partial x} dx \wedge dy \wedge dz + \frac{\partial f_3}{\partial x} dx \wedge dy \wedge dz \\
&= \left(\frac{\partial f_1}{\partial x} + \frac{\partial f_2}{\partial x} + \frac{\partial f_3}{\partial x}\right) dx \wedge dy \wedge dz.
\end{aligned}$$

Now we see the divergence appearing. In fact, apply the star operator one more time and we get the function $\operatorname{div} X = \frac{\partial f_1}{\partial x} + \frac{\partial f_2}{\partial x} + \frac{\partial f_3}{\partial x}$. We are thus lead to the formula $*d * X^\flat = \operatorname{div} X$.

[3] As defined more generally on a Riemannian manifold of dimension n the star operator maps $\Omega^k(M)$ to $\Omega^{n-k}(M)$.

What about the curl? For this we just take $dX^\flat$ to get

$$\begin{aligned}
&d\left(f_1dx + f_2dy + f_3dz\right)\\
&= df_1 \wedge dx + df_2 \wedge dy + df_3 \wedge dz\\
&= \left(\frac{\partial f_1}{\partial x}dx + \frac{\partial f_2}{\partial y}dy + \frac{\partial f_3}{\partial z}dz\right) \wedge dx + \text{the obvious other two terms}\\
&= \frac{\partial f_2}{\partial y}dy \wedge dx + \frac{\partial f_3}{\partial z}dz \wedge dx + \frac{\partial f_1}{\partial x}dx \wedge dy + \frac{\partial f_3}{\partial z}dz \wedge dy\\
&+ \frac{\partial f_1}{\partial x}dx \wedge dz + \frac{\partial f_2}{\partial y}dy \wedge dz\\
&= \left(\frac{\partial f_2}{\partial y} - \frac{\partial f_3}{\partial z}\right) dy \wedge dz + \left(\frac{\partial f_3}{\partial z} - \frac{\partial f_1}{\partial x}\right) dz \wedge dx + \left(\frac{\partial f_1}{\partial x} - \frac{\partial f_2}{\partial y}\right) dx \wedge dy
\end{aligned}$$

and then apply the star operator to get back to vector fields obtaining $\left(\frac{\partial f_2}{\partial y} - \frac{\partial f_3}{\partial z}\right)\mathbf{i}+\left(\frac{\partial f_3}{\partial z} - \frac{\partial f_1}{\partial x}\right)\mathbf{j}+\left(\frac{\partial f_1}{\partial x} - \frac{\partial f_2}{\partial y}\right)\mathbf{k} = \operatorname{curl} X$. In short we have

$$*dX^\flat = \operatorname{curl} X$$

Exercise 8.14 *Show that the fact that $dd = 0$ leads to each of the following the familiar facts:*

$$\begin{aligned}
\operatorname{curl}(\operatorname{grad} f) &= 0\\
\operatorname{div}(\operatorname{curl} X) &= 0
\end{aligned}$$

Recall that for general curvilinear coordinates $\left(u^i\right)$ on a domain in $\mathbb{R}^3$ we have defined $g_{ij} := \langle \frac{\partial}{\partial u^i}, \frac{\partial}{\partial u^j}\rangle$ and also g^{ij} is defined so that $g^{ir}g_{rj} = \delta^i_j$. Notice that in the special case that u^1, u^2, u^3 are rectangular orthonormal coordinates such as x, y, z then $g_{ij} = \delta_{ij}$. The index placement is important in general but for the Kronecker delta we simply have $\delta_{ij} = \delta^i_j = \delta^{ij} = 1$ if $i = j$ and 0 otherwise.

Let g denote the determinant of the matrix $[g_{ij}]$. The 3-form $dx \wedge dy \wedge dz$ is called the (oriented) volume element of R^3 for obvious reasons. Every 3-form is a function time this volume form and integration of a $3-$form over sufficiently nice subset (say open) is given by $\int_D \omega = \int_D f dx \wedge dy \wedge dz = \int_D f dxdydz$ (usual Riemann integral). Let us denote $dx \wedge dy \wedge dz$ this by $\mathrm{d}V$. Of course, $\mathrm{d}V$ is not to be considered as the exterior derivative of some object V. Now we will show that in curvilinear coordinates $\mathrm{d}V = \sqrt{g}du^1 \wedge du^2 \wedge du^3$. In any case there must be some function f such that $\mathrm{d}V = f du^1 \wedge du^2 \wedge du^3$. Let's discover this function.

$$\begin{aligned}
dV = dx \wedge dy \wedge dz &= \frac{\partial x}{\partial u^i}du^i \wedge \frac{\partial y}{\partial u^j}du^j \wedge \frac{\partial z}{\partial u^k}du^k\\
&= du^1 \wedge du^2 \wedge du^3\\
&\frac{\partial x}{\partial u^i}\frac{\partial y}{\partial u^j}\frac{\partial z}{\partial u^k}\epsilon^{123}_{ijk}du^1 \wedge du^2 \wedge du^3
\end{aligned}$$

where ϵ_{123}^{ijk} is the sign of the permutation $123 \mapsto ijk$. Now we see that $\frac{\partial x}{\partial u^i}\frac{\partial y}{\partial u^j}\frac{\partial z}{\partial u^k}\epsilon_{123}^{ijk}$ is $\det(\frac{\partial x^i}{\partial u^j})$. On the other hand, $g_{ij} = \frac{\partial x^k}{\partial u^i}\frac{\partial x^k}{\partial u^j}$ or $(g_{ij}) = (\frac{\partial x^i}{\partial u^i})(\frac{\partial x^i}{\partial u^j})$. Thus $g := \det(g_{ij}) = \det(\frac{\partial x^i}{\partial u^i})\det(\frac{\partial x^i}{\partial u^j}) = \left(\det(\frac{\partial x^i}{\partial u^i})\right)^2$. From this we get $\det(\frac{\partial x^i}{\partial u^i}) = \sqrt{g}$ and so $\mathrm{d}V = \sqrt{g}du^1 \wedge du^2 \wedge du^3$. A familiar example is the case when (u^1, u^2, u^3) is spherical coordinates ρ, θ, ϕ then

$$\mathrm{d}V = \rho^2 \sin\theta d\rho \wedge d\theta \wedge d\phi$$

and if an open set $U \subset \mathbb{R}^3$ is parameterized by these coordinates then

$$\begin{aligned}\int_U f\mathrm{d}V &= \int_U f(\rho,\theta,\phi)\rho^2 \sin\theta d\rho \wedge d\theta \wedge d\phi \\ &= \int_U f(\rho,\theta,\phi)\rho^2 \sin\theta d\rho d\theta d\phi \text{ (now a Riemann integral)} \\ &\text{which is equal to } \int_U f(x,y,z)dxdydz.\end{aligned}$$

Exercise 8.15 *Show that if* $\mathbf{B} = f_1\mathbf{i} + f_2\mathbf{j} + f_3\mathbf{k}$ *is given in general curvilinear coordinates by*

$$\mathbf{B} = b^i \frac{\partial}{\partial u^i}$$

then $\operatorname{div}\mathbf{B} = \frac{1}{\sqrt{g}}\frac{\partial(\sqrt{g}b^i)}{\partial u^i}$ *(summation implied).*

Exercise 8.16 *Discover the local expression for* $\nabla^2 f = \operatorname{div}(\operatorname{grad} f)$ *in any coordinates by first writing* $\operatorname{div}(\operatorname{grad} f)$. *Notice that* $\nabla^2 f = *d*(df)^\natural$.

The reader should have notices that (for the case of $\mathbb{R}^3$) we have explained how to integrate 1-forms along curves, 2−forms over surfaces, and 3 forms over open domains in $\mathbb{R}^3$. To round things out define the "integral" of a function over and ordered pair of points (p_1, p_2) as $f(p_2) - f(p_1)$. Let M denote one of the following: (1) A curve oriented by a parameterization, (2) a surface oriented by a parameterization and having a smooth curve as boundary, (3) a domain in $\mathbb{R}^3$ oriented by the usual orientation on $\mathbb{R}^3$ and having a smooth surface as a boundary. Now let ∂M denote in the first case the set $\{c(a), c(b)\}$ of beginning and ending points of the curve in the first case, the counterclockwise traversed boundary curve of M if M is a surface as in the second case, or finally the surface which is the boundary of M (assumed to be nice) when M is a domain of $\mathbb{R}^3$. Finally, let ω be a 0-form in case (1), a 1−form in case (2) and a 2 form in case (3).As a special case of Stokes' theorem on manifolds we have the following.

$$\int_M d\omega = \int_{\partial M} d\omega$$

The three cases become

$$\int_{M=c} df = \int_{\partial M=\{c(a),c(b)\}} f \text{ (which is} = \int_a^b f(c(t))dt = f(c(b)) - f(c(a)))$$
$$\int_S d\alpha = \int_{\partial M=c} \alpha$$
$$\int_D d\omega = \int_{S=\partial D} \omega$$

If we go to the trouble of writing these in terms of vector fields associated to the forms in an appropriate way we get the following familiar theorems (using standard notation):

$$\int_c \nabla f \cdot d\mathbf{r} = f(\mathbf{r}(b)) - f(\mathbf{r}(a))$$
$$\iint_S curl(\mathbf{X}) \times \mathbf{dS} = \oint_c X \cdot d\mathbf{r} \text{ (Stokes' theorem)}$$
$$\iiint_D div\mathbf{X}dV = \iint_S \mathbf{X} \cdot \mathbf{dS} \text{ (Divergence theorem)}$$

Similar and simpler things can be done in $\mathbb{R}^2$ leading for example to the following version of Green's theorem for a planar domain D with (oriented) boundary $c = \partial D$.

$$\int_D \left(\frac{\partial M}{\partial y} - \frac{\partial M}{\partial y} \right) dx \wedge dy = \int_D d(Mdx + Ndy) = \int_c Mdx + Ndy$$

All of the standard integral theorems from vector calculus are special cases of the general Stokes' theorem that we introduce later in this chapter.

8.6.4 Differential forms on a general differentiable manifold

Now let us give a presentation of the general situation of differential forms on a manifold. By now it should be clear that algebraic constructions can be done fiberwise for the tangent bundle or some other vector bundle and then by considering fields of such constructions as we move to the module level. Alternatively we can start with of sections of vector bundles over a given manifold and work immediately in the realm of $C^\infty(M)$. The two approaches are equivalent for finite rank vector bundles over finite dimensional manifolds. In the following we try to hedge our bet a little attempting to handle both cases at once by just working with abstract modules which later may turn out to be vector spaces or modules of section of a vector bundle.

Definition 8.10 *Let* V *and* F *be modules over a ring* R *or an* $\mathbb{F}-$*algebra* R *where* $\mathbb{F} = \mathbb{R}$ *or* $\mathbb{C}$. *A k-multilinear map* $\alpha : \mathrm{V} \times \cdots \times \mathrm{V} \to \mathrm{W}$ *is called* ***alternating*** *if* $\alpha(\mathrm{v}_1, ..., \mathrm{v}_k) = 0$ *whenever* $\mathrm{v}_i = \mathrm{v}_j$ *for some* $i \neq j$. *The space of all* ***alternating***

k-multilinear maps into W *will be denoted by* $L^k_{alt}(\mathrm{V};\mathrm{W})$ *or by* $L^k_{alt}(\mathrm{W})$ *if the* $\mathrm{F} = \mathsf{R}$.

Since we are dealing in the cases where R is either one of the fields $\mathbb{R}$ and $\mathbb{C}$ (which have characteristic zero) or an algebra of real or complex valued functions, it is easy to see that **alternating** k-multilinear maps are the same as (completely) **antisymmetric** k-multilinear maps which are defined by the property that for any permutation σ of the letters $1, 2, ..., k$ we have

$$\omega(\mathrm{w}_1, \mathrm{w}_2, .., \mathrm{w}_k) = sgn(\sigma)\omega(\mathrm{w}_{\sigma_1}, \mathrm{w}_{\sigma_2}, .., \mathrm{w}_{\sigma_k}).$$

Definition 8.11 *The* ***antisymmetrization map*** $Alt^k : T^0{}_k(\mathrm{W}) \to L^k_{alt}(\mathrm{W})$ *is defined by* $Alt(\omega)(\mathrm{w}_1, \mathrm{w}_2, .., \mathrm{w}_k) := \frac{1}{k!}\sum_\sigma sgn(\sigma)\omega(\mathrm{w}_{\sigma_1}, \mathrm{w}_{\sigma_2}, .., \mathrm{w}_{\sigma_k})$.

Now given $\omega \in L^r_{alt}(\mathrm{V})$ and $\eta \in L^s_{alt}(\mathrm{V})$ we define their **exterior product** or **wedge product** $\omega \wedge \eta \in L^{r+s}_{alt}(\mathrm{V})$ by the formula

$$\omega \wedge \eta := \frac{(r+s)!}{r!s!} Alt^{r+s}(\omega \otimes \eta)$$

Written out this is

$$\omega \wedge \eta(\mathrm{v}_1, ..., \mathrm{v}_r, \mathrm{v}_{r+1}, ..., \mathrm{v}_{r+s}) := \frac{1}{r!s!}\sum_\sigma \mathrm{sgn}(\sigma)\omega(\mathrm{v}_{\sigma_1}, ..., \mathrm{v}_{\sigma_r})\eta(\mathrm{v}_{\sigma_{r+1}}, ..., \mathrm{v}_{\sigma_{r+s}})$$

It is an exercise in combinatorics that we also have

$$\omega\wedge\eta(\mathrm{v}_1, ..., \mathrm{v}_r, \mathrm{v}_{r+1}, ..., \mathrm{v}_{r+s}) := \sum_{r,s-\text{shuffles }\sigma} \mathrm{sgn}(\sigma)\omega(\mathrm{v}_{\sigma_1}, ..., \mathrm{v}_{\sigma_r})\eta(\mathrm{v}_{\sigma_{r+1}}, ..., \mathrm{v}_{\sigma_{r+s}}).$$

In the latter formula we sum over all permutations such that $\sigma_1 < \sigma_2 < ... < \sigma_r$ and $\sigma_{r+1} < \sigma_{r+2} < .. < \sigma_{r+s}$. This kind of permutation is called an $r, s-$**shuffle** as indicated in the summation. The most important case is for $\omega, \eta \in L^1_{alt}(\mathrm{V})$ in which case

$$(\omega \wedge \eta)(v, w) = \omega(v)\eta(w) - \omega(w)\eta(v)$$

This is clearly an antisymmetric multilinear map, which is just what we call antisymmetric in the case of two 2 variables.

If we use a basis $\varepsilon^1, \varepsilon^2, ..., \varepsilon^n$ for V^* it is easy to show that the set of all elements of the form $\varepsilon^{i_1} \wedge \varepsilon^{i_2} \wedge \cdots \wedge \varepsilon^{i_k}$ with $i_1 < i_2 < ... < i_k$ is a basis for $L^k_{alt}(\mathrm{V})$. Thus for any $\omega \in L^k_{alt}(\mathrm{V})$

$$\omega = \sum_{i_1 < i_2 < ... < i_k} a_{i_1 i_2, .., i_k}\varepsilon^{i_1} \wedge \varepsilon^{i_2} \wedge \cdots \wedge \varepsilon^{i_k}$$

Remark 8.3 *In order to facilitate notation we will abbreviate a sequence of k integers, say $i_1, i_2, ..., i_k$, from the set $\{1, 2, ..., \dim(\mathrm{V})\}$ as I and $\varepsilon^{i_1}\wedge\varepsilon^{i_2}\wedge\cdots\wedge\varepsilon^{i_k}$ is written as ε^I. Also, if we require that $i_1 < i_2 < ... < i_k$ we will write $\vec{I}$. We*

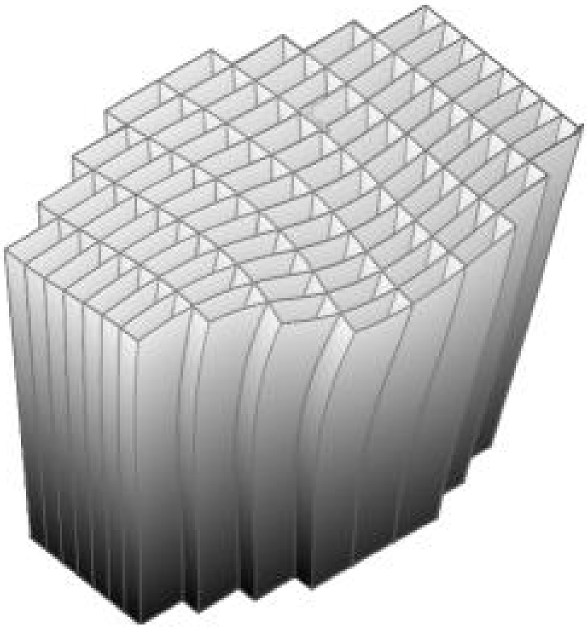


Figure 8.1: 2-form as tubes forming honeycomb.

will freely use similar self explanatory notation as we go along without further comment. For example, the above equation can be written as

$$\omega = \sum a_{\vec{I}} \varepsilon^{\vec{I}}$$

Lemma 8.1 $L^k_{alt}(\mathrm{V}) = 0$ *if* $k > n = \dim(\mathrm{V})$.

Proof. Easy exercise. ■

If one defines $L^0_{alt}(\mathrm{V})$ to be the scalars R and recalling that $L^1_{alt}(\mathrm{V}) = \mathrm{V}^*$ then the sum

$$L_{alt}(\mathrm{V}) = \bigoplus_{k=0}^{\dim(M)} L^k_{alt}(\mathrm{V})$$

is made into an algebra via the wedge product just defined.

Proposition 8.5 *For* $\omega \in L^r_{alt}(V)$ *and* $\eta \in L^s_{alt}(V)$ *we have* $\omega \wedge \eta = (-1)^{rs} \eta \wedge \omega \in L^{r+s}_{alt}(V)$.

Let M be a smooth manifold modeled on M. Let $n = \dim M$ and let us agree that $n := \infty$ in case M is an infinite dimensional Banach space. We now bundle together the various spaces $L^k_{alt}(T_pM)$. That is we form the natural bundle $L^k_{alt}(TM)$ that has as its fiber at p the space $L^k_{alt}(T_pM)$. Thus $L^k_{alt}(TM) = \bigsqcup_{p\in M} L^k_{alt}(T_pM)$.

Exercise 8.17 *Exhibit, simultaneously the smooth structure and vector bundle structure on $L^k_{alt}(TM) = \bigsqcup_{p\in M} L^k_{alt}(T_pM)$.*

Let the smooth sections of this bundle be denoted by

$$\Omega^k(M) = \Gamma(M; L^k_{alt}(TM)). \tag{8.3}$$

and sections over $U \subset M$ by $\Omega^k_M(U)$. This space is a module over the algebra of smooth functions $C^\infty(M) = \mathcal{F}(U)$. We have the *direct* sum

$$\Omega_M(U) = \textstyle\sum_{n=0}^{\dim M} \Omega^k_M(U) = \Gamma\left(U, \sum_{n=0}^{\dim M} L^k_{alt}(TM)\right).$$

which is a $\mathbb{Z}^+$-graded algebra under the **exterior product** which is defined using the exterior product on each fiber:

$$(\omega \wedge \eta)(p) := \omega(p) \wedge \eta(p).$$

Definition 8.12 *The sections of the bundle $\Omega_M(U)$ are called* ***differential forms*** *on U. We identify $\Omega^k_M(U)$ with the obvious subspace of $\Omega_M(U) = \sum_{n=0}^{\dim M} \Omega^k_M(U)$. A differential form in $\Omega^k_M(U)$ is said to be* ***homogeneous*** *of* ***degree*** *k and is referred to a "k-form". If $U = M$ we write $\Omega^k(M)$*

Whenever convenient we may extend this to a sum over all $n \in \mathbb{Z}$ by defining (as before) $\Omega^k_M(U) := 0$ for $n < 0$ and $\Omega^k_M(U) := 0$ if $n > \dim(M)$. Of course, we have made the trivial extension of $\wedge$ to the **$\mathbb{Z}$-graded algebra** by declaring that $\omega \wedge \eta = 0$ if either η or ω is homogeneous of negative degree.

The assignment $U \mapsto \Omega^k_M(U)$ is a presheaf of modules over C^∞_M and $\Omega_M(U)$ is a sheaf of graded algebras. Sections $\Omega^k_M(U)$ are called **smooth differential forms** over U of degree k ($k-$forms for short).

Just as a tangent vector is the infinitesimal version of a (parameterized) curve through a point $p \in M$ so a covector at $p \in M$ is the infinitesimal version of a function defined near p. At this point one must be careful. It is true that for any single covector $\alpha_p \in T_pM$ there always exists a function f such $df_p = \alpha_p$. But if $\alpha \in \Omega^1_M(U)$ then it is not necessarily true that there is a function $f \in C^\infty(U)$ such that $df = \alpha$. Now if $f_1, f_2, ..., f_k$ are smooth functions then one way to picture the situation is by think of the intersecting family of level sets of the functions $f_1, f_2, ..., f_k$ which generically form a sort of "egg crate" structure. The infinitesimal version of this is the picture of a sort of straightened out "linear egg crate structure" which may be thought of as existing in the tangent space at a point. This is the rough intuition for $df_1|_p \wedge ... \wedge df_k|_p$ and the $k-$form $df_1 \wedge ... \wedge df_k$ is a field of such structures which somehow fit the level sets of the family $f_1, f_2, ..., f_k$. Of course, $df_1 \wedge ... \wedge df_k$ is a very special kind of $k-$form. In general a $k-$form over U may not arise from a family of functions. The extent to which something like this would be true involves integrability and is one approach to the Frobenius' integrability theory to be studied later in the book.

Pull-back of a differential form.

Given any smooth map $f : M \to N$ we can define the **pull-back** map $f^* : \Omega(N) \to \Omega(M)$ as follows:

Definition 8.13 *Let $\eta \in \Omega^k(N)$. For vectors $v_1, ..., v_k \in T_pM$ define*

$$(f^*\eta)(p)(v_1, ..., v_k) = \eta_{f(p)}(T_pv_1, ..., T_pv_k)$$

*then the map $f^*\eta : p \to (f^*\eta)(p)$ is a differential form on M. $f^*\eta$ is called the **pullback of** η **by** f.*

The pull-back is a very natural operation as exhibited by the following propositions.

Proposition 8.6 *Let $f : M \to N$ smooth map and $\eta_1, \eta_2 \in \Omega(N)$ we have*

$$f^*(\eta_1 \wedge \eta_2) = f^*\eta_1 \wedge f^*\eta_2$$

Proof: Exercise

Proposition 8.7 *Let $f_1 : M \to N$ and $f_2 : N \to P$ be smooth maps. Then for any smooth differential form $\eta \in \Omega(P)$ we have $(f_1 \circ f_2)^* \eta = f_2^*(f_1^*\eta)$. Thus $(f_1 \circ f_2)^* = f_2^* \circ f_1^*$ and so the assignments $f \mapsto f^*$ and $M \mapsto \Omega(M)$ give us a contravariant functor in the smooth category.*

In case S is a regular submanifold of M then the inclusion map $\iota : S \hookrightarrow M$ which maps $p \in S$ to the very same point $p \in S$ is just a fancy way of saying that every point of S is also a point of M. Similarly, any smooth curve $c : I \to S$ is also a smooth curve into M which could be written $\iota \circ c : I \to M$. It seems silly to distinguish between $\iota \circ c$ and c and indeed one usually identifies these. In the same way, we should find it natural to identify T_pS with $T_p\iota(T_pS)$ for any $p \in S$. In other words, we normally do not distinguish between a vector $v_p \in T_pS$ and $T_p\iota(v_p)$. Both are then written as v_p. So then the tangent bundle of S is actually a subset (in fact, a subbundle) of TM. With this in mind we must realize that for any $\alpha \in \Omega^k(M)$ the form $\iota^*\alpha$ is just the restriction of α to vectors tangent to S. In particular, if $U \subset M$ is open and $\iota : U \hookrightarrow M$ then $\iota^*\alpha = \alpha|_U$.

The local expression for the pullback is described as follows. First we define

$$\begin{aligned} \epsilon^I_L &= \epsilon^{i_1 \dots i_k}_{l_1 \dots l_k} \\ &= \begin{cases} 0 & \text{if} \quad I \text{ is not a permutation of } J \\ 1 & \text{if} \quad I \text{ is an even permutation of } J \\ -1 & \text{if} \quad I \text{ is an odd permutation of } J \end{cases} \end{aligned}$$

Let $U, \mathtt{x}$ be a chart on M and $V, \mathtt{y}$ a chart on N with $\mathtt{x}(U) \subset V$ then writing $\eta = \sum b_{\vec{J}} dy^{\vec{J}}$ and abbreviating $\frac{\partial\left(y^{j_1} \circ f\right)}{\partial x^{i_1}}$ to simply $\frac{\partial y^{j_1}}{\partial x^{i_1}}$ etc., we have

$$
\begin{aligned}
f^*\eta &= \sum b_{\vec{J}} \circ f \; d\left(y^{\vec{J}} \circ f\right) \\
&= \sum_{\vec{J}} b_{\vec{J}} \circ f \; \left(\sum_{i_1} \frac{\partial y^{j_1}}{\partial x^{i_1}} dx^{i_1}\right) \wedge \cdots \wedge \left(\sum_{i_k} \frac{\partial y^{j_1}}{\partial x^{i_1}} dx^{i_k}\right) \\
&= \sum_{\vec{J}} \sum_{I} b_{\vec{J}} \circ f \left(\frac{\partial y^{j_1}}{\partial x^{i_1}} \cdots \frac{\partial y^{j_1}}{\partial x^{i_1}}\right) \epsilon^I_L dx^{l_1} \wedge \cdots \wedge dx^{l_k} \\
&= \sum b_{\vec{J}} \circ f \frac{dy^{\vec{J}}}{dx^{\vec{L}}} dx^{\vec{L}}
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{dy^{\vec{J}}}{dx^{\vec{L}}} &= \frac{dy^{j_1} \wedge \cdots \wedge dy^{j_k}}{dx^{l_1} \wedge \cdots \wedge dx^{l_k}} \\
&= \det \begin{bmatrix} \frac{\partial y^{j_1}}{\partial x^{l_1}} & \cdots & \frac{\partial y^{j_1}}{\partial x^{l_k}} \\ \vdots & & \vdots \\ \frac{\partial y^{j_k}}{\partial x^{l_1}} & \cdots & \frac{\partial y^{j_k}}{\partial x^{l_k}} \end{bmatrix}
\end{aligned}
$$

Since this highly combinatorial notation is a bit intimidating at first sight we work out the case where $\dim M = 2$, $\dim N = 3$ and $k = 2$. As a warm up notice that since $dx^1 \wedge dx^1 = 0$ we have

$$
\begin{aligned}
d\left(y^2 \circ f\right) \wedge d\left(y^3 \circ f\right) &= \frac{\partial y^2}{\partial x^i} dx^i \wedge \frac{\partial y^3}{\partial x^j} dx^j \\
&= \frac{\partial y^2}{\partial x^i} \frac{\partial y^3}{\partial x^j} dx^i \wedge dx^j \\
&= \frac{\partial y^2}{\partial x^1} \frac{\partial y^3}{\partial x^2} dx^1 \wedge dx^2 + \frac{\partial y^2}{\partial x^2} \frac{\partial y^3}{\partial x^1} dx^2 \wedge dx^1 \\
&= \left(\frac{\partial y^2}{\partial x^1} \frac{\partial y^3}{\partial x^2} - \frac{\partial y^2}{\partial x^2} \frac{\partial y^3}{\partial x^1}\right) dx^2 \wedge dx^3 \\
&= \frac{d\left(y^2 \circ f\right) \wedge d\left(y^3 \circ f\right)}{dx^2 \wedge dx^3} = \frac{dy^2 \wedge dy^3}{dx^2 \wedge dx^3}
\end{aligned}
$$

Using similar expressions, we have

$$
\begin{aligned}
f^*\eta &= f^*\left(b_{23} dy^2 \wedge dy^3 + b_{13} dy^1 \wedge dy^3 + b_{12} dy^1 \wedge dy^2\right) \\
&= b_{23} \circ f \sum \frac{\partial y^2}{\partial x^i} dx^i \wedge \frac{\partial y^3}{\partial x^j} dx^j + b_{13} \circ f \sum \frac{\partial y^1}{\partial x^i} dx^i \wedge \frac{\partial y^3}{\partial x^j} dx^j \\
&+ b_{12} \circ f \sum \frac{\partial y^1}{\partial x^i} dx^i \wedge \frac{\partial y^2}{\partial x^j} dx^j \\
&= b_{23} \circ f \frac{dy^2 \wedge dy^3}{dx^1 \wedge dx^2} + b_{13} \circ f \frac{dy^1 \wedge dy^3}{dx^1 \wedge dx^2} + b_{12} \circ f \frac{dy^1 \wedge dy^2}{dx^1 \wedge dx^2}.
\end{aligned}
$$

Remark 8.4 *Notice the space $\Omega^0_M(U)$ is just the space of smooth functions $C^\infty(U)$ and so unfortunately we have several notations for the same set: $C^\infty(U) = \Omega^0_M(U) = \mathfrak{T}^0_0(U)$. The subscript M will be omitted where confusion is unlikely to result.*

All that follows and much of what we have done so far works well for $\Omega_M(U)$ whether $U = M$ or not and will also respect restriction maps. Thus we will simply write Ω_M instead of $\Omega_M(U)$ or $\Omega(M)$ and $\mathfrak{X}_M$ instead of $\mathfrak{X}(U)$ so forth (recall remark **??**). In fact, the exterior derivative d defined below commutes with restrictions and so is really a presheaf map.

The algebra of smooth differential forms $\Omega(U)$ is an example of a $\mathbb{Z}$ graded algebra over the ring $C^\infty(U)$ and is also a graded vector space over $\mathbb{R}$. We have for each $U \subset M$

1) The direct sum decomposition

$$\Omega(U) = \cdots \oplus \Omega^{-1}(U) \oplus \Omega^0(U) \oplus \Omega^1(U) \oplus \Omega^2(U) \cdots$$

where $\Omega^k(U) = 0$ if $k < 0$ or if $k > \dim(U)$;

2) The exterior product is a graded product:

$$\alpha \wedge \beta \in \Omega^{k+l}(U) \text{ for } \alpha \in \Omega^k(U) \text{ and } \beta \in \Omega^l(U)$$

that is

3) graded commutative: $\alpha \wedge \beta = (-1)^{kl} \beta \wedge \alpha$ for $\alpha \in \Omega^k(U)$ and $\beta \in \Omega^l(U)$.

Each of these is natural with respect to restriction and so we have a **presheaf of graded algebras**.

8.6.5 Exterior Derivative

Here we will define and study the exterior derivative d. First a useful general definition:

Definition 8.14 *A graded derivation of degree r on $\Omega := \Omega_M$ is a (pre)sheaf map $\mathcal{D} : \Omega_M \to \Omega_M$ such that for each $U \subset M$,*

$$\mathcal{D} : \Omega^k_M(U) \to \Omega^{k+r}_M(U)$$

and such that for $\alpha \in \Omega^k(U)$ and $\beta \in \Omega(U)$ we have graded (anti) commutativity:

$$\mathcal{D}(\alpha \wedge \beta) = \mathcal{D}\alpha \wedge \beta + (-1)^{kr} \alpha \wedge \mathcal{D}\beta.$$

Along lines similar to our study of tensor derivations one can show that a graded derivation of Ω_M is completely determined by, and can be defined by it action on 0-forms (functions) and 1-forms. In fact, since every form can be locally built out of functions and exact one forms, i.e. differentials, we only need to know the action on 0-forms and exact one forms to determine the graded derivation.

Lemma 8.2 *Suppose $\mathcal{D}_1 : \Omega^k_M(U) \to \Omega^{k+r}_M(U)$ and $\mathcal{D}_2 : \Omega^k_M(U) \to \Omega^{k+r}_M(U)$ are defined for each open set $U \subset M$ and satisfy the following for $i = 1$ and 2 and all k:*

1. *$\mathcal{D}_i$ is $\mathbb{R}$ linear;*
2. *$\mathcal{D}_i(\alpha \wedge \beta) = \mathcal{D}_i\alpha \wedge \beta + (-1)^{kr}\alpha \wedge \mathcal{D}_i\beta$;*
3. *$\mathcal{D}_i$ is natural with respect to restriction:*

$$\begin{array}{ccc} \Omega^k(U) & \stackrel{\mathcal{D}_i}{\to} & \Omega^{k+r}(U) \\ \downarrow & & \downarrow \\ \Omega^k(V) & \stackrel{\mathcal{D}_i}{\to} & \Omega^{k+r}(V) \end{array}$$

for all open $V \subset U$;

Then if $\mathcal{D}_1 = \mathcal{D}_2$ on Ω^0 and Ω^1 then $\mathcal{D}_1 = \mathcal{D}_2$. If Ω^1 is generated by $\mathcal{D}_1(\Omega^0)$ and if $\mathcal{D}_1 = \mathcal{D}_2$ on Ω^0 then $\mathcal{D}_1 = \mathcal{D}_2$.

Exercise 8.18 *Prove Lemma 8.2 above.*

The differential d defined by

$$df(X) = Xf \text{ for } X \in \mathfrak{X}_M \tag{8.4}$$

is a map $\Omega^0_M \to \Omega^1_M$. This map can be extended to a degree one map from the graded space Ω_M to itself. Degree one means that writing $d : \Omega^0_M \to \Omega^1_M$ as $d_0 : \Omega^0_M \to \Omega^1_M$ we can find maps $d_i : \Omega^i_M \to \Omega^{i+1}_M$ that combine to give d.

Theorem 8.2 *Let M a finite dimensional smooth manifold. There is a unique degree 1 graded derivation $\Omega_M \to \Omega_M$ such that for each $f \in C^\infty(U) = \Omega^0_M(U)$ we have that df coincides with the usual differential and such that $dd = 0$.*

Proof. Let $U, \mathtt{x}$ be a coordinate system. On a 0-form (a smooth function) we just define $d_\mathtt{x} f$ to be the usual differential given by $df = \sum \frac{\partial f}{\partial x^i} dx^i$. Now for $\alpha \in \Omega^k_M(U)$ we have $\alpha = \sum a_{\vec{I}} dx^{\vec{I}}$ we define $d_\mathtt{x}\alpha = \sum d_\mathtt{x} a_{\vec{I}} \wedge dx^{\vec{I}}$. To show the graded commutativity consider $\alpha = \sum a_{\vec{I}} dx^{\vec{I}} \in \Omega^k_M(U)$ and $\beta = \sum b_{\vec{J}} dx^{\vec{J}} \in \Omega^l_M(U)$. Then

$$\begin{aligned} d_\mathtt{x}(\alpha \wedge \beta) &= d_\mathtt{x}\left(\sum a_{\vec{I}} dx^{\vec{I}} \wedge \sum b_{\vec{J}} dx^{\vec{J}}\right) \\ &= d_\mathtt{x}\left(\sum a_{\vec{I}} b_{\vec{J}} dx^{\vec{I}} \wedge dx^{\vec{J}}\right) \\ &= \sum \left(\left(da_{\vec{I}}\right) b_{\vec{J}} + a_{\vec{I}}\left(db_{\vec{J}}\right)\right) dx^{\vec{I}} \wedge dx^{\vec{J}} \\ &= \left(\sum_{\vec{I}} da_{\vec{I}} \wedge dx^{\vec{I}}\right) \wedge \sum_{\vec{J}} b_{\vec{J}} dx^{\vec{J}} \\ &\quad + \sum_{\vec{I}} a_{\vec{I}} dx^{\vec{I}} \wedge \left((-1)^k \sum_{\vec{J}} db_{\vec{J}} \wedge dx^{\vec{J}}\right) \end{aligned}$$

since $db_{\vec{J}} \wedge dx^{\vec{I}} = (-1)^k dx^{\vec{I}} \wedge db_{\vec{J}}$ due to the k interchanges of the basic differentials dx^i. This means that the commutation rule holds at least in local coordinates. Also, easily verified in local coordinates is that for any function f we have $d_{\mathtt{x}} d_{\mathtt{x}} f = d_{\mathtt{x}} df = \sum_{ij} \left(\frac{\partial^2 f}{\partial x^i \partial x^j} \right) dx^i \wedge dx^j = 0$ since $\frac{\partial^2 f}{\partial x^i \partial x^j}$ is symmetric in ij and $dx^i \wedge dx^j$ is antisymmetric in i, j. More generally, for any functions f, g we have $d_{\mathtt{x}}(df \wedge dg) = 0$ because of the graded commutativity. Inductively we get $d_{\mathtt{x}}(df_1 \wedge df_2 \wedge \cdots \wedge df_k) = 0$ for any functions $f_i \in C^\infty(U)$. From this it easily follows that for any $\alpha = \sum a_{\vec{I}} dx^{\vec{I}} \in \Omega_M^k(U)$ we have $d_{\mathtt{x}} d_{\mathtt{x}} \sum a_{\vec{I}} dx^{\vec{I}} = d_{\mathtt{x}} \sum d_{\mathtt{x}} a_{\vec{I}} \wedge dx^{\vec{I}} = 0$. We have now defined an operator $d_{\mathtt{x}}$ for each coordinate chart $U, \mathtt{x}$, that clearly has the desired properties on that chart. Consider to different charts $U, \mathtt{x}$ and $V, \mathtt{y}$ such that $U \cap V \neq \emptyset$. We need to show that $d_{\mathtt{x}}$ restricted to $U \cap V$ coincides with $d_{\mathtt{y}}$ restricted to $U \cap V$ but is clear that $d_{\mathtt{x}}$ and $d_{\mathtt{y}}$ satisfy the hypothesis of Lemma 8.2 and so they must agree on $U \cap V$.

It is now clear that the individual operators on coordinate charts fit together to give a well defined operator with the desired properties. ■

The degree 1 graded derivation just introduced is called the **exterior derivative**. Another approach to the existence of the exterior derivative is to exhibit a global coordinate free formula. One can then verify that the global formula reproduces the correct local formula. This approach is given in G.2. The formula is as follows: If $\omega \in \Omega^k(U)$ and $X_0, X_1, ..., X_k \in \mathfrak{X}_M(U)$ then

$$\begin{aligned} d\omega(X_0, X_1, ..., X_k) = &\sum_{0 \le i \le k} (-1)^i X_i(\omega(X_0, ..., \widehat{X_i}, ..., X_k)) \\ &+ \sum_{0 \le i < j \le k} (-1)^{i+j} \omega([X_i, X_j], X_0, ..., \widehat{X_i}, ..., \widehat{X_j}, ..., X_k). \end{aligned}$$

Lemma 8.3 *Given any smooth map $f : M \to N$ we have that d is natural with respect to the pull-back:*

$$f^*(d\eta) = d(f^*\eta)$$

Definition 8.15 *A smooth differential form α is called **closed** if $d\alpha = 0$ and **exact** if $\alpha = d\beta$ for some differential form β.*

Corollary 8.4 *Every exact form is closed.*

The converse is not true in general. The extent to which it fails is a topological property of the manifold. This is the point of the De Rham cohomology to be studied in detail in chapter **??**. Here we just give the following basic definition:

Definition 8.16 *Since the exterior derivative operator is a graded map of degree one with $d^2 = 0$ we have, for each i, the **de Rham cohomology group** (actually vector spaces) given by*

$$H^i(M) = \frac{ker(d : \Omega^i(M) \to \Omega^{i+1}(M))}{Im(d : \Omega^{i-1}(M) \to \Omega^i(M))}. \tag{8.5}$$

In other words, we look at closed forms and identify any two whose difference is an exact form.

8.6.6 Vector Valued and Algebra Valued Forms.

For $\alpha \in L^k_{skew}$ (V; W) and $\beta \in L^k_{skew}$ (V; W) we define the wedge product using the same formula as before except that we use the tensor product so that $\alpha \wedge \beta \in L^k_{skew}$ (V; W $\otimes$ W) :

$$\begin{aligned}&(\omega \wedge \eta)(v_1, v_2, .., v_r, v_{r+1}, v_{r+2}, .., v_{r+s})\\&= \sum_{r,s-\text{shuffles } \sigma} \operatorname{sgn}(\sigma)\omega(v_{\sigma_1}, v_{\sigma_2}, .., v_{\sigma_r}) \otimes \eta(v_{\sigma_{r+1}}, v_{\sigma_{r+2}}, .., v_{\sigma_{r+s}})\end{aligned}$$

Globalizing this algebra as usual we get a vector bundle W $\otimes (\bigwedge^k T^*M)$ that in turn gives rise to a space of sections $\Omega^k(M, \mathrm{W})$ (and a presheaf $U \mapsto \Omega^k(U, \mathrm{W})$) and exterior product $\Omega^k(U, \mathrm{W}) \times \Omega^l(U, \mathrm{W}) \to \Omega^{k+l}(U, \mathrm{W} \otimes \mathrm{W})$. The space $\Omega^k(U, \mathrm{W})$ is a module over $\mathcal{C}^\infty_M(U)$. We still have a pullback operation defined as before and also a natural exterior derivative

$$d : \Omega^k(U, \mathrm{W}) \to \Omega^{k+1}(U, \mathrm{W})$$

defined by the formula

$$\begin{aligned}&d\omega(X_0, ..., X_k)\\&= \sum_{1 \le i \le k} (-1)^i X_i(\omega(X_0, ..., \widehat{X_i}, ..., X_k))\\&+ \sum_{1 \le i < j \le k} (-1)^{i+j} \omega(X_0, ..., [X_i, X_j], ..., X_k)\end{aligned}$$

where now $\omega(X_0, ..., \widehat{X_i}, ..., X_k)$ is a W-valued function so we take

$$\begin{aligned}&X_i(\omega(X_0, ..., \widehat{X_i}, ..., X_k))(p)\\&= D\omega(X_0, ..., \widehat{X_i}, ..., X_k)\Big| \cdot X_i(p)\\&= d(\omega(X_0, ..., \widehat{X_i}, ..., X_k))(X_i(p))\end{aligned}$$

which is an element of W under the usual identification of W with any of its tangent spaces.

To give a local formula valid for finite dimensional M, we let $f_1, ..., f_n$ be a basis of W and $(x^1, ..., x^n)$ local coordinates defined on U. For $\omega = \sum a^j_{\vec{I}}, f_j \otimes dx^{\vec{I}}$ we have

$$\begin{aligned}d\omega &= d(f_j \otimes \sum a_{\vec{I},j} dx^{\vec{I}})\\&= \sum (f_j \otimes da_{\vec{I},j} \wedge dx^{\vec{I}}).\end{aligned}$$

The elements $f_j \otimes dx_p^{\vec{I}}$ form a basis for the vector space $\mathrm{W} \otimes (\bigwedge^k T_p^* M)$ for every $p \in U$.

Now if W happens to be an algebra then the algebra product $\mathrm{W} \times \mathrm{W} \to \mathrm{W}$ is bilinear and so gives rise to a linear map $m{:}\mathrm{W} \otimes \mathrm{W} \to \mathrm{W}$. We compose the exterior product with this map to get a wedge product $\overset{m}{\wedge} : \Omega^k(U, \mathrm{W}) \times \Omega^l(U, \mathrm{W}) \to \Omega^{k+l}(U, \mathrm{W})$

$$
\begin{aligned}
&(\omega \overset{m}{\wedge} \eta)(v_1, v_2, .., v_r, v_{r+1}, v_{r+2}, .., v_{r+s}) \\
&= \sum_{r,s-\text{shuffles } \sigma} \operatorname{sgn}(\sigma) m \left(\omega(v_{\sigma_1}, v_{\sigma_2}, .., v_{\sigma_r}) \otimes \eta(v_{\sigma_{r+1}}, v_{\sigma_{r+2}}, .., v_{\sigma_{r+s}})\right) \\
&= \sum_{r,s-\text{shuffles } \sigma} \operatorname{sgn}(\sigma) \omega(v_{\sigma_1}, v_{\sigma_2}, .., v_{\sigma_r}) \cdot \eta(v_{\sigma_{r+1}}, v_{\sigma_{r+2}}, .., v_{\sigma_{r+s}})
\end{aligned}
$$

A particularly important case is when W is a Lie algebra $\mathfrak{g}$ with bracket $[.,.]$. Then we write the resulting product $\overset{m}{\wedge}$ as $[.,.]_\wedge$ or just $[.,.]$ when there is no risk of confusion. Thus if $\omega, \eta \in \Omega^1(U, \mathfrak{g})$ are Lie algebra valued 1-forms then

$$[\omega, \eta](X) = [\omega(X), \eta(Y)] + [\eta(X), \omega(Y)].$$

In particular, $\frac{1}{2}[\omega, \omega](X, Y) = [\omega(X), \omega(Y)]$ which might not be zero in general!

8.6.7 Vector Bundle Valued Forms.

It will occur in several contexts to have on hand the notion of a differential form with values in a vector bundle.

Definition 8.17 *Let $\xi = (\mathbb{F}^k \hookrightarrow E \to M)$ be a smooth vector bundle. A differential $p-$form with values in ξ (or values in E) is a smooth section of the bundle $E \otimes \wedge^p T^* M$. These are denoted by $\Omega^p(M; E)$.*

Remark 8.5 *The reader should avoid confusion between $\Omega^p(M; E)$ and the space of sections $\Gamma(M, \wedge^p E)$.*

In order to get a grip on the meaning of the bundle let exhibit transition functions. For a vector bundle, knowing the transition functions is tantamount to knowing how local expressions with respect to a frame transform as we change frame. A frame for $E \otimes \wedge^p T^* M$ is given by combining a local frame for E with a local frame for $\wedge^p TM$. Of course we must choose a common refinement of the VB-charts in order to do this but this is obviously no problem. Let $(e_1, ..., e_k)$ be a frame field defined on an open set U. We may as well take U to also be a chart domain for the manifold M. Then any local section of $\Omega^p(\xi)$ defined on U has the form

$$\sigma = \sum a_{\vec{I}}^j e_j \otimes dx^I$$

for some smooth functions $a_{\vec{I}}^j = a_{i_1 \ldots i_p}^j$ defined in U. Then for a new local set up with frames $(f_1, ..., f_k)$ and $dy^{\vec{I}} = dy^{i_1} \wedge \cdots \wedge dy^{i_p}$ $(i_1 < ... < i_p)$ then

$$\sigma = \sum \acute{a}_{\vec{I}}^j f_j \otimes dy^{\vec{I}}$$

we get the transformation law

$$\acute{a}^{j}_{\bar{I}} = a^{i}_{\bar{J}} C^{j}_{i} \frac{\partial x^{\bar{J}}}{\partial y^{\bar{I}}}$$

and where C^j_i is defined by $f_s C^s_j = e_j$.

Exercise 8.19 *Derive the above transformation law.*

Solution 8.1 $\sum a^j_{\bar{I}} e_j \otimes dx^I = \sum a^j_{\bar{I}} f_s C^s_j \otimes \frac{\partial x^{\bar{J}}}{\partial y^{\bar{I}}} dy^I$ *etc.*

A more elegant way of describing the transition functions is just to recall that anytime we have two vector bundles over the same base space and respective typical fibers V and W then the respective transition functions $g_{\alpha\beta}$ and $h_{\alpha\beta}$ (on a common cover) combine to give $g_{\alpha\beta} \otimes h_{\alpha\beta}$ where for a given $x \in U_{\alpha\beta}$

$$g_{\alpha\beta}(x) \otimes h_{\alpha\beta}(x) : \mathrm{V} \otimes \mathrm{W} \to \mathrm{V} \otimes \mathrm{W}$$
$$g_{\alpha\beta}(x) \otimes h_{\alpha\beta}(x).(v, w) = g_{\alpha\beta}(x)v \otimes h_{\alpha\beta}(x)w.$$

At any rate, these transformation laws fade into the background since if all our expressions are manifestly invariant (or invariantly defined in the first place) then we don't have to bring them up. A more important thing to do is to get used to calculating.

Now we want to define an important graded module structure on $\Omega(M; E) = \sum_{p=0}^{n} \Omega^p(M; E)$. This will be a module over the graded algebra $\Omega(M)$. The action of $\Omega(M)$ on $\Omega(M; E)$ is given by maps $\overset{\wedge}{\otimes} : \Omega^k(M) \times \Omega^l(M; E) \to \Omega^{k+l}(M; E)$ which in turn are defined by extending the following rule linearly:

$$\mu^1 \overset{\wedge}{\otimes} (\sigma \otimes \mu^2) := \sigma \otimes \mu^1 \wedge \mu^2$$

If the vector bundle is actually an algebra bundle then (naming the bundle $\mathcal{A} \to M$ now for "algebra") we may turn $\mathcal{A} \otimes \wedge T^*M := \sum_{p=0}^{n} \mathcal{A} \otimes \wedge^p T^*M$ into an algebra bundle by defining

$$(v_1 \otimes \mu^1) \wedge (v_2 \otimes \mu^2) := v_1 v_2 \otimes \mu^1 \wedge \mu^2$$

and then extending linearly:

$$(a^i_j v_i \otimes \mu^j) \wedge (b^k_l v_k \otimes \mu^l) := v_i v_j \otimes \mu^j \wedge \mu^l$$

From this the sections $\Omega(M, \mathcal{A}) = \Gamma(M, \mathcal{A} \otimes \wedge T^*M)$ become an algebra over the ring of smooth functions. For us the most important example is where $\mathcal{A} = \mathrm{End}(E)$. Locally, say on U, sections σ_1 and σ_2 of $\Omega(M, \mathrm{End}(E))$ take the form $\sigma_1 = A_i \otimes \alpha^i$ and $\sigma_2 = B_i \otimes \beta^i$ where A_i and B_i are maps $U \to \mathrm{End}(E)$. Thus for each $x \in U$, the A_i and B_i evaluate to give $A_i(x), B_i(x) \in \mathrm{End}(E_x)$. The multiplication is then

$$(A_i \otimes \alpha^i) \wedge (B_j \otimes \beta^j) = A_i B_j \otimes \alpha^i \wedge \beta^j$$

where the $A_i B_j : U \to \mathrm{End}(E)$ are local sections given by composition:

$$A_i B_j : x \mapsto A_i(x) \circ B_j(x).$$

Exercise 8.20 *Show that $\Omega(M, \mathrm{End}(E))$ acts on $\Omega(M, E)$ making $\Omega(M, E)$ a bundle of modules over the bundle of algebras $\Omega(M, \mathrm{End}(E))$.*

If this seems all to abstract to the newcomer perhaps it would help to think of things this way: We have a cover of a manifold M by open sets $\{U_\alpha\}$ that simultaneously trivialize both E and TM. Then these give also trivializations on these open sets of the bundles $\mathrm{Hom}(E, E)$ and $\wedge TM$. Associated with each is a frame field for $E \to M$ say $(e_1, ..., e_k)$ which allows us to associate with each section $\sigma \in \Omega^p(M, E)$ a $k-$tuple of $p-$forms $\sigma_U = (\sigma_U^i)$ for each U. Similarly, a section $A \in \Omega^q(M, \mathrm{End}(E))$ is equivalent to assigning to each open set $U \in \{U_\alpha\}$ a matrix of $q-$forms A_U. The algebra structure on $\Omega(M, \mathrm{End}(E))$ is then just matrix multiplication were the entries are multiplied using the wedge product $A_U \wedge B_U$ where

$$(A_U \wedge B_U)^i_j = A^i_k \wedge B^k_j$$

The module structure is given locally by $\sigma_U \mapsto A_U \wedge \sigma_U$. Where did the bundle go? The global topology is now encoded in the transformation laws which tell us what the same section looks like when we change to a new frame field on an overlap $U_\alpha \cap U_\beta$. In this sense the bundle is a combinatorial recipe for pasting together local objects.

8.7 Lie derivative, interior product and exterior derivative.

The Lie derivative acts on differential forms since the latter are, from one viewpoint, tensors. When we apply the Lie derivative to a differential form we get a differential form so we should think about the Lie derivative in the context of differential forms.

Lemma 8.4 *For any $X \in \mathfrak{X}(M)$ and any $f \in \Omega^0(M)$ we have $\mathcal{L}_X df = d\mathcal{L}_X f$.*

Proof. For a function f we compute as

$$\begin{aligned}
&(\mathcal{L}_X df)(Y) \\
&= (\frac{d}{dt}(Fl_t^X)^* df)(Y) = \frac{d}{dt} df(TFl_t^X \cdot Y) \\
&= \frac{d}{dt} Y((Fl_t^X)^* f) = Y(\frac{d}{dt}(Fl_t^X)^* f) \\
&= Y(\mathcal{L}_X f) = d(\mathcal{L}_X f)(Y)
\end{aligned}$$

where $Y \in \mathfrak{X}(M)$ is arbitrary. ∎

Exercise 8.21 *Show that $\mathcal{L}_X(\alpha \wedge \beta) = \mathcal{L}_X \alpha \wedge \beta + \alpha \wedge \mathcal{L}_X \beta$.*

We now have two ways to differentiate sections in $\Omega(M)$. Once again we write Ω_M instead of $\Omega(U)$ or $\Omega(M)$ since every thing works equally well in either case. In other words we are thinking of the presheaf $\Omega_M : U \mapsto \Omega(U)$. First, there is the Lie derivative which turns out to be a graded derivation of degree zero;

$$\mathcal{L}_X : \Omega^i_M \to \Omega^i_M. \tag{8.6}$$

Second, there is the exterior derivative d which is a graded derivation of degree 1. In order to relate the two operations we need a third map which, like the Lie derivative, is taken with respect to a given field $X \in \Gamma(U; TM)$. This map is a degree -1 graded derivation and is defined by

$$\iota_X \omega(X_1, \ldots, X_{i-1}) = \omega(X, X_1, \ldots, X_{i-1}) \tag{8.7}$$

where we view $\omega \in \Omega^i_M$ as a skew-symmetric multi-linear map from $\mathfrak{X}_M \times \cdots \times \mathfrak{X}_M$ to $\mathcal{C}^\infty_M$. We could also define ι_X as that unique operator that satisfies

$$\iota_X \theta = \theta(X) \text{ for } \theta \in \Omega^1_M \text{ and } X \in \mathfrak{X}_M$$

$$\iota_X(\alpha \wedge \beta) = (\iota_X \alpha) \wedge \beta + (-1)^k \wedge \alpha \wedge (\iota_X \beta) \text{ for } \alpha \in \Omega^k_M.$$

In other word, ι_X is the graded derivation of Ω_M of degree -1 determined by the above formulas.

In any case, we will call this operator the **interior product** or **contraction operator**.

Notation 8.1 *Other notations for $\iota_X \omega$ include $X \lrcorner \omega = \langle X, \omega \rangle$. These notations make the following theorem look more natural:*

Theorem 8.3 *The Lie derivative is a derivation with respect to the pairing $\langle X, \omega \rangle$. That is*

$$\mathcal{L}_X \langle X, \omega \rangle = \langle \mathcal{L}_X X, \omega \rangle + \langle X, \mathcal{L}_X \omega \rangle$$

or

$$\mathcal{L}_X (X \lrcorner \omega) = (\mathcal{L}_X X) \lrcorner \omega + X \lrcorner (\mathcal{L}_X \omega)$$

Using the "ι_X" notation: $\mathcal{L}_X(\iota_X \omega) = \iota_{\mathcal{L}_X X} \omega + \iota_X \mathcal{L}_X \omega$ (not as pretty).

Proof. Exercise. ■

Now we can relate the Lie derivative, the exterior derivative and the contraction operator.

Theorem 8.4 *Let $X \in \mathfrak{X}_M$. Then we have Cartan's homotopy formula;*

$$\mathcal{L}_X = d \circ \iota_X + \iota_X \circ d \tag{8.8}$$

Proof. One can check that both sides define derivations and so we just have to check that they agree on functions and exact 1-forms. On functions we have $\iota_X f = 0$ and $\iota_X df = Xf = \mathcal{L}_X f$ so the formula holds. On differentials of functions we have

$$(d \circ \iota_X + \iota_X \circ d)df = (d \circ \iota_X)df = d\mathcal{L}_X f = \mathcal{L}_X df$$

where we have used lemma 8.4 in the last step. ■

As a corollary can now extend lemma 8.4:

Corollary 8.5 $d \circ \mathcal{L}_X = \mathcal{L}_X \circ d$

Proof.

$$\begin{aligned} d\mathcal{L}_X \alpha &= d(d\iota_X + \iota_X d)(\alpha) \\ &= d\iota_X d\alpha = d\iota_X d\alpha + \iota_X dd\alpha = \mathcal{L}_X \circ d \end{aligned}$$

■

Corollary 8.6 *We have the following formulas:*

1) $\iota_{[X,Y]} = \mathcal{L}_X \circ \iota_Y + \iota_Y \circ \mathcal{L}_X$

2) $\mathcal{L}_{fX}\omega = f\mathcal{L}_X\omega + df \wedge \iota_X\omega$ *for all* $\omega \in \Omega(M)$.

Proof.

Corollary 8.7 *Exercise.*

■

8.8 Orientation

A rank n vector bundle $E \to M$ is called oriented if every fiber E_p is given a smooth choice of orientation. There are several equivalent ways to make a rigorous definition:

1. A vector bundle is orientable if and only if it has an atlas of bundle charts (local trivializations) such that the corresponding transition maps take values in $\mathrm{GL}^+(n, \mathbb{R})$ the group of positive determinant matrices. If the vector bundle is orientable then this divides the set of all bundle charts into two classes. Two bundle charts are in the same orientation class the transition map takes values in $\mathrm{GL}^+(n, \mathbb{R})$. If the bundle is not orientable there is only one class.

2. If there is a smooth global section s on the bundle $\bigwedge^n E \to M$ then we say that this determines an orientation on E. A frame $(f_1, ..., f_n)$ of fiber E_p is positively oriented with respect to s if and only if $f_1 \wedge ... \wedge f_n = as(p)$ for a positive real number $a > 0$.

3. If there is a smooth global section ω on the bundle $\bigwedge^n E^* \cong L^k_{alt}(E) \to M$ then we say that this determines an orientation on E. A frame $(f_1, ..., f_n)$ of fiber E_p is positively oriented with respect to ω if and only if $\omega(p)(f_1, ..., f_n) > 0$.

Exercise 8.22 *Show that each of these three approaches is equivalent.*

Now let M be an n-dimensional manifold. Let U be some open subset of M which may be all of M. Consider a **top form**, i.e. an n-form $\varpi \in \Omega^n_M(U)$ where $n = \dim(M)$ and assume that ϖ is never zero on U. In this case we will say that ϖ is nonzero or that ϖ is a **volume form**. Every other top form μ is of the form $\mu = f\varpi$ for some smooth function f. This latter fact follows easily from $\dim(\bigwedge^n T_pM) = 1$ for all p. If $\varphi : U \to U$ is a diffeomorphism then we must have that $\varphi^*\varpi = \delta\varpi$ for some $\delta \in C^\infty(U)$ that we will call the **Jacobian determinant** of φ with respect to the volume element ϖ:

$$\varphi^*\varpi = J_\varpi(\varphi)\varpi$$

Proposition 8.8 *The sign of $J_\varpi(\varphi)$ is independent of the choice of volume form ϖ.*

Proof. Let $\varpi' \in \Omega^n_M(U)$. We have

$$\varpi = a\varpi'$$

for some function a that is never zero on U. We have

$$\begin{aligned} J(\varphi)\varpi &= (\varphi^*\varpi) = (a \circ \varphi)(\varphi^*\varpi') \\ &= (a \circ \varphi)J_{\varpi'}(\varphi)\varpi' = \frac{a \circ \varphi}{a}\varpi \end{aligned}$$

and since $\frac{a\circ\varphi}{a} > 0$ and ϖ is nonzero the conclusion follows. ∎

Let us consider a very important special case of this: Suppose that $\varphi : U \to U$ is a diffeomorphism and $U \subset \mathbb{R}^n$. Then letting $\varpi_0 = du^1 \wedge \cdots \wedge du^n$ we have

$$\begin{aligned} \varphi^*\varpi_0(x) &= \varphi^*du^1 \wedge \cdots \wedge \varphi^*du^n(x) \\ &= \left(\sum \left.\frac{\partial(u^1 \circ \varphi)}{\partial u^{i_1}}\right|_x du^{i_1} \right) \wedge \cdots \wedge \left(\sum \left.\frac{\partial(u^n \circ \varphi)}{\partial u^{i_n}}\right|_x du^{i_n} \right) \\ &= \det\left(\frac{\partial(u^i \circ \varphi)}{\partial u^j}(x) \right) = J\varphi(x). \end{aligned}$$

so in this case $J_{\varpi_0}(\varphi)$ is just the usual Jacobian determinant of φ.

Definition 8.18 *A diffeomorphism $\varphi : U \to U \subset \mathbb{R}^n$ is said to be positive or orientation preserving if* $\det(T\varphi) > 0$.

More generally, let a nonzero top form ϖ be defined on $U \subset M$ and let ϖ' be another defined on $U' \subset N$. Then we say that a diffeomorphism $\varphi : U \to U'$ is orientation preserving (or positive) with respect to the pair ϖ, ϖ' if the unique function $J_{\varpi,\varpi'}$ such that $\varphi^*\varpi' = J_{\varpi,\varpi'}\varpi$ is strictly positive on U.

Definition 8.19 *A differentiable manifold M is said to be* ***orientable*** *if and only if there is an atlas of admissible charts such that for any pair of charts ψ_α, U_α and ψ_β, U_β from the atlas with $U_\alpha \cap U_\beta \neq \emptyset$, the transition map $\psi_\beta \circ \psi_\alpha^{-1}$ is orientation preserving. Such an atlas is called an orienting atlas.*

Exercise 8.23 *The tangent bundle is a vector bundle. Show that this last definition agrees with our definition of an orientable vector bundle in that M is an orientable manifold in the current sense if and only if TM is an orientable vector bundle.*

Let $\mathcal{A}_M$ be the maximal atlas for an orientable differentiable manifold M. Then there are two subatlases $\mathcal{A}$ and $\mathcal{A}'$ with $\mathcal{A} \cup \mathcal{A}' = \mathcal{A}_M$, $\mathcal{A} \cap \mathcal{A}' = \emptyset$ and such that the transition maps for charts from $\mathcal{A}$ are all positive and similarly the transition maps of $\mathcal{A}'$ are all positive.. Furthermore if $\psi_\alpha, U_\alpha \in \mathcal{A}$ and $\psi_\beta, U_\beta \in \mathcal{A}'$ then $\psi_\beta \circ \psi_\alpha^{-1}$ is negative (orientation reversing). A choice of one these two atlases is called an **orientation on** M. Every orienting atlas is a subatlas of exactly one of $\mathcal{A}$ or $\mathcal{A}'$. If such a choice is made then we say that M is **oriented**. Alternatively, we can use the following proposition to specify an orientation on M:

Proposition 8.9 *Let $\varpi \in \Omega^n(M)$ be a volume form on M, i.e. ϖ is a nonzero top form. Then ϖ determines an orientation by determining an (orienting) atlas $\mathcal{A}$ by the rule*

$$\psi_\alpha, U_\alpha \in \mathcal{A} \Longleftrightarrow \psi_\alpha \text{ is orientation preserving resp. } \varpi, \varpi_0$$

where ϖ_0 is the standard volume form on $\mathbb{R}^n$ introduced above.

Exercise 8.24 *Prove the last proposition and then prove that we can use an orienting atlas to construct a volume form on an orientable manifold that gives the same orientation as the orienting atlas.*

We now construct a two fold covering manifold $\mathrm{Or}(M)$ for any finite dimensional manifold called the orientation cover. The orientation cover will itself always be orientable. Consider the vector bundle $\bigwedge^n T^*M$ and remove the zero section to obtain

$$(\bigwedge^n T^*M)^\times := \bigwedge^n T^*M - \{\text{zero section}\}$$

Define an equivalence relation on $(\bigwedge^n T^*M)^\times$ by declaring $\nu_1 \sim \nu_2$ if and only if ν_1 and ν_2 are in the same fiber and if $\nu_1 = a\nu_2$ with $a > 0$. The space of equivalence classes is denoted $\mathrm{Or}(M)$. There is a unique map π_{Or} making the following diagram commute:

$$\begin{array}{ccc}
\mathrm{Or}(M) & \to & (\bigwedge^n T^*M)^\times \\
 & & \downarrow \\
\pi_{Or} & \searrow & \\
 & & M
\end{array}$$

$\mathrm{Or}(M) \to M$ is a two fold covering space with the quotient topology and in fact is a differentiable manifold.

8.8.1 Orientation of manifolds with boundary

Recall that a **half space chart** ψ_α for a manifold with boundary M is a bijection (actually diffeomorphism) of an open subset U_α of M onto an open subset of M_λ^-. A C^r **half space atlas** is a collection ψ_α, U_α of such charts such that for any two; ψ_α, U_α and ψ_β, U_β, the map $\psi_\alpha \circ \psi_\beta^{-1}$ is a C^r diffeomorphism on its natural domain (if non-empty). Note: "Diffeomorphism" means in the extended sense of a being homeomorphism such that both $\psi_\alpha \circ \psi_\beta^{-1} :: \mathsf{M}_\lambda^- \to \mathsf{M}$ and its inverse are C^r in the sense of definition 3.5.

Let us consider the case of finite dimensional manifolds. Then letting $\mathsf{M} = \mathbb{R}^n$ and $\lambda = u^1 : \mathbb{R}^n \to \mathbb{R}$ we have the half space $\mathsf{M}_\lambda^- = \mathbb{R}^n_{u^1 \le 0}$. The funny choice of sign is to make $\mathsf{M}_\lambda^- = \mathbb{R}^n_{u^1 \le 0}$ rather than $\mathbb{R}^n_{u^1 \ge 0}$. The reason we do this is to be able to get the right induced orientation on ∂M without introducing a minus sign into our Stoke's formula proved below. The reader may wish to re-read remark 3.6 at this time.

Now, imitating our previous definition we define an oriented (or orienting) atlas for a finite dimensional manifold with boundary to be an atlas of half-space charts such that the overlap maps $\psi_\alpha \circ \psi_\beta^{-1} :: \mathbb{R}^n_{u^1 \le 0} \to \mathbb{R}^n_{u^1 \le 0}$ are orientation preserving. A manifold with boundary together a choice of (maximal) oriented atlas is called an **oriented manifold with boundary**. If there exists an orienting atlas for M then we say that M is **orientable** just as the case of a manifold without boundary.

Now if $\mathcal{A} = \{(\psi_\alpha, U_\alpha)\}_{\alpha \in A}$ is an orienting atlas for M as above with domains in $\mathbb{R}^n_{u^1 \le 0}$ then the induced atlas $\{(\psi_\alpha|_{U_\alpha \cap \partial M}, U_\alpha \cap \partial M)\}_{\alpha \in A}$ is an orienting atlas for the manifold ∂M and the resulting choice of orientation is called the induced orientation on ∂M. If M is oriented we will always assume that ∂M is given this induced orientation.

Definition 8.20 *A basis $f_1, f_2, ..., f_n$ for the tangent space at a point p on an oriented manifold (with or without boundary) is called positive if whenever $\psi_\alpha = (x^1, ..., x^n)$ is an oriented chart on a neighborhood of p then $(dx^1 \wedge ... \wedge dx^n)(f_1, f_2, ..., f_n) > 0$.*

Recall that vector v in T_pM for a point p on the boundary ∂M is called outward pointing if $T_p\psi_\alpha \cdot v \in \mathsf{M}_\lambda^-$ is outward pointing in the sense that $\lambda(T_p\psi_\alpha \cdot v) < 0$.

Since we have chosen $\lambda = u^1$ and hence $\mathsf{M}_\lambda^- = \mathbb{R}^n_{u^1 \le 0}$ for our definition in choosing the orientation on the boundary we have that in this case v is outward pointing if and only if $T_p\psi_\alpha \cdot v \in \mathbb{R}^n_{u^1 \le 0}$.

Definition 8.21 *A **nice chart** on a smooth manifold (possibly with boundary) is a chart ψ_α, U_α where ψ_α is a diffeomorphism onto $\mathbb{R}^n_{u^1 \le 0}$ if $U_\alpha \cap \partial M \neq \emptyset$ and a diffeomorphism onto the interior $\mathbb{R}^n_{u^1 < 0}$ if $U_\alpha \cap \partial M = \emptyset$.*

Lemma 8.5 *Every (oriented) smooth manifold has an (oriented) atlas consisting of nice charts.*

Proof. If ψ_α, U_α is an oriented chart with range in the interior of the left half space $\mathbb{R}^n_{u^1\leq 0}$ then we can find a ball B inside $\psi_\alpha(U_\alpha)$ in $\mathbb{R}^n_{u^1<0}$ and then we form a new chart on $\psi_\alpha^{-1}(B)$ with range B. But a ball is diffeomorphic to $\mathbb{R}^n_{u^1<0}$. So composing with such a diffeomorphism we obtain the nice chart. If ψ_α, U_α is an oriented chart with range meeting the boundary of the left half space $\mathbb{R}^n_{u^1\leq 0}$ then we can find a half ball B_- in $\mathbb{R}^n_{u^1\leq 0}$ with center on $\mathbb{R}^n_{u^1=0}$. Reduce the chart domain as before to have range equal to this half ball. But every half ball is diffeomorphic to the half space $\mathbb{R}^n_{u^1\leq 0}$ so we can proceed by composition as before. ■

Exercise 8.25 *If $\psi_\alpha = (x^1, ..., x^n)$ is an oriented chart on a neighborhood of p on the boundary of an oriented manifold with boundary then the vectors $\frac{\partial}{\partial x^2}, ..., \frac{\partial}{\partial x^n}$ form a positive basis for $T_p\partial M$ with respect to the induced orientation on ∂M. More generally, if f_1 is outward pointing and $f_1, f_2, ..., f_n$ is positive on M at p, then $f_2, ..., f_n$ will be positive for ∂M at p.*

Exercise 8.26 *Show that if M is simply connected then it must be orientable.*

Chapter 9

Integration and Stokes' Theorem

In this chapter we explore a fundamental aspect of differential forms that has not been mentioned yet. Namely, a differential $k-$form "wants" to be integrated over a k dimensional submanifold. The submanifold should be parameterized so as to allow us to perform the integration but the result doesn't depend on the parameterization except that a parameterization gives an orientation to the submanifold. The result of integrating a form over a submanifold does depend on the choice of orientation and so we should refine our statement above to say that a differential k-form "wants" to be integrated over an oriented submanifold. Actually, we will be able to make sense of integrating a $k-$form over certain smooth maps. Since a submanifold is given by an embedding the idea of integrating over a map is really a generalization of the case of integration over a submanifold.

9.1 Integration of Differential Forms.

First we will talk about integrating 1-forms over a parameterized curve. The idea is already familiar from the calculus of several variables and is none other than the familiar line integral. For example, consider the curve $\gamma : [0, \pi] \to \mathbb{R}^3$

given by $\gamma(t) := (\cos t, \sin t, t)$. If $\omega = xydx + dy + xzdz$ then the integral is

$$\begin{aligned}\int_\gamma \omega &:= \int xydx + dy + xzdz \\ &= \int_{[0,\pi]} \gamma^*\omega \text{ (this step is to be a definition)} \\ &\int_0^\pi (x(t)y(t)dx/dt + dydt + x(t)z(t)dz/dt)dt \\ &= \int_0^\pi (\cos t \sin t \sin t - \cos t + t\cos t)dt = -2\end{aligned}$$

Notice the analogy in the fact that a $1-$form at a point takes a vector (infinitesimal curve) and gives a number while the $1-$form globally takes a curve and via integration also yields a number.

Since we want to be precise about what we are doing and also since the curves we integrate over may have singularities we now put forth various definitions regarding curves.

The Let O be an open set in $\mathbb{R}$. A continuous map $c : O \to M$ is C^k if the $k-$th derivative $(\mathtt{x} \circ c)^{(k)}$ exist and is continuous on $c^{-1}(U)$ for every coordinate chart $\mathtt{x}, U$ such that $U \cap c(I) \neq \emptyset$. If I is a subset of $\mathbb{R}$ then a map $c : I \to M$ is said to be C^k if there exists a C^k extension $\widetilde{c} : O \to M$ for some open set O containing I. We are particularly interested in the case where I is an interval. This interval may be open, closed, half open etc. We also allow intervals for which one of the "end points" a or b is $\pm\infty$.

Definition 9.1 *Let $O \subset \mathbb{R}$ be open. A continuous map $c : O \to M$ is said to be piecewise C^k if there exists a discrete sequence of points $\{t_n\} \subset O$ with $t_i < t_{i+1}$ such that c restricted to each $(t_i, t_{i+1}) \cap O$ is C^k. If I is subset of $\mathbb{R}$, then a map $c : I \to M$ is said to be piecewise C^k if there exists a piecewise extension of c to some open set containing I.*

Definition 9.2 *A **parametric curve** in M is a piecewise differentiable map $c : I \to M$ where I is either an interval or finite union of intervals. If c is an interval then we say that c is a **connected parametric curve**.*

Definition 9.3 *A elementary parametric curve is a regular curve $c : I \to \mathbb{R}^n$ such that I is an open connected interval.*

Definition 9.4 *If $c : I_1 \to M$ and $b : I_2 \to M$ are curves then we say that b is a positive (resp. negative) reparametrization of c if there exists a bijection $h : I_2 \to I_1$ with $c \circ h = b$ such that h is smooth and $h'(t) > 0$ (resp. $h'(t) > 0$) for all $t \in I_2$.*

We distinguish between a C^k map $c : I \to M$ and its image (or trace) $c(I)$ as a subset of M. The geometric curve is the set $c(I)$ itself while the parameterized curve is the map in question. The set of all parameterizations of a curve fall into two classes according to whether they are positive reparametrizations of each other or not. A choice of one of these classes gives and orientation to the geometric curve. Once the orientation is fixed then we may integrate a $1-$form over this oriented geometric curve using any parameterization in the chosen class.

A 1-form on an interval $I = [a, b]$ may always be given as $f\,dt$ for some smooth function on I. The integral of $f\ dt$ over $[a, b]$ is just the usual Riemann integral $\int_{[a,b]} f(t)\,dt$. If α is a $1-$form on M and $\gamma : [a, b] \to M$ is a parameterized curve whose image happens to be contained in a coordinate chart U,then the line integral of α along γ is defined as

$$\int_\gamma \alpha := \int_{[a,b]} \gamma^* \alpha$$

If γ is continuous but merely piecewise smooth (C^1 is enough) then we just integrate α along each smooth piece and add the results:

$$\int_\gamma \alpha := \sum_i \int_{[t_i, t_{i+1}]} \gamma^* \alpha$$

Next we move to 2 dimensions. We start we another simple example from calculus. Let $\sigma(u, v) = (\sin u \cos v, \cos u \cos v, \cos u)$ for $(\phi, \theta) \in (0, \pi] \times (0, 2\pi)$. This gives a parameterization of the sphere S^2. There are a few places where our parameterization is not quite perfect but those are measure zero sets so it will not make an difference when we integrate. Now we need something to integrate; say $\omega := zdy \wedge dz + xdx \wedge dy$. The integration is done by analogy with what we did before. We pull the form back to the uv space and then integrate just as one normally would in calculus of several variables:

$$\begin{aligned}
\int_\sigma \omega &:= \int_{(0,\pi] \times (0,2\pi)} \sigma^* \omega \\
&= \int_{(0,\pi] \times (0,2\pi)} \left(z(u, v) \frac{dy \wedge dz}{du \wedge dv} + x(u, v) \frac{dx \wedge dy}{du \wedge dv} \right) du \wedge dv \\
&= \int_0^\theta \int_0^{2\pi} \big[\cos u (\sin^2 u \cos v - \cos u \sin v \sin u) \\
&+ \sin u \cos v \ (\cos^2 u \cos^2 v - \sin^2 u \sin v \cos v) \big] \, dudv
\end{aligned}$$

Rather than finishing the above calculation, let us apply a powerful theorem that we will prove below (Stokes' theorem). The reason this will work depends on the fact that our map σ gives a good enough parameterization of the sphere that we are justified in interpreting the integral $\int_\sigma \omega$ as being an integral over the sphere $\int_{S^2} \omega$. This is a special case of integrating an $n-$form over a compact n-dimensional manifold. We will get back to this shortly. As far as the integral above goes, the Stokes' theorem is really a reformulation of the Stoke's theorem from calculus of 3 variables. The theorem instructs us that if we take the exterior derivative of ω and integrate that over the ball which is the interior of the sphere we should get the same thing. It is important to realize that this only works because our map σ parameterizes the whole sphere and the sphere is the boundary of the ball. Now a quick calculation of gives $d\omega = 0$ and so no matter what the integral would be zero and so we get zero for the above complicated integral too! In summary, what Stokes' theorem gives us is

$$\int_{S^2} \omega \overset{Stokes}{=} \int_B d\omega = \int_B 0 = 0$$

Let us now make sense out the more general situation. A smooth $k-$form on an open subset $O \subset \mathbb{R}^k$ can always be written in the form

$$f \; dx^1 \wedge \cdots \wedge dx^k$$

for some smooth function. It is implicit that $\mathbb{R}^k$ is oriented by $dx^1 \wedge \cdots \wedge dx^k$. We then simply define $\int_S f \, dx^1 \wedge \cdots \wedge dx^k := \int_S f \, dx^1 \cdots dx^k$ where the integral is the Riemann integral and S is any reasonable set on which such an integration makes sense. We could also take the integral to be the Lebesgue integral and S a Borel set. So integration of a k-form over a set in Euclidean space of the same dimension k turns out to be just ordinary integration as long as we write our k form as a function times the basic "volume form" $dx^1 \wedge \cdots \wedge dx^k$.

Let O be an bounded open subset of $\mathbb{R}^k$. A parameterized patch in M is a smooth map $\phi : O \to M$. If α is a smooth $k-$form then we define

$$\int_\phi \alpha := \int_O \phi^* \alpha$$

Notice that what we have done is integrate over a map. However, it was important that the map had as its domain a domain in a Euclidean space of a dimension that matches the degree of the form. We will also integrate over sets (a point we already made) by parameterizing the set. The reason this will work is that the integral is independent of the parameterization of the set as long as we keep the orientation induced by the parameterization is kept the same.

A variation on the idea of integration over a map is the following:

Definition 9.5 *A smooth map $\sigma : S \to M$ is called a singular simplex if $S = \{x \in \mathbb{R}^k : x^i \geq 0$ and $\sum x^i \leq 1\}$. A smooth map $\sigma : S \to M$ is called a singular cube if $S = \{x \in \mathbb{R}^k : 0 \leq x^i \leq 1\}$. In either case, we define the integral of a $k-$form α as*

$$\int_\sigma \alpha := \int_S \phi^* \alpha$$

Now lets integrate over manifolds (rather than maps). Let M be a smooth n-manifold possibly with boundary ∂M and assume that M is oriented and that ∂M has the induced orientation. From our discussion on orientation of manifolds with boundary and by general principles it should be clear that we may assume that all the charts in our orienting atlas have range in the left half space $\mathbb{R}^n_{u^1 \leq 0}$. If $\partial M = \emptyset$ then the ranges will just be in the interior $\mathbb{R}^n_{u^1 < 0} \subset \mathbb{R}^n_{u^1 \leq 0}$.

Definition 9.6 *A the* ***support*** *of a differential form $\alpha \in \Omega(M)$ is the* closure *of the set $\{p \in M : \alpha(p) \neq 0\}$ and is denoted by* $\mathrm{supp}(\alpha)$. *The set of all k-forms $\alpha^{(k)}$ that have compact support contained in $U \subset M$ is denoted by $\Omega_c^k(U)$.*

Let us return to the case of a k-form $\alpha^{(k)}$ on an open subset U of $\mathbb{R}^k$. If $\alpha^{(k)}$ has compact support in U we may define the integral $\int_U \alpha^{(k)}$ by

$$\begin{aligned}\int_U \alpha^{(k)} &= \int_U a(u)du^1 \wedge \cdots \wedge du^k \\ &:= \int_U a(u) \left|du^1 \cdots du^k\right|\end{aligned}$$

where this latter integral is the Riemann (or Lebesgue) integral of $a(u)$. We have written $\left|du^1 \cdots du^k\right|$ instead of $du^1 \cdots du^k$ to emphasize that the order of the du^i does not matter as it does for $du^1 \wedge \cdots \wedge du^k$. Of course, we would get the wrong answer if the order of the $u's$ did not give the standard orientation on $\mathbb{R}^k$.

Now consider an oriented $n-$dimensional manifold M and let $\alpha \in \Omega_M^n$. If α has compact support inside U_α for some chart $\mathtt{x}_\alpha, U_\alpha$ compatible with the orientation then $\mathtt{x}_\alpha^{-1} : \mathtt{x}_\alpha(U_\alpha) \to U_\alpha$ and $(\mathtt{x}_\alpha^{-1})^*\alpha$ has compact support in $\mathtt{x}_\alpha(U_\alpha) \subset \mathbb{R}^n_{u^1 \leq 0}$. We define

$$\int \alpha := \int_{\mathtt{x}_\alpha(U_\alpha)} (\mathtt{x}_\alpha^{-1})^*\alpha.$$

The standard change of variables formula show that this definition is independent of the oriented chart chosen. Now if $\alpha \in \Omega^n(M)$ does not have support contained in some chart domain then we choose a locally finite cover of M by oriented charts $(\mathtt{x}_i, U_i)$ and a smooth partition of unity (ρ_i, U_i) , $\mathrm{supp}(\rho_i) \subset U_i$. Then we define

$$\int \alpha := \sum_i \int_{\mathtt{x}_i(U_i)} (\mathtt{x}_i^{-1})^*(\rho_i\alpha)$$

Proposition 9.1 *The above definition is independent of the choice of the charts $\mathtt{x}_i, U_i$ and smooth partition of unity ρ_i, U_i.*

Proof. Let $(\bar{\mathtt{x}}_i, V_i)$, and $\overline{\rho}_i$ be another such choice. Then we have

$$\begin{aligned}
\int \alpha &:= \sum_i \int_{\mathtt{x}_i(U_i)} (\mathtt{x}_i^{-1})^*(\rho_i \alpha) \\
&= \sum_i \int_{\mathtt{x}_i(U_i)} (\mathtt{x}_i^{-1})^*(\rho_i \sum_j \overline{\rho}_j \alpha) \\
&\sum_i \sum_j \int_{\mathtt{x}_i(U_i \cap U_j)} (\mathtt{x}_i^{-1})^*(\rho_i \overline{\rho}_j \alpha) \\
&= \sum_i \sum_j \int_{\bar{\mathtt{x}}_j(U_i \cap U_j)} (\bar{\mathtt{x}}_j^{-1})^*(\rho_i \overline{\rho}_j \alpha) \\
&= \sum_j \int_{\bar{\mathtt{x}}_i(U_i)} (\bar{\mathtt{x}}_j^{-1})^*(\overline{\rho}_j \alpha)
\end{aligned}$$

■

9.2 Stokes' Theorem

Let us start with a couple special cases .

Case 9.1 (1) *Let $\omega_j = f du^1 \wedge \cdots \wedge \widehat{du^j} \wedge \cdots \wedge du^n$ be a smooth $n-1$ form with compact support contained in the interior of $\mathbb{R}^n_{u^1 \le 0}$ where the hat symbol over the du^j means this j-th factor is omitted. All $n-1$ forms on $\mathbb{R}^n_{u^1 \le 0}$ are sums of forms of this type. Then we have*

$$\begin{aligned}
\int_{\mathbb{R}^n_{u^1 \le 0}} d\omega_j &= \int_{\mathbb{R}^n_{u^1 \le 0}} d(f du^1 \wedge \cdots \wedge \widehat{du^j} \wedge \cdots \wedge du^n) \\
&= \int_{\mathbb{R}^n_{u^1 \le 0}} (df \wedge du^1 \wedge \cdots \wedge \widehat{du^j} \wedge \cdots \wedge du^n) \\
&= \int_{\mathbb{R}^n_{u^1 \le 0}} (\sum_k \frac{\partial f}{\partial u^k} du^k \wedge du^1 \wedge \cdots \wedge \widehat{du^j} \wedge \cdots \wedge du^n) \\
&= \int_{\mathbb{R}^n_{u^1 \le 0}} (-1)^{j-1} \frac{\partial f}{\partial u^j} du^1 \wedge \cdots \wedge du^n = \int_{\mathbb{R}^n} (-1)^{j-1} \frac{\partial f}{\partial u^j} du^1 \cdots du^n \\
&= 0
\end{aligned}$$

by the fundamental theorem of calculus and the fact that f has compact support.

Case 9.2 (2) *Let $\omega_j = f du^1 \wedge \cdots \wedge \widehat{du^j} \wedge \cdots \wedge du^n$ be a smooth $n-1$ form with* *compact support meeting* $\partial\mathbb{R}^n_{u^1\leq 0} = \mathbb{R}^n_{u^1=0} = 0\times\mathbb{R}^{n-1}$ *then*

$$
\begin{aligned}
\int_{\mathbb{R}^n_{u^1\leq 0}} d\omega_j &= \int_{\mathbb{R}^n_{u^1\leq 0}} d(f du^1 \wedge \cdots \wedge \widehat{du^j} \wedge \cdots \wedge du^n) \\
&= \int_{\mathbb{R}^{n-1}} (-1)^{j-1} \left(\int_{-\infty}^{\infty} \frac{\partial f}{\partial u^j} du^j \right) du^1 \cdots \widehat{du^j} \cdots du^n = \\
&= 0 \text{ if } j \neq 1 \text{ and if } j = 1 \text{ we have } \int_{\mathbb{R}^n_{u^1\leq 0}} d\omega_1 = \\
&= \int_{\mathbb{R}^{n-1}} (-1)^{j-1} \left(\int_{-\infty}^{0} \frac{\partial f}{\partial u^1} du^1 \right) du^2 \wedge \cdots \wedge du^n \\
&= \int_{\mathbb{R}^{n-1}} f(0, u^2, ..., u^n) du^2 \cdots du^n \\
&= \int_{\partial\mathbb{R}^n_{u^1\leq 0}} f(0, u^2, ..., u^n) du^2 \wedge \cdots \wedge du^n = \int_{\partial\mathbb{R}^n_{u^1\leq 0}} \omega_1
\end{aligned}
$$

Now since clearly $\int_{\partial\mathbb{R}^n_{u^1\leq 0}} \omega_j = 0$ *if* $j \neq 1$ *or if* ω_j *has support that doesn't meet* $\partial\mathbb{R}^n_{u^1\leq 0}$ *we see that in any case* $\int_{\mathbb{R}^n_{u^1\leq 0}} d\omega_j = \int_{\partial\mathbb{R}^n_{u^1\leq 0}} \omega_j$. *Now as we said all* $n-1$ *forms on* $\mathbb{R}^n_{u^1\leq 0}$ *are sums of forms of this type and so summing such we have for any smooth* $n-1$ *form on* $\mathbb{R}^n_{u^1\leq 0}$.

$$
\int_{\mathbb{R}^n_{u^1\leq 0}} d\omega = \int_{\partial\mathbb{R}^n_{u^1\leq 0}} \omega.
$$

Now we define integration on a manifold (possibly with boundary). Let $\mathcal{A}_M = (\mathtt{x}_\alpha, U_\alpha)_{\alpha\in A}$ be an oriented atlas for a smooth orientable n-manifold M consisting of nice charts so either $\mathtt{x}_\alpha : U_\alpha \cong \mathbb{R}^n$ or $\mathtt{x}_\alpha : U_\alpha \cong \mathbb{R}^n_{u^1\leq 0}$. Now let $\{\rho_\alpha\}$ be a smooth partition of unity subordinate to $\{U_\alpha\}$. Notice that $\{\rho_\alpha|_{U_\alpha\cap\partial M}\}$ is a partition of unity for the cover $\{U_\alpha \cap \partial M\}$ of ∂M. Then for $\omega \in \Omega^{n-1}(M)$ we have that

$$
\begin{aligned}
\int_M d\omega &= \int_{U_\alpha} \sum_\alpha d(\rho_\alpha\omega) = \sum_\alpha \int_{U_\alpha} d(\rho_\alpha\omega) \\
&= \sum_\alpha \int_{\mathtt{x}_\alpha(U_\alpha)} \mathtt{x}_\alpha^* d(\rho_\alpha\omega) = \sum_\alpha \int_{\psi\mathtt{x}_\alpha(U_\alpha)} d(\mathtt{x}_\alpha^*\rho_\alpha\omega) \\
&= \sum_\alpha \int_{\mathtt{x}_\alpha(U_\alpha)} d(\mathtt{x}_\alpha^*\rho_\alpha\omega) = \sum_\alpha \int_{\partial\{\mathtt{x}_\alpha(U_\alpha)\}} (\mathtt{x}_\alpha^*\rho_\alpha\omega) \\
&= \sum_\alpha \int_{\partial U_\alpha} \rho_\alpha\omega = \int_{\partial M} \omega
\end{aligned}
$$

so we have proved

Theorem 9.1 (Stokes' Theorem) *Let M be an oriented manifold with boundary (possibly empty) and give ∂M the induced orientation. Then for any $\omega \in \Omega^{n-1}(M)$ we have*

$$\int_M d\omega = \int_{\partial M} \omega$$

9.3 Differentiating integral expressions

Suppose that $S \subset M$ is a regular submanifold with boundary ∂S (possibly empty) and Φ_t is the flow of some vector field $X \in \mathfrak{X}(M)$. In this case $S_t := \Phi_t(S)$ is also an regular submanifold with boundary. We then consider $\frac{d}{dt}\int_S \Phi_t^*\eta$. We have

$$\begin{aligned}
\frac{d}{dt}\int_S \Phi_t^*\eta &= \lim_{h\to 0}\frac{1}{h}\left[\int_S \Phi_{t+h}^*\eta - \int_S \Phi_t^*\eta\right] \\
&= \lim_{h\to 0}\frac{1}{h}\left[\Phi_t^*\int_S (\Phi_h^*\eta - \eta)\right] \\
&= \lim_{h\to 0}\left[\Phi_t^*\int_S \frac{1}{h}(\Phi_h^*\eta - \eta)\right] \\
&= \left[\int_{\Phi_t S} \lim_{h\to 0}\frac{1}{h}(\Phi_h^*\eta - \eta)\right] \\
&= \int_{S_t} \mathcal{L}_X\eta
\end{aligned}$$

But also $\int_S \Phi_t^*\eta = \int_{S_t}\eta$ and the resulting formula $\frac{d}{dt}\int_{S_t}\eta = \int_{\Phi_t S}\mathcal{L}_X\eta$ is quite useful. As a special case ($t = 0$) we have

$$\left.\frac{d}{dt}\right|_{t=0}\int_{S_t}\eta = \int_S \mathcal{L}_X\eta$$

We can go farther using Cartan's formula $\mathcal{L}_X = \iota_X \circ d + d \circ \iota_X$. We get

$$\begin{aligned}
\frac{d}{dt}\int_{S_t}\eta = \frac{d}{dt}\int_S \Phi_t^*\eta &= \int_{\Phi_t S}\iota_X d\eta + \int_{\Phi_t S} d\iota_X\eta \\
&= \int_{\Phi_t S}\iota_X d\eta + \int_{\partial S_t}\iota_X\eta.
\end{aligned}$$

This becomes particularly interesting in the case that $S = \Omega$ is an open submanifold of M with compact closure and smooth boundary and vol is a volume form on M. We then have $\frac{d}{dt}\int_{\Omega_t} vol = \int_{\partial\Omega_t}\iota_X vol$ and then

$$\left.\frac{d}{dt}\right|_{t=0}\int_{\Omega_t} vol = \int_{\partial\Omega}\iota_X vol$$

Definition 9.7 *If vol is a volume form orienting a manifold M then $\mathcal{L}_X vol = (\operatorname{div} X)\, vol$ for a unique function $\operatorname{div} X$ called the divergence of X with respect to the volume form vol.*

We have

$$\left.\frac{d}{dt}\right|_{t=0}\int_{\Omega_t} vol = \int_{\partial\Omega} \iota_X vol$$

and

$$\left.\frac{d}{dt}\right|_{t=0}\int_{\Omega_t} vol = \int_{\Omega} \mathcal{L}_X vol = \int_{\Omega} (\operatorname{div} X)\, \mathrm{d}\, V.$$

Now let $E_1, ..., E_n$ be a local frame field on $U \subset M$ and $\varepsilon^1, ..., \varepsilon^n$ the dual frame field. Then for some smooth function ρ we have

$$vol = \rho\varepsilon^1 \wedge \cdots \wedge \varepsilon^n$$

and so

$$\begin{aligned}
\mathcal{L}_X vol &= \mathcal{L}_X \left(\rho\varepsilon^1 \wedge \cdots \wedge \varepsilon^n\right) = d\iota_X \left(\rho\varepsilon^1 \wedge \cdots \wedge \varepsilon^n\right) \\
&= d\sum_{j=1}^{n}(-1)^{j-1}\rho\varepsilon^1 \wedge \cdots \iota_X\varepsilon^k \wedge \cdots \wedge \varepsilon^n \\
&= d\sum_{j=1}^{n}(-1)^{j-1}\rho\varepsilon^1 \wedge \cdots \wedge \varepsilon^k(X) \wedge \cdots \wedge \varepsilon^n \\
&= d\sum_{j=1}^{n}(-1)^{j-1}\rho\varepsilon^1 \wedge \cdots \wedge \varepsilon^k(\sum_{r=1}^{n} X^r\varepsilon_r) \wedge \cdots \wedge \varepsilon^n \\
&= d\sum_{k=1}^{n}(-1)^{j-1}\rho X^k\varepsilon^1 \wedge \cdots \wedge \widehat{\varepsilon^k} \wedge \cdots \wedge \varepsilon^n \\
&= \sum_{k=1}^{n}(-1)^{j-1}d(\rho X^k) \wedge \varepsilon^1 \wedge \cdots \wedge \widehat{\varepsilon^k} \wedge \cdots \wedge \varepsilon^n \\
&= \sum_{k=1}^{n}(-1)^{j-1}\sum_{i=1}^{n}(\rho X^k)_i\varepsilon^i \wedge \varepsilon^1 \wedge \cdots \wedge \widehat{\varepsilon^k} \wedge \cdots \wedge \varepsilon^n \\
&= \sum_{k=1}^{n}\left(\frac{1}{\rho}(\rho X^k)_k\right)\rho\varepsilon^1 \wedge \cdots \wedge \varepsilon^n \\
&\sum_{k=1}^{n}\left(\frac{1}{\rho}(\rho X^k)_k\right) vol
\end{aligned}$$

where $(\rho X^k)_k := d(\rho X^k)(E_k)$. Thus

$$\operatorname{div} X = \sum_{k=1}^{n}\frac{1}{\rho}(\rho X^k)_k$$

In particular, if $E_k = \frac{\partial}{\partial x^k}$ for some chart $U, \mathtt{x} = (x^1, ..., x^n)$ then

$$\operatorname{div} X = \sum_{k=1}^{n}\frac{1}{\rho}\frac{\partial}{\partial x^k}(\rho X^k)$$

Now if we were to replace the volume form vol by $-vol$ then divergence with respect to that volume form would be given locally by $\sum_{k=1}^{n} \frac{1}{-\rho} \frac{\partial}{\partial x^k}(-\rho X^k) = \sum_{k=1}^{n} \frac{1}{\rho} \frac{\partial}{\partial x^k}(\rho X^k)$ and so the orientation seems superfluous! What if M isn't even orientable? In fact, since divergence is a local concept and orientation a global concept it seems that we should replace the volume form in the definition by something else that makes sense even on a nonorientable manifold. But what? This brings us to our next topic.

9.4 Pseudo-forms

Let M be a smooth manifold and $\{U_\alpha, \mathtt{x}_\alpha\}$ be an atlas for M. The flat orientation line bundle $O_{flat}(TM)$ is the vector bundle constructed from the local bundles $U_\alpha \times \mathbb{R}$ by the transition functions $g_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(1, \mathbb{R})$ defined by

$$g_{\alpha\beta}(p) := \frac{\det\left(\mathtt{x}_\alpha \circ \mathtt{x}_\beta^{-1}\right)}{\left|\det\left(\mathtt{x}_\alpha \circ \mathtt{x}_\beta^{-1}\right)\right|} = \pm 1$$

For every chart $(U_\alpha, \mathtt{x}_\alpha)$ there is a frame field for $O_{flat}(TM)$ consisting of a single section $o_{\mathtt{x}_\alpha}$ over U_α. A pseudo-k-form is a cross section of $O_{flat}(TM) \otimes \wedge^k T^*M$. The set of all pseudo-$k$-forms will be denoted by $\Omega_o^k(M)$. Now we can extend the exterior product to maps $\wedge : \Omega_o^k(M) \times \Omega^l(M) \to \Omega_o^{k+l}(M)$ by the rule

$$(o_1 \otimes \theta_1) \wedge \theta_2 = o_1 \otimes \theta_1 \wedge \theta_2$$

with a similar and obvious map $\wedge : \Omega_o^k(M) \times \Omega^l(M) \to \Omega_o^{k+l}(M)$. Similarly, we have a map $\wedge : \Omega_o^k(M) \times \Omega_o^l(M) \to \Omega_o^{k+l}(M)$ given by

$$(o_1 \otimes \theta_1) \wedge (o_2 \otimes \theta_2) = o_1 o_2 \theta_1 \wedge \theta_2$$

and the map $o_1 o_2 \in C^\infty(M)$ is defined by fiberwise multiplication. Now we can extend the exterior algebra to $\sum_{k,l=0}^{n} \left(\Omega^k(M) \oplus \Omega_o^l(M)\right)$. If $\omega \in \Omega_o^k(M)$ then with respect to the chart $(U_\alpha, \mathtt{x}_\alpha)$, α has the local expression

$$\omega = o_{\mathtt{x}_\alpha} \otimes a_{\vec{I}}^\alpha dx_\alpha^{\vec{I}}$$

and if $\omega = o_{\mathtt{x}_\beta} \otimes a_{\vec{J}}^\beta dx_\beta^{\vec{J}}$ for some other chart $(U_\beta, \mathtt{x}_\beta)$ then $a_{\vec{I}}^\alpha = \frac{\det\left(\mathtt{x}_\alpha \circ \mathtt{x}_\beta^{-1}\right)}{\left|\det\left(\mathtt{x}_\alpha \circ \mathtt{x}_\beta^{-1}\right)\right|} \frac{dx^{\vec{J}}}{dx_\beta^{\vec{I}}} a_{\vec{J}}^\beta$. In particular if ω is a pseudo-n-form (a volume pseudo-form) then $\vec{I} = (1, 2, ..., n) = \vec{J}$ and $\frac{dx^{\vec{J}}}{dx_\beta^{\vec{I}}} = \det\left(\mathtt{x}_\alpha \circ \mathtt{x}_\beta^{-1}\right)$ and so that rule becomes

$$a_{12..n}^\alpha = \frac{\det\left(\mathtt{x}_\alpha \circ \mathtt{x}_\beta^{-1}\right)}{\left|\det\left(\mathtt{x}_\alpha \circ \mathtt{x}_\beta^{-1}\right)\right|} \det\left(\mathtt{x}_\alpha \circ \mathtt{x}_\beta^{-1}\right) a_{12..n}^\beta$$

or

$$a_{12..n}^\alpha = \left|\det\left(\mathtt{x}_\alpha \circ \mathtt{x}_\beta^{-1}\right)\right| a_{12..n}^\beta.$$

There is another way to think about pseudo-forms that has the advantage of having a clearer global description. Recall that the set of frames $\{e_1, ..., e_n\}$ in T_pM are divided into two equivalence classes called orientations. Equivalently, an orientation at p is an equivalence class of elements of $\wedge^n T_p^* M$. For each p there are exactly two such orientations and the set of all orientations $\mathrm{Or}(M)$ at all points $p \in M$ form a set with a differentiable structure and the map $\wp_{Or}$ that takes both orientations at p the point p is a two fold covering map. The group $\mathbb{Z}/2\mathbb{Z}$ actions as deck transformations on $\mathrm{Or}(M)$ sending each orientation to its opposite. Denote the action of $g \in \mathbb{Z}/2\mathbb{Z}$ by $l_g : \mathrm{Or}(M) \to \mathrm{Or}(M)$. Now we think of a pseudo-k-form as being nothing more than a k-form η on the manifold $\mathrm{Or}(M)$ with the property that $l_g^* \eta = \eta$ for each $g \in \mathbb{Z}/2\mathbb{Z}$. Now we would like to be able to integrate a k-form over a map $h : N \to M$ where N is a $k-$dimensional manifold. By definition h is said to be orientable if there is a lift $\widetilde{h} : \mathrm{Or}(N) \to \mathrm{Or}(M)$

$$
\begin{array}{ccc}
\mathrm{Or}(N) & \overset{\widetilde{h}}{\to} & \mathrm{Or}(M) \\
\downarrow \wp_{Or_N} & & \downarrow \wp_{Or_M} \\
N & \overset{h}{\to} & M
\end{array}
$$

We will say that $\widetilde{h}$ is said to orient the map. In this case we define the integral of a pseudo-k-form η over h to be

$$\int_h \eta := \frac{1}{2} \int_{\mathrm{Or}(N)} \widetilde{h}^* \eta$$

Now there is clearly another lift $\widetilde{h_-}$ which sends each $\widetilde{n} \in \mathrm{Or}(N)$ the opposite orientation of $\widetilde{h}(\widetilde{n})$. This is nothing more that saying $\widetilde{h_-} = l_g^* \widetilde{h}$.

Exercise 9.1 *Show that there are at most two such lifts $\widetilde{h}$.*

Now

$$\int_{\mathrm{Or}(N)} \widetilde{h}_-^* \eta = \int_{\mathrm{Or}(N)} \widetilde{h}^* l_g^* \eta = \int_{\mathrm{Or}(N)} \widetilde{h}^* \eta$$

and so the definition of $\int_h \eta$ is independent of the lift $\widetilde{h}$.

If $S \subset M$ is a regular k-submanifold and if the inclusion map $\iota_S : S \hookrightarrow M$ map is orientable than we say that S has a transverse orientation in M. In this case we define the integral of a pseudo-k-form η over S to be

$$\int_S \eta := \int_{\iota_S} \eta = \frac{1}{2} \int_{\mathrm{Or}(N)} \widetilde{\iota_S}^* \eta$$

Exercise 9.2 *Show that the identity map $id_M : M \to M$ is orientable:*

$$
\begin{array}{ccc}
\mathrm{Or}(M) & \overset{\widetilde{id}}{\to} & \mathrm{Or}(M) \\
\downarrow \wp_{Or_M} & & \downarrow \wp_{Or_M} \\
M & \overset{id}{\to} & M
\end{array}
$$

$\widetilde{id} = \pm id_{Or(M)}$.

Now finally, if ω is a pseudo-n-form on M then by definition

$$\int_M \omega := \frac{1}{2}\int_{\mathrm{Or}(M)} \widetilde{id}^{*}\omega = \frac{1}{2}\int_{\mathrm{Or}(M)} \omega$$

If $(U, \mathtt{x})$ is a chart then the map $\sigma_{\mathtt{x}} : p \mapsto [dx^1 \wedge \cdots \wedge dx^n]$ is a local cross section of the covering $\wp_{Or} : \mathrm{Or}(M) \to M$ meaning that the following diagram commutes

$$\begin{array}{ccc} & & \mathrm{Or}(M) \\ & \overset{\sigma_{\mathtt{x}}}{\nearrow} & \downarrow \wp_{Or} \\ U & \hookrightarrow & M \end{array}$$

and we can define the integral of ω locally using a partition of unity $\{\rho_\alpha\}$ subordinate to a cover U_α, ρ_α :

$$\int_M j\omega := \sum_\alpha \int_{U_\alpha} \rho_\alpha \sigma_{\mathtt{x}_a}^{*}\omega.$$

Now suppose we have a vector field $X \in \mathfrak{X}(M)$. Since $\wp_{Or} : \mathrm{Or}(M) \to M$ is a surjective local diffeomorphism there is a vector field $\widetilde{X} \in \mathfrak{X}(\widetilde{M})$ such that $T\wp \cdot \widetilde{X}_{\widetilde{p}} = X_p$ (where $\wp_{\mathrm{Or}}(\widetilde{p}) = p$) for all p. Similarly, if vol is a volume form on M then there is a volume pseudo-form $\widetilde{vol}$ on M, i.e. a $\mathbb{Z}/2\mathbb{Z}$ invariant $n-$form $\widetilde{vol}$ on $\mathrm{Or}(M)$ such $\widetilde{vol} = \wp_{\mathrm{Or}}^{*} vol$. In this case, it is easy to show that the divergence $\operatorname{div}\widetilde{X}$ of $\widetilde{X}$ with respect to $\widetilde{vol}$ is the lift of $\operatorname{div} X$ (with respect to vol). Thus if M is not orientable and so has no volume form we may still define $\operatorname{div} X$ (with respect to the pseudo-volume form $\widetilde{vol}$) to be the unique vector field on M which is $\wp_{\mathrm{Or}}-$related to $\operatorname{div}\widetilde{X}$ (with respect to volume form $\widetilde{vol}$ on $\mathrm{Or}(M)$).

Chapter 10

Immersion and Submersion.

The following proposition follows from the inverse function theorem:

Proposition 10.1 *If $f : M \to N$ is a smooth map such that $T_p f : T_p M \to T_q N$ is an isomorphism for all $p \in M$ then $f : M \to N$ is a local diffeomorphism.*

Definition 10.1 *Let $f : M \to N$ be C^r-map and $p \in M$ we say that p is a* ***regular point*** *for the map f if $T_p f$ is a splitting surjection (see 1.13) and is called a* ***singular point*** *otherwise. For finite dimensional manifolds this amounts to the requirement that $T_p f$ have full rank. A point q in N is called a* ***regular value*** *of f if every point in the inverse image $f^{-1}\{q\}$ is a regular point for f. A point of N that is not regular is called a* ***critical value****.*

It is a very useful fact that regular values are easy to come by in that most values are regular. In order to make this precise we will introduce the notion of measure zero on a manifold. It is actually no problem to define a Lebesgue measure on a manifold but for now the notion of measure zero is all we need.

Definition 10.2 *A set A in a smooth finite dimensional manifold M is said to be of* ***measure zero*** *if for every admissible chart U, ϕ the set $\phi(A \cap U)$ has Lebesgue measure zero in $\mathbb{R}^n$ where $\dim M = n$.*

In order for this to be a reasonable definition the manifold must be second countable so that every atlas has a countable subatlas. This way we may be assured that every set that we have defined to be measure zero is the countable union of sets that are measure zero as viewed in a some chart. We also need to know that the local notion of measure zero is independent of the chart. This follows from

Lemma 10.1 *Let M be a n-manifold. The image of a measure zero set under a differentiable map is of measure zero.*

Proof. We are assuming, of course that M is Hausdorff and second countable. Thus any set is contained in the countable union of coordinate charts. We may assume that $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ and A is some measure zero subset of U. In fact, since A is certainly contained in the countable union of compact balls (all of which are translates of a ball at the origin) we may as well assume that $U = B(0,r)$ and that A is contained in a slightly smaller ball $B(0, r-\delta) \subset B(0,r)$. By the mean value theorem, there is a constant c depending only on f and its domain such that for $x, y \in B(0,r)$ we have $|f(y) - f(x)| \le c\,|x - y|$. Let $\epsilon > 0$ be given. Since A has measure zero there is a sequence of balls $B(x_i, \epsilon_i)$ such that $A \subset \bigcup B(x_i, \epsilon_i)$ and

$$\sum vol(B(x_i, \epsilon_i)) < \frac{\epsilon}{2^n c^n}$$

Thus $f(B(x_i, \epsilon_i)) \subset B(f(x_i), 2c\epsilon_i)$ and while $f(A) \subset \bigcup B(f(x_i), 2c\epsilon_i)$ we also have

$$\begin{aligned}
&\mathrm{vol}\left(\bigcup B(f(x_i), 2c\epsilon_i)\right) \le \\
&\sum \mathrm{vol}(B(f(x_i), 2c\epsilon_i)) \le \sum \mathrm{vol}(B_1)(2c\epsilon_i)^n \\
&\qquad \le 2^n c^n \sum \mathrm{vol}(B(x_i, \epsilon_i)) \\
&\qquad \le \epsilon.
\end{aligned}$$

Thus the measure of A is less than or equal to ϵ. Since ϵ was arbitrary it follows that A has measure zero. ∎

Corollary 10.1 *Given a fixed* $\mathcal{A} = \{U_\alpha, \mathtt{x}_\alpha\}$ *atlas for* M*, if* $\mathtt{x}_\alpha(A \cap U_\alpha)$ *has measure zero for all* α *then* A *has measure zero.*

Theorem 10.1 (Sard) *Let* N *be an* n*-manifold and* M *an* m*-manifold (Hausdorff and second countable). For a smooth map* $f : N \to M$ *the set of critical values has Lebesgue measure zero.*

Proof. Through the use of a countable cover of the manifolds in question by charts we may immediately reduce to the problem of showing that for a smooth map $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ the set of critical values $C \subset U$ has image $f(C)$ of measure zero. We will use induction on the dimension n. For $n = 0$, the set $f(C)$ is just a point (or empty) and so has measure zero. Now assume the theorem is true for all dimensions $j \le n - 1$. We seek to show that the truth of the theorem follows for $j = n$ also.

Let us use the following common notation: For any k-tuple of nonnegative integers $\alpha = (i_1, ..., i_k)$ we let

$$\frac{\partial^{|\alpha|} f}{\partial x^\alpha} := \frac{\partial^{i_1 + \ldots + i_k} f}{\partial x^{i_1} \cdots \partial x^{i_k}}$$

where $|\alpha| := i_1 + ... + i_k$. Now let

$$C_i := \{x \in U : \ \frac{\partial^{|\alpha|} f}{\partial x^\alpha}(x) = 0 \text{ for all } |\alpha| \le i\}.$$

Then

$$C = (C \backslash C_1) \cup (C_1 \backslash C_2) \cup \cdots \cup (C_{k-1} \backslash C_k) \cup C_k$$

so we will be done if we can show that

a) $f(C \backslash C_1)$ has measure zero,

b) $f(C_{j-1} \backslash C_j)$ has measure zero and

c) $f(C_k)$ has measure zero for some sufficiently large k.

Proof of a): We may assume that $m \geq 2$ since if $m = 1$ we have $C = C_1$. Now let $x \in C \backslash C_1$ so that some first partial derivative is not zero at $x = a$. By reordering we may assume that this partial is $\frac{\partial f}{\partial x^1}$ and so the

$$(x^1, ..., x^n) \mapsto (f(x), x^2, ..., x^n)$$

map restricts to a diffeomorphism ϕ on some open neighborhood containing x. Since we may always replace f by the equivalent map $f \circ \phi^{-1}$ we may go ahead and assume without loss of generality that f has the form

$$f : x \mapsto (x^1, f^2(x), ..., f^m(x)) := (x^1, h(x))$$

on some perhaps smaller neighborhood V containing a. The Jacobian matrix for f in V is of the form

$$\begin{bmatrix} 1 & 0 \\ * & Dh \end{bmatrix}$$

and so $x \in V$ is critical for f if and only if it is critical for h. Now $h(C \cap V) \subset \mathbb{R}^{m-1}$ and so by the induction hypothesis $h(C \cap V)$ has measure zero in $\mathbb{R}^{m-1}$. Now $f(C \cap V) \cap (\{x\} \times \mathbb{R}^{m-1}) \subset \{x\} \times h(C \cap V)$ which has measure zero in $\{x\} \times \mathbb{R}^{m-1} \cong \mathbb{R}^{m-1}$ and so by Fubini's theorem $f(C \cap V)$ has measure zero. Since we may cover C by a countable number of sets of the form $C \cap V$ we conclude that $f(C)$ itself has measure zero.

Proof of (b): The proof of this part is quite similar to the proof of (a). Let $a \in C_{j-1} \backslash C_j$. It follows that some k-th partial derivative is 0 and after some permutation of the coordinate functions we may assume that

$$\frac{\partial}{\partial x^1} \frac{\partial^{|\beta|} f^1}{\partial x^\beta}(a) \neq 0$$

for some $j-1$- tuple $\beta = (i_1, ..., i_{j-1})$ where the function $g := \frac{\partial^{|\beta|} f^1}{\partial x^\beta}$ is zero at a since a is in C_{k-1}. Thus as before we have a map

$$x \mapsto (g(x), x^2, ..., x^n)$$

which restricts to a diffeomorphism ϕ on some open set V. We use ϕ, V as a chart about a. Notice that $\phi(C_{j-1} \cap V) \subset 0 \times \mathbb{R}^{n-1}$. We may use this chart ϕ to replace f by $g = f \circ \phi^{-1}$ which has the form

$$x \mapsto (x^1, h(x))$$

for some map $h : V \to \mathbb{R}^{m-1}$. Now by the induction hypothesis the restriction of g to

$$g_0 : \{0\} \times \mathbb{R}^{n-1} \cap V \to \mathbb{R}^m$$

has a set of critical values of measure zero. But each point from $\phi(C_{j-1} \cap V) \subset 0 \times \mathbb{R}^{n-1}$ is critical for g_0 since diffeomorphisms preserve criticality. Thus $g \circ \phi(C_{j-1} \cap V) = f(C_{j-1} \cap V)$ has measure zero.

Proof of (c): Let $I^n(r) \subset U$ be a cube of side r. We will show that if $k > (n/m) - 1$ then $f(I^n(r) \cap C_k)$ has measure zero. Since we may cover by a countable collection of such V the result follows. Now Taylor's theorem gives that if $a \in I^n(r) \cap C_k$ and $a + h \in I^n(r)$ then

$$|f(a+h) - f(a)| \le c\,|h|^{k+1} \tag{10.1}$$

for some constant c that depends only on f and $I^n(r)$. We now decompose the cube $I^n(r)$ into R^n cubes of side length r/R. Suppose that we label these cubes which contain critical points of f as $D_1, ...D_N$. Let D_i contain a critical point a of f. Now if $y \in D$ then $|y - a| \le \sqrt{n}r/R$ so using the Taylor's theorem remainder estimate above (10.1) with $y = a + h$ we see that $f(D_i)$ is contained in a cube $\widetilde{D}_i \subset \mathbb{R}^m$ of side

$$2c\left(\frac{\sqrt{n}r}{R}\right)^{k+1} = \frac{b}{R^{k+1}}$$

where the constant $b := 2c(\sqrt{n}r)^{k+1}$ is independent of the particular cube D from the decomposition and depends only on f and $I^n(r)$. The sum of the volumes of all such cubes $\widetilde{D}_i$ is

$$S \le R^n \left(\frac{b}{R^{k+1}}\right)^m$$

which, under the condition that $m(k+1) > n$, may be made arbitrarily small be choosing R large (refining the decomposition of $I^n(r)$). The result now follows. ∎

Corollary 10.2 *If M and N are finite dimensional manifolds then the regular values of a smooth map $f : M \to N$ are dense in N.*

10.1 Immersions

Definition 10.3 *A map $f : M \to N$ is called an* ***immersion at*** *$p \in M$ if and only if $T_p f : T_p M \to T_{f(p)} N$ is a (bounded) splitting injection (see 1.12) at p. $f : M \to N$ is called an* ***immersion*** *if f is an immersion at every $p \in M$.*

Figure 10.1 shows a simple illustration of an immersion of $\mathbb{R}^2$ into $\mathbb{R}^3$. This example is also an injective immersion (as far as is shown) but an immersion can come back and cross itself. Being an immersion at p only requires that the

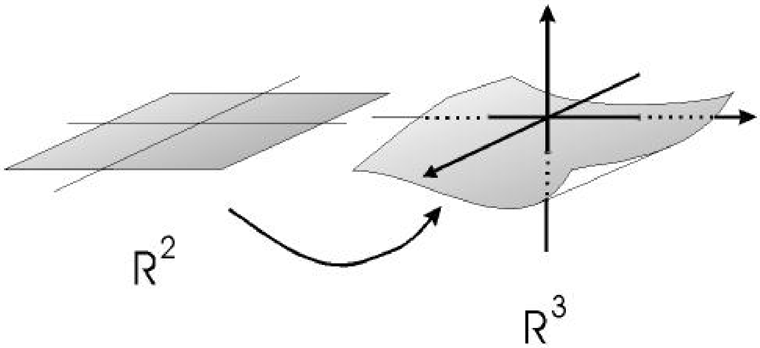


Figure 10.1: Embedding of the plane into 3d space.

restriction of the map to some small open neighborhood of p is injective. If an immersion is (globally) injective then we call it an immersed submanifold (see the definition 10.4 below).

Theorem 10.2 *Let $f : M \to N$ be a smooth function that is an immersion at p. Then there exist charts $\mathtt{x} :: (M, p) \to (\mathsf{M}, 0)$ and $\mathtt{y} :: (N, f(p)) \to (\mathsf{N}, 0) = (\mathsf{M} \times \mathsf{N}_2, 0)$ such that*

$$\mathtt{y} \circ f \circ \mathtt{x}^{-1} :: \mathsf{M} \to \mathsf{N} = \mathsf{M} \times \mathsf{N}_2$$

is given by $x \mapsto (x, 0)$ near 0. In particular, there is a open set $U \subset M$ such that $f(U)$ is a submanifold of N. In the finite dimensional case this means we can find charts on M and N containing p and $f(p)$, so that the coordinate expression for f is $(x^1, ..., x^k) \mapsto (x^1, ..., x^k, 0, ..., 0) \in \mathbb{R}^{k+m}$. Here, $k + m = n$ is the dimension of N and k is the rank of $T_p f$.

Proof. Follows easily from theorem 1.9. ∎

Theorem 10.3 *If $f : M \to N$ is an immersion (so an immersion at every point) and if f is a homeomorphism onto its image $f(M)$ using the relative topology, then $f(M)$ is a regular submanifold of N. In this case we call $f : M \to N$ an **embedding**.*

Proof. Follows from the last theorem plus a little point set topology. ∎

10.2 Immersed Submanifolds and Initial Submanifolds

Definition 10.4 *If $I : S \to M$ is an injective immersion then (S, I) is called an **immersed submanifold**.*

Exercise 10.1 *Show that every injective immersion of a compact manifold is an embedding.*

Theorem 10.4 *Suppose that M is an $n-$dimensional smooth manifold that has a finite atlas. Then there exists an injective immersion of M into $\mathbb{R}^{2n+1}$. Consequently, every* compact *$n-$dimensional smooth manifold can be embedded into $\mathbb{R}^{2n+1}$.*

Proof. Let M be a smooth manifold. Initially, we will settle for an immersion into $\mathbb{R}^D$ for some possibly very large dimension D. Let $\{O_i, \varphi_i\}_{i \in N}$ be an atlas with cardinality $N < \infty$. The cover $\{O_i\}$ cover may be refined to two other covers $\{U_i\}_{i \in N}$ and $\{V_i\}_{i \in N}$ such that $\overline{U_i} \subset V_i \subset \overline{V_i} \subset O_i$. Also, we may find smooth functions $f_i : M \to [0,1]$ such that

$$f_i(x) = 1 \text{ for all } x \in U_i$$
$$\text{supp}(f_i) \subset O_i.$$

Next we write $\varphi_i = (x_i^1, ...x_i^n)$ so that $x_i^j : O_i \to \mathbb{R}$ is the $j-$th coordinate function of the i-th chart and then let

$$f_{ij} := f_i x_i^j \text{ (no sum)}$$

which is defined and smooth on all of M after extension by zero.

Now we put the functions f_i together with the functions f_{ij} to get a map $i : M \to \mathbb{R}^{n+Nn}$:

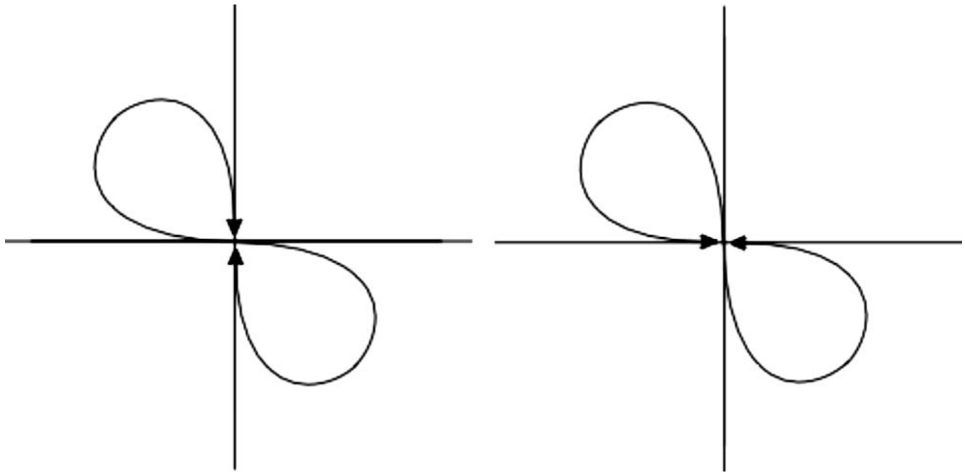
$$i = (f_1, ..., f_n, f_{11}, f_{12}, ..., f_{21}, ..., f_{nN}).$$

Now we show that i is injective. Suppose that $i(x) = i(y)$. Now $f_k(x)$ must be 1 for some k since $x \in U_k$ for some k. But then also $f_k(y) = 1$ also and this means that $y \in V_k$ (why?). Now then, since $f_k(x) = f_k(y) = 1$ it follows that $f_{kj}(x) = f_{kj}(y)$ for all j. Remembering how things were defined we see that x and y have the same image under $\varphi_k : O_k \to \mathbb{R}^n$ and thus $x = y$.

To show that $T_x i$ is injective for all $x \in M$ we fix an arbitrary such x and then $x \in U_k$ for some k. But then near this x the functions $f_{k1,} f_{k2}, ..., f_{kn,}$ are equal to $x_k^1, ...x_k^n$ and so the rank of i must be *at least* n and in fact equal to n since $\dim T_x M = n$.

So far we have an injective immersion of M into $\mathbb{R}^{n+Nn}$.

We show that there is a projection $\pi : \mathbb{R}^D \to L \subset \mathbb{R}^D$ where $L \cong \mathbb{R}^{2n+1}$ is a $2n+1$ dimensional subspace of $\mathbb{R}^D$, such that $\pi \circ f$ is an injective immersion. The proof of this will be inductive. So suppose that there is an injective immersion f of M into $\mathbb{R}^d$ for some d with $D \geq d > 2n+1$. We show that there is a projection $\pi_d : \mathbb{R}^d \to L^{d-1} \cong \mathbb{R}^{d-1}$ such that $\pi_d \circ f$ is still an injective immersion. To this end, define a map $h : M \times M \times \mathbb{R} \to \mathbb{R}^d$ by $h(x, y, t) := t(f(x) - f(y))$. Now since $d > 2n+1$, Sard's theorem implies that there is a vector $y \in \mathbb{R}^d$ which is neither in the image of the map h nor in the image of the map $df : TM \to \mathbb{R}^d$. This y cannot be 0 since 0 is certainly in the image of both of these maps. Now if $pr_{\perp y}$



is projection onto the orthogonal complement of y then $pr_{\perp y} \circ f$ is injective; for if $pr_{\perp y} \circ f(x) = pr_{\perp y} \circ f(y)$ then $f(x) - f(y) = ay$ for some $a \in \mathbb{R}$. But suppose $x \neq y$. then since f is injective we must have $a \neq 0$. This state of affairs is impossible since it results in the equation $h(x, y, 1/a) = y$ which contradicts our choice of y. Thus $pr_{\perp y} \circ f$ is injective.

Next we examine $T_x(pr_{\perp y} \circ f)$ for an arbitrary $x \in M$. Suppose that $T_x(pr_{\perp y} \circ f)v = 0$. Then $d(pr_{\perp y} \circ f)|_x v = 0$ and since $pr_{\perp y}$ is linear this amounts to $pr_{\perp y} \circ df|_x v = 0$ which gives $df|_x v = ay$ for some number $a \in \mathbb{R}$, and which cannot be 0 since f is assumed an immersion. But then $df|_x \frac{1}{a} v = y$ which also contradicts our choice of y.

We conclude that $pr_{\perp y} \circ f$ is an injective immersion. Repeating this process inductively we finally get a composition of projections $pr : \mathbb{R}^D \to \mathbb{R}^{2n+1}$ such that $pr \circ f : M \to \mathbb{R}^{2n+1}$ is an injective immersion. ■

It might surprise the reader that an immersed submanifold *does not* necessarily have the following property:

Criterion 10.1 *Let S and M be a smooth manifolds. An injective immersion $I : S \to M$ is called* ***smoothly universal*** *if for any smooth manifold N, a mapping $f : N \to S$ is smooth if and only if $I \circ f$ is smooth.*

To see what goes wrong, imagine the map corresponding to superimposing one of the figure eights shown in figure 10.2 onto the other. If $I : S \to M$ is an embedding then it is also smoothly universal but this is too strong of a condition for our needs. For that we make the following definitions which will be especially handy when we study foliations.

Definition 10.5 *Let S be any subset of a smooth manifold M. For any $x \in S$, denote by $C_x(S)$ the set of all points of S that can be connected to x by a smooth curve with image entirely inside S.*

Definition 10.6 *A subset $S \subset M$ is called an* ***initial submanifold*** *if for each $s_0 \in S$ there exists a chart $U, \mathtt{x}$ centered at s_0 such that $\mathtt{x}(C_{s_0}(U \cap S)) = \mathtt{x}(U) \cap (\mathsf{F} \times \{0\})$ for some splitting of the model space $\mathsf{M} = \mathsf{F} \times \mathsf{V}$ (that is independent of s_0).*

The definition implies that if S is an initial submanifold of M then it has a unique smooth structure as a manifold modelled on F and it is also not hard to see that any initial submanifold S has the property that the inclusion map $S \hookrightarrow M$ is smoothly universal. Conversely, we have the following

Theorem 10.5 *If an injective immersion $I : S \to M$ is smoothly universal then the image $f(S)$ is an initial submanifold.*

Proof. Choose $s_0 \in S$. Since I is an immersion we may pick a coordinate chart $w : W \to \mathsf{F}$ centered at s_0 and a chart $v : V \to \mathsf{M} = \mathsf{F} \times \mathsf{V}$ centered at $I(s_0)$ such that we have

$$v \circ I \circ w^{-1}(y) = (y, 0).$$

Choose an $r > 0$ small enough that $B(0, r) \subset w(U)$ and $B(0, 2r) \subset w(V)$. Let $U_0 = v^{-1}(B(0, r))$ and $W_0 = w^{-1}(V)$. We show that the coordinate chart $V_0, u := \varphi|_{V_0}$ satisfies the property of lemma 10.6.

$$\begin{aligned} u^{-1}(u(U_0) \cap (\mathsf{F} \times \{0\})) &= u^{-1}\{(y, 0) : \|y\| < r\} \\ &= I \circ w^{-1} \circ (u \circ I \circ w^{-1})^{-1}(\{(y, 0) : \|y\| < r\}) \\ &= I \circ w^{-1}(\{y : \|y\| < r\}) = I(W_0) \end{aligned}$$

Now $I(W_0) \subset U_0 \cap I(S)$ and since $I(W_0)$ is contractible we have $I(W_0) \subset C_{f(s_0)}(U_0 \cap I(S))$. Thus $u^{-1}(u(U_0) \cap (\mathsf{F} \times \{0\})) \subset C_{f(s_0)}(U_0 \cap I(S))$ or

$$u(U_0) \cap (\mathsf{F} \times \{0\}) \subset u(C_{I(s_0)}(U_0 \cap I(S))).$$

Conversely, let $z \in C_{I(s_0)}(U_0 \cap I(S))$. By definition there must be a smooth curve $c : [0, 1] \to S$ starting at $I(s_0)$, ending at z and $c([0, 1]) \subset U_0 \cap I(S)$. Since $I : S \to M$ is smoothly universal there is a unique smooth curve $c_1 : [0, 1] \to S$ with $I \circ c_1 = c$.
Claim: $c_1([0, 1]) \subset W_0$.
Assume not. Then there is some number $t \in [0, 1]$ with $c_1(t) \in w^{-1}(\{r \leq \|y\| < 2r\})$. Then

$$\begin{aligned} (v \circ I)(c_1(t)) &\in (v \circ I \circ w^{-1})(\{r \leq \|y\| < 2r\}) \\ &= \{(y, 0) : r \leq \|y\| < 2r\} \subset \{z \in \mathsf{M} : r \leq \|y\| < 2r\}. \end{aligned}$$

Now this implies that $(v \circ I \circ c_1)(t) = (v \circ c)(t) \in \{z \in \mathsf{M} : r \leq \|y\| < 2r\}$ which in turn implies the contradiction $c(t) \notin U_0$. The claim is proven.

Now the fact that $c_1([0,1]) \subset W_0$ implies $c_1(1) = I^{-1}(z) \in W_0$ and so $z \in I(W_0)$. As a result we have $C_{I(s_0)}(U_0 \cap I(S)) = I(W_0)$ which together with the first half of the proof gives the result:

$$\begin{aligned} I(W_0) = u^{-1}(u(U_0) \cap (\mathsf{F} \times \{0\})) \subset C_{f(s_0)}(U_0 \cap I(S)) = I(W_0) \\ \implies \quad u^{-1}(u(U_0) \cap (\mathsf{F} \times \{0\})) = C_{f(s_0)}(U_0 \cap I(S)) \\ \implies \quad u(U_0) \cap (\mathsf{F} \times \{0\}) = u(C_{f(s_0)}(U_0 \cap I(S))). \end{aligned}$$

■

We say that two immersed submanifolds (S_1, I_1) and (S_2, I_2) are **equivalent** if there exists a diffeomorphism $\Phi : S_1 \to S_2$ such that $I_2 \circ \Phi = I_1$; i.e. so that the following diagram commutes

$$\begin{array}{ccc} & \Phi & \\ S_1 & \to & S_2 \\ & \searrow \quad \swarrow & \\ & M & \end{array} .$$

Now if $I : S \to M$ is smoothly universal so that $f(S)$ is an initial submanifold then it is not hard to see that (S, I) is equivalent to $(f(S), \iota)$ where ι is the inclusion map and we give $f(S)$ the unique smooth structure guaranteed by the fact that it is an initial submanifold. Thus we may as well be studying initial submanifolds rather than injective immersions that are smoothly universal. For this reason we will seldom have occasion to even use the terminology "smoothly universal" which the author now confesses to be a nonstandard terminology anyway.

10.3 Submersions

Definition 10.7 *A map $f : M \to N$ is called a **submersion at** $p \in M$ if and only if $T_p f : T_p M \to T_{f(p)} N$ is a (bounded) splitting surjection (see 1.12). $f : M \to N$ is called a **submersion** if f is a submersion at every $p \in M$.*

Example 10.1 *The map of the punctured space $\mathbb{R}^3 - \{0\}$ onto the sphere S^2 given by $x \mapsto |x|$ is a submersion. To see this use spherical coordinates and the map becomes $(\rho, \phi, \theta) \mapsto (\phi, \theta)$. Here we ended up with a projection onto a second factor $\mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2$ but this is clearly good enough.*

Theorem 10.6 *Let $f : M \to N$ be a smooth function that is an submersion at p. Then there exist charts $\mathtt{x} :: (M, p) \to (\mathsf{N} \times \mathsf{M}_2, 0) = (\mathsf{M}, 0)$ and $\mathtt{y} :: (N, f(p)) \to (\mathsf{N}, 0)$ such that*

$$\mathtt{y} \circ f \circ \mathtt{x}^{-1} :: (\mathsf{N} \times \mathsf{M}_2, 0) \to (\mathsf{N}, 0)$$

is given by $(x, y) \mapsto x$ near $0 = (0, 0)$. In the finite dimensional case this means we can find charts on M and N containing p and $f(p)$, so that the coordinate *expression for f is $(x^1, ..., x^k, ..., x^n) \mapsto (x^1, ..., x^k) \in \mathbb{R}^k$. Here k is both the dimension of N and the rank of $T_p f$.*

Proof. Follows directly from theorem 1.11. ∎

Corollary 10.3 (Submanifold Theorem I) *Consider any smooth map $f : M \to N$ then if $q \in N$ is a regular value the inverse image set $f^{-1}(q)$ is a regular submanifold.*

Proof. If $q \in N$ is a regular value then f is a submersion at every $p \in f^{-1}(q)$. Thus for any $p \in f^{-1}(q)$ there exist charts $\psi :: (M, p) \to (\mathsf{N} \times \mathsf{M}_2, 0) = (\mathsf{M}, 0)$ and $\phi :: (N, f(p)) \to (\mathsf{N}, 0)$ such that

$$\phi \circ f \circ \psi^{-1} :: (\mathsf{N} \times \mathsf{M}_2, 0) \to (\mathsf{N}, 0)$$

is given by $(x, y) \mapsto x$ near $\mathsf{0}$. We may assume that domains are of the nice form

$$U' \times V' \overset{\psi^{-1}}{\to} U \overset{f}{\to} V \overset{\phi}{\to} V'.$$

But $\phi \circ f \circ \psi^{-1}$ is just projection and since q corresponds to $\mathsf{0}$ under the diffeomorphism we see that $\psi(U \cap f^{-1}(q)) = (\phi \circ f \circ \psi^{-1})^{-1}(0) = pr_1^{-1}(0) = U' \times \{0\}$ so that $f^{-1}(q)$ has the submanifold property at p. Now $p \in f^{-1}(q)$ was arbitrary so we have a cover of $f^{-1}(q)$ by submanifold charts. ∎

Example 10.2 (The unit sphere) *The set $S^{n-1} = \{x \in \mathbb{R}^n : x \cdot x = 1\}$ is a codimension 1 submanifold of $\mathbb{R}^n$ since we can use the map $(x, y) \mapsto x^2 + y^2$ as our map and let $q = 1$.*

Given k functions $F^j(x, y)$ on $\mathbb{R}^n \times \mathbb{R}^k$ we define the locus

$$M := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^k : F^j(x, y) = c^j\}$$

where each c^j is a fixed number in the range of F^j. If the Jacobian determinant at $(x_0, y_0) \in M$;

$$\left| \frac{\partial(F^1, ..., F^k)}{\partial(y^1, ..., y^k)} \right| (x_0, y_0)$$

is not zero then near (x_0, y_0) then we can apply the theorem. We can see things more directly: Since the Jacobian determinant is nonzero, we can solve the equations $F^j(x, y) = c^j$ for $y^1, ..., y^k$ in terms of $x^1, ..., x^k$:

$$y^1 = f^1(x^1, ..., x^k)$$
$$y^2 = f^2(x^1, ..., x^k)$$
$$y^k = f^k(x^1, ..., x^k)$$

and this parameterizes M near (x_0, y_0) in such a nice way that the inverse is a chart for M. This latter statement is really the content of the inverse mapping theorem in this case. If the Jacobian determinant never vanishes on M then we have a cover by charts and M is a submanifold of $\mathbb{R}^n \times \mathbb{R}^k$.

It may help the understanding to recall that if $F^j(x, y)$ and (x_0, y_0) are as above then we can differentiate the expressions $F^j(x, y(x)) = c^j$ to get

$$\frac{\partial F^i}{\partial x^j} + \sum_s \frac{\partial F^i}{\partial y^s}\frac{\partial y^s}{\partial x^j} = 0$$

and then solve

$$\frac{\partial y^j}{\partial x^i} = -\sum [J^{-1}]^j_s \frac{\partial F^s}{\partial x^i}$$

where $[J^{-1}]^j_s$ is the matrix inverse of the Jacobian $\left[\frac{\partial F}{\partial y}\right]$ evaluated at points near (x_0, y_0).

Example 10.3 *The set of all square matrices $\mathbb{M}_{n\times n}$ is a manifold by virtue of the obvious isomorphism $\mathbb{M}_{n\times n} \cong \mathbb{R}^{n^2}$. The set $\mathfrak{sym}(n, \mathbb{R})$ of all symmetric matrices is an $n(n+1)/2$-dimensional manifold by virtue of the obvious 1-1 correspondence $\mathfrak{sym}(n, \mathbb{R}) \cong \mathbb{R}^{n(n+1)/2}$ given by using $n(n+1)/2$ independent entries in the upper triangle of the matrix as coordinates.*
Now the set $\mathrm{O}(n, \mathbb{R})$ of all $n \times n$ orthogonal matrices is a submanifold of $\mathbb{M}_{n\times n}$. We can show this using Theorem 10.3 as follows. Consider the map $f : \mathbb{M}_{n\times n} \to \mathfrak{sym}(n, \mathbb{R})$ given by $A \mapsto A^t A$. Notice that by definition of $\mathrm{O}(n, \mathbb{R})$ we have $f^{-1}(I) = \mathrm{O}(n, \mathbb{R})$. Let us compute the tangent map at any point $Q \in f^{-1}(I) = \mathrm{O}(n, \mathbb{R})$. The tangent space of $\mathfrak{sym}(n, \mathbb{R})$ at I is $\mathfrak{sym}(n, \mathbb{R})$ itself since $\mathfrak{sym}(n, \mathbb{R})$ is a vector space. Similarly, $\mathbb{M}_{n\times n}$ is its own tangent space. Under the identifications of section 4.6 we have

$$T_Q f \cdot v = \frac{d}{ds}(Q^t + sv^t)(AQ + sv) = v^t Q + Q^t v.$$

Now this map is clearly surjective onto $\mathfrak{sym}(n, \mathbb{R})$ when $Q = I$. On the other hand, for any $Q \in \mathrm{O}(n, \mathbb{R})$ consider the map $L_{Q^{-1}} : \mathbb{M}_{n\times n} \to \mathbb{M}_{n\times n}$ given by $L_{Q^{-1}}(B) = Q^{-1}B$. The map $T_Q L_{Q^{-1}}$ is actually just $T_Q L_{Q^{-1}} \cdot v = Q^{-1}v$ which is a linear isomorphism since Q is a nonsingular. We have that $f \circ L_Q = f$ and so by the chain rule

$$\begin{aligned} T_Q f \cdot v &= T_I f \circ T_Q(L_{Q^{-1}}) \cdot v \\ &= T_I f \cdot Q^{-1} v \end{aligned}$$

which shows that $T_Q f$ is also surjective.

The following proposition shows an example of the simultaneous use of Sard's theorem and theorem 10.3.

Proposition 10.2 *Let M be a connected submanifold of $\mathbb{R}^n$ and let S be a linear subspace of $\mathbb{R}^n$. Then there exist a vector $v \in \mathbb{R}^n$ such that $(v + S) \cap M$ is a submanifold of M.*

Proof. Start with a line l through the origin that is normal to S. Let $pr : \mathbb{R}^n \to S$ be orthogonal projection onto l . The restriction $\pi := pr|_M \to l$ is easily seen to be smooth. If $\pi(M)$ were just a single point x then $\pi^{-1}(x)$ would be all of M. Now $\pi(M)$ is connected and a connected subset of $l \cong \mathbb{R}$ must contain an interval which means that $\pi(M)$ has positive measure. Thus by Sard's theorem there must be a point $v \in l$ that is a regular value of π. But then 10.3 implies that $\pi^{-1}(v)$ is a submanifold of M. But this is the conclusion since $\pi^{-1}(v) = (v + S) \cap M$. ∎

10.4 Morse Functions

If we consider a smooth function $f : M \to \mathbb{R}$ and assume that M is a compact manifold (without boundary) then f must achieve both a maximum at one or more points of M and a minimum at one or more points of M. Let x_e be one of these points. The usual argument shows that $df|_e = 0$ (Recall that under the usual identification of $\mathbb{R}$ with any of its tangent spaces we have $df|_e = T_e f$). Now let x_0 be some point for which $df|_{x_0} = 0$. Does f achieve either a maximum or a minimum at x_0? How does the function behave in a neighborhood of x_0? As the reader may well be aware, these questions are easier to answer in case the second derivative of f at x_0 is nondegenerate. But what is the second derivative in this case? One could use a Riemannian metric and the corresponding Levi-Civita connection (to be introduced later) to given an invariant notion but for us it will suffice to work in a local coordinate system and restrict our attention to critical points. Under these conditions the following definition of nondegeneracy is well defined independent of the choice of coordinates:

Definition 10.8 *The **Hessian** matrix of f at one of its critical points x_0 and with respect to coordinates $\psi = (x^1, ..., x^n)$ is the matrix of second partials:*

$$H = \begin{bmatrix} \frac{\partial^2 f\circ\psi^{-1}}{\partial x^1 \partial x^1}(x_0) & \cdots & \frac{\partial^2 f\circ\psi^{-1}}{\partial x^1 \partial x^n}(x_0) \\ \vdots & & \vdots \\ \frac{\partial^2 f\circ\psi^{-1}}{\partial x^n \partial x^1}(x_0) & \cdots & \frac{\partial^2 f\circ\psi^{-1}}{\partial x^n \partial x^n}(x_0) \end{bmatrix}$$

The critical point is called nondegenerate if H is nonsingular.

Now any such matrix H is symmetric and by Sylvester's law of inertia this matrix is equivalent to a diagonal matrix whose diagonal entries are either 1 or -1. The number of -1 occurring is called the **index** of the critical point.

Exercise 10.2 *Show that the nondegeneracy is well defined.*

Exercise 10.3 *Show that nondegenerate critical points are isolated. Show by example that this need not be true for general critical points.*

The structure of a function near one of its nondegenerate critical points is given by the following famous theorem of M. Morse:

Theorem 10.7 (Morse Lemma) *If $f : M \to \mathbb{R}$ is a smooth function and x_0 is a nondegenerate critical point for f of index i. Then there is a local coordinate system $U, \mathtt{x}$ containing x_0 such that the local representative $f_U := f \circ \mathtt{x}^{-1}$ for f has the form*

$$f_U(x^1, ..., x^n) = f(x_0) + \sum h_{ij} x^i x^j$$

and where it may be arranged that the matrix $h = (h_{ij})$ is a diagonal matrix of the form $\mathrm{diag}(-1, ... - 1, 1, ..., 1)$ *for some number (perhaps zero) of ones and minus ones. The number of minus ones is exactly the index i.*

Proof. This is clearly a local problem and so it suffices to assume $f ::$ $\mathbb{R}^n \to \mathbb{R}$ and also that $f(0) = 0$. Then our task is to show that there exists a diffeomorphism $\phi : \mathbb{R}^n \to \mathbb{R}^n$ such that $f \circ \phi(x) = x^t h x$ for a matrix of the form described. The first step is to observe that if $g : U \subset \mathbb{R}^n \to \mathbb{R}$ is any function defined on a convex open set U and $g(0) = 0$ then

$$\begin{aligned} g(u_1, ..., u_n) &= \int_0^1 \frac{d}{dt} g(tu_1, ..., tu_n) dt \\ &= \int_0^1 \sum_{i=1}^n u_i \partial_i g(tu_1, ..., tu_n) dt \end{aligned}$$

Thus g is of the form $g = \sum_{i=1}^n u_i g_i$ for certain smooth functions $g_i, 1 \le i \le n$ with the property that $\partial_i g(0) = g_i(0)$. Now we apply this procedure first to f to get $f = \sum_{i=1}^n u_i f_i$ where $\partial_i f(0) = f_i(0) = 0$ and then apply the procedure to each f_i and substitute back. The result is that

$$f(u_1, ..., u_n) = \sum_{i,j=1}^n u_i u_j h^{ij}(u_1, ..., u_n) \tag{10.2}$$

for some functions h^{ij} with the property that $h^{ij}()$ is nonsingular at and therefore near 0. Next we symmetrize (h^{ij}) by replacing it with $\frac{1}{2}(h^{ij} + h^{ji})$ if necessary. This leaves the expression 10.2 untouched. Now the index of the matrix $(h^{ij}(0))$ is i and this remains true in a neighborhood of 0. The trick is to find, for each x in the neighborhood a matrix $C(x)$ that effects the diagonalization guaranteed by Sylvester's theorem: $D = C(x)h(x)C(x)^{-1}$. The remaining details, including the fact that the matrix $C(x)$ may be chosen to depend smoothly on x, is left to the reader. ∎

We can generalize theorem 10.3 using the concept of transversality .

Definition 10.9 *Let $f : M \to N$ be a smooth map and $S \subset N$ a submanifold of N. We say that f is* ***transverse to*** *S if for every $p \in f^{-1}(S)$ the image of T_pM under the tangent map T_pf and $T_{f(p)}S$ together span all of $T_{f(p)}N$:*

$$T_{f(p)}N = T_{f(p)}S + T_pf(T_pM).$$

If f is ***transverse to*** *S we write $f \pitchfork S$.*

Theorem 10.8 *Let $f : M \to N$ be a smooth map and $S \subset N$ a submanifold of N and suppose that $f \pitchfork S$. Then $f^{-1}(S)$ is a submanifold of M. Furthermore we have $T_p(f^{-1}(S)) = T_{f(p)}f^{-1}(T_{f(p)}S)$ for all $p \in f^{-1}(S)$ and* $\operatorname{codim}(f^{-1}(S)) = \operatorname{codim}(S)$.

10.5 Problem set

1. Show that a submersion always maps open set to open set (it is an open mapping). Further show that if M is compact and N connected then a submersion $f : M \to N$ must be surjective.

2. Show that the set of all symmetric matrices $Sym_{n\times n}(\mathbb{R})$ is a submanifold of $\mathbb{M}_{n\times n}(\mathbb{R})$. Under the canonical identification of $T_S\mathbb{M}_{n\times n}(\mathbb{R})$ with $\mathbb{M}_{n\times n}(\mathbb{R})$ the tangent space of $Sym_{n\times n}(\mathbb{R})$ at the symmetric matrix S becomes what subspace of $\mathbb{M}_{n\times n}(\mathbb{R})$? Hint: It's kind of a trick question.

3. Prove that in each case below the subset $f^{-1}(p_0)$ is a submanifold of M:

   a) $f : M = \mathbb{M}_{n\times n}(\mathbb{R}) \to Sym_{n\times n}(\mathbb{R})$ and $f(Q) = Q^tQ$ with $p_0 = I$ the identity matrix.

   b) $f : M = \mathbb{R}^n \to \mathbb{R}^2$ and $f(x_1, ..., x_n) = (\,x_1{}^2 + \cdots + x_n{}^2\,, -\,x_1{}^2 + \cdots + x_n{}^2)$ with $p_0 = (0, 0)$.

4. Show that if $p(x) = p(x_1, ..., x_n)$ is a homogeneous polynomial so that for some $m \in \mathbb{Z}_+$
$$p(tx_1, ..., tx_n) = t^m p(x_1, ..., x_n)$$
then as long as $c \neq 0$ the set $p^{-1}(c)$ is a $n-1$ dimensional submanifold of $\mathbb{R}^n$.

5. Suppose that $g : M \to N$ is transverse to a submanifold $W \subset N$. For another smooth map $f : Y \to M$ show that $f \pitchfork g^{-1}(N)$ if and only if $(g \circ f) \pitchfork W$.

6. Suppose that $c : [a, b] \to M$ is a smooth map. Show that given any compact subset $C \subset (a, b)$ and any $\epsilon > 0$ there is an immersion $\gamma : (a, b) \to M$ that agrees with c on the set C and such that
$$|\gamma(t) - c(t)| \leq \epsilon \text{ for all } t \in (a, b).$$

7. Show that there is a continuous map $f : \mathbb{R}^2 \to \mathbb{R}^2$ such that $f(B(0, 1)) \subset f(B(0, 1))$, $f(\mathbb{R}^2\backslash B(0, 1)) \subset f(\mathbb{R}^2\backslash B(0, 1))$ and $f_{\partial B(0,1)} = \mathrm{id}_{\partial B(0,1)}$ and with the properties that f is C^∞ on $\overline{B}(0, 1)$ and on $\mathbb{R}^2\backslash B(0, 1)$ while f is not C^∞ on $\mathbb{R}^2$.

8. Construct an embedding of $\mathbb{R} \times S^n$ into $\mathbb{R}^{n+1}$

9. Construct an embedding of $G(n, k)$ into $G(n, k + l)$ for each $l \geq 1$.

Chapter 11

Fiber Bundles

11.1 General Fiber Bundles

This chapter is about a special kind of submersion called a general fiber bundle. We have already examples since every vector bundle is an example of a general fiber bundle.

Definition 11.1 *Let F, M and Æ be C^r manifolds and let $\pi : E \to M$ be a C^r map. The quadruple (E, π, M, F) is called a (locally trivial) C^r* ***fiber bundle*** *with typical fiber F if for each point $p \in M$ there is an open set U containing p and a homeomorphism $\phi : \pi^{-1}(U) \to U \times F$ such that the following diagram commutes:*

$$\begin{array}{ccc} \pi^{-1}(U) & \xrightarrow{\phi} & U \times F \\ \pi \searrow & & \swarrow pr_1 \\ & U & \end{array}$$

As with vector bundles, the space M is called the base space, E is called the total space and for each $p \in M$ the $\pi^{-1}\{p\}$ is called the fiber over p and shall also be denoted by E_p.

Exercise 11.1 *Show that if (E, π, M, F) is a (locally trivial) C^r fiber bundle with $r > 1$ then $\pi : E \to M$ is a submersion and each fiber $\pi^{-1}\{p\}$ is diffeomorphic to F.*

A bundle equivalence between two bundles (E_1, π_1, M, F) and (E_2, π_2, M, F) over the same base space M is a C^r diffeomorphism $h : E_1 \to E_2$ such that the following diagram commutes:

$$\begin{array}{ccc} E_1 & \xrightarrow{h} & E_2 \\ \pi_1 \searrow & & \swarrow \pi_2 \\ & M & \end{array}$$

The simplest example is the product bundle $(M \times F, pr_1, M, F)$ and any bundle which is equivalent to a product bundles is said to be trivial. The definition of a fiber bundle as we have given it includes the requirement that a bundle be locally trivial and the maps $\phi : \pi^{-1}(U) \to U \times F$ occurring in the definition are said to be trivializations over U. It is easy to see that each such trivialization must be a map of the form $\phi = (\pi, \Phi)$ where $\Phi : E \to F$ is a map with the property that $\Phi|_{E_p} : E_p \to F$ is a diffeomorphism. Giving a family of trivializations ϕ_α over corresponding open sets U_α which cover M is said to be a bundle atlas and the existence of such an atlas is enough to give the bundle structure. Each (ϕ_α, U_α) is called a bundle chart. Given two such trivializations or bundle charts (ϕ_α, U_α) and (ϕ_β, U_β) we have $\phi_\alpha = (\pi, \Phi_\alpha) : \pi^{-1}(U_\alpha) \to U_\alpha \times F$ and $\phi_\beta = (\pi, \Phi_\beta) : \pi^{-1}(U_\beta) \to U_\beta \times F$ and then for each $p \in U_\alpha \cap U_\beta$ there is a diffeomorphism $\Phi_{\alpha\beta}|_p$: defined by the composition $\Phi_\alpha|_{E_p} \circ \Phi_\beta|_{E_p}^{-1} : F \to F$. These are called **transition maps** or **transition functions**. Now allowing p vary we get a map $U_\alpha \cap U_\beta \to \mathrm{Diff}(F)$ defined by $p \mapsto \Phi_{\alpha\beta}|_p$. I could be that we may find a bundle atlas such that all of the transition maps are elements of a subgroup $G \subset \mathrm{Diff}(F)$. Also, G may be (isomorphic to) Lie group. In this case we say we have a $G-$bundle. From this point of view a vector bundle is a $Gl(n, \mathbb{R})$ bundle. It is sometimes convenient to be a bit more general and consider rather than subgroups of $\mathrm{Diff}(F)$ representations of Lie groups in $\mathrm{Diff}(F)$. Recall that a group action $\rho : G \times F \to F$ is equivalently thought of as a representation $\overline{\rho} : G \to \mathrm{Diff}(F)$ given by $\overline{\rho}(g)(f) = \rho(g, f)$. We will forgo the separate notation $\overline{\rho}$ and simple write ρ for the action and the corresponding representation. Returning to our discussion of fiber bundles, suppose that there is a Lie group action $\rho : G \times F \to F$ and cover of E by trivializations $\{U_\alpha, \phi_\alpha\}$ such that for each α, β we have

$$\Phi_{\alpha\beta}(p)(f) = \rho(g_{\alpha\beta}(p), f) := g_{\alpha\beta}(p) \cdot f$$

for some smooth map $g_{\alpha\beta} : U_\alpha \cap U_\beta \to G$ then we say that we have presented (π, E, M, F) as G **bundle under the representation** ρ. We call (π, E, M, F, G) a (G, ρ) **bundle** or just a $G-$bundle if the representation is understood or standard in some way. It is common to call G the **structure group** but since the action in question may not be effective we should really refer to the **structure group representation (or action)** ρ. We also say that the transition functions **live in** G (via ρ). In many but not all cases the representation ρ will be faithful, i.e. the action will be **effective** and so G can be considered as a subgroup of $\mathrm{Diff}(F)$ and the term 'G bundle' becomes more accurate. A notable exception is the case of spin bundles where the representation has a nontrivial (but finite) kernel.

A fiber bundle is determined if we are given an open cover $\mathcal{U} = \{U_\alpha\}$ and maps $\Phi_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{diff}(F)$ such that for all α, β, γ

$$\begin{aligned} \Phi_{\alpha\alpha}(p) &= \mathrm{id}\, \text{for } p \in U_\alpha \\ \Phi_{\alpha\beta}(p) &= \Phi_{\beta\alpha}^{-1}(p) \text{ for } p \in U_\alpha \cap U_\beta \\ \Phi_{\alpha\beta}(p) \circ \Phi_{\beta\gamma}(p) \circ \Phi_{\gamma\alpha}(p) &= \ \mathrm{id} \ \text{ for } p \in U_\alpha \cap U_\beta \cap U_\gamma \end{aligned}$$

If we want a (G, ρ) bundle then we further require that $\Phi_{\alpha\beta}(p)(f) = \rho(g_{\alpha\beta}(p))(f)$ as above and that the maps $g_{\alpha\beta}$ themselves satisfy the cocycle condition:

$$\begin{aligned} g_{\alpha\alpha}(p) &= \mathrm{id} \text{ for } p \in U_\alpha \\ g_{\alpha\beta}(p) &= g_{\beta\alpha}^{-1}(p) \text{ for } p \in U_\alpha \cap U_\beta \\ g_{\alpha\beta}(p) \circ g_{\beta\gamma}(p) \circ g_{\gamma\alpha}(p) &= \mathrm{id} \text{ for } p \in U_\alpha \cap U_\beta \cap U_\gamma \end{aligned} \tag{11.1}$$

We shall also call the maps $g_{\alpha\beta}$ **transition functions** or **transition maps**. Notice that if ρ is effective the last condition follows from the first. The family $\{U_\alpha\}$ together with the maps $\Phi_{\alpha\beta}$ form a **cocycle** and we can construct a bundle by taking the disjoint union $\bigsqcup(U_\alpha \times F) = \bigcup U_\alpha \times F \times \{\alpha\}$ and then taking the equivalence classes under the relation $(p, f, \beta) \backsim (p, \Phi_{\alpha\beta}(p)(f), \alpha)$ so that

$$E = \left(\bigcup U_\alpha \times F \times \{\alpha\}\right) / \backsim$$

and $\pi([p, f, \beta]) = p$.

Let $H \subset G$ be a closed subgroup. Suppose that we can, by throwing out some of the elements of $\{ U_\alpha, \Phi_\alpha\}$ arrange that all of the transition functions live in H. That is, suppose we have that $\Phi_{\alpha\beta}(p)(f) = \rho(g_{\alpha\beta}(p), f)$ where $g_{\alpha\beta} : U_\alpha \cap U_\beta \to H$. Then we have a **reduction the structure group (**or **reduction of the structure representation** in case the action needs to be specified).

Next, suppose that we have an surjective Lie group homomorphism $h : \overline{G} \to G$ **.** We then have the lifted representation $\overline{\rho} : \overline{G} \times F \to F$ given by $\overline{\rho}(\overline{g}, f) = \rho(h(\overline{g}), f)$. We may be able to lift the maps $g_{\alpha\beta}$ to maps $\overline{g}_{\alpha\beta} : U_\alpha \cap U_\beta \to \overline{G}$ and, under suitable topological conditions, by choosing a subfamily we can even arrange that the $\overline{g}_{\alpha\beta}$ satisfy the cocycle condition. Note that $\Phi_{\alpha\beta}(p)(f) = \rho(g_{\alpha\beta}, f) = \rho(h(\overline{g}_{\alpha\beta}), f) = \overline{\rho}(\overline{g}_{\alpha\beta}(p), f)$. In this case we say that we have **lifted** the structure representation to $\overline{\rho}$.

Example 11.1 *As we have said, the simplest class of examples of fiber bundles over a manifold M are the product bundles. These are just Cartesian products $M \times F$ together with the projection map $pr_1 : M \times F \to M$. Here, the structure group can be reduced to the trivial group $\{e\}$ acting as the identity map on F.*

Example 11.2 *A covering manifold $\pi : \widetilde{M} \to M$ is a G-bundle where G is the group of deck transformations. In this example the group G is a discrete (0-dimensional) Lie group.*

Example 11.3 (The Hopf Bundle) *Identify S^1 as the group of complex numbers of unit modulus. Also, we consider the sphere S^3 as it sits in $\mathbb{C}^2$:*

$$S^3 = \{(z_1, z_2) \in \mathbb{C}^2 : |z_1|^2 + |z_2|^2 = 1\}.$$

The group S^1 acts on S^2 by $u \cdot (z_1, z_2) = (uz_1, uz_2)$. Next we get S^2 into the act. We want to realize S^2 as the sphere of radius 1/2 in $\mathbb{R}^3$ and having two *coordinate maps coming from stereographic projection from the north and south poles onto copies of* $\mathbb{C}$ *embedded as planes tangent to the sphere at the two poles. The chart transitions then have the form* $w = 1/z$. *Thus we may view* S^2 *as two copies of* $\mathbb{C}$, *say the* z *plane* $\mathbb{C}_1$ *and the* w *plane* $\mathbb{C}_2$ *glued together under the identification* $\phi : z \mapsto 1/z = w$

$$S^2 = \mathbb{C}_1 \cup_\phi \mathbb{C}_2.$$

With this in mind define a map $\pi : S^3 \subset \mathbb{C}^2 \to S^2$ *by*

$$\pi(z_1, z_2) = \begin{cases} z_2/z_1 \in \mathbb{C}_2 \subset \mathbb{C}_1 \cup_\phi \mathbb{C}_2 & \textit{if} \quad z_1 \neq 0 \\ z_1/z_2 \in \mathbb{C}_1 \subset \mathbb{C}_1 \cup_\phi \mathbb{C}_2 & \textit{if} \quad z_2 \neq 0 \end{cases}.$$

Note that this gives a well defined map onto S^2.

Claim 11.1 $u \cdot (z_1, z_2) = u \cdot (w_1, w_2)$ *if and only if* $\pi(z_1, z_2) = \pi(w_1, w_2)$.

Proof. If $u \cdot (z_1, z_2) = u \cdot (w_1, w_2)$ and $z_1 \neq 0$ then $w_1 \neq 0$ and $\pi(w_1, w_2) = w_2/w_1 = uw_2/uw_1 = \pi(w_1, w_2) = \pi(z_1, z_2) = uz_2/uz_1 = z_2/z_1 = \pi(z_1, z_2)$. A similar calculation show applies when $z_2 \neq 0$. On the other hand, if $\pi(w_1, w_2) = \pi(z_1, z_2)$ then by a similar chain of equalities we also easily get that $u \cdot (w_1, w_2) = ... = \pi(w_1, w_2) = \pi(z_1, z_2) = ... = u \cdot (z_1, z_2)$. ■

Using these facts we see that there is a fiber bundle atlas on $\pi_{Hopf} = \pi : S^3 \to S^2$ given by the following trivializations:

$$\begin{aligned} \varphi_1 &: \pi^{-1}(C_1) \to C_1 \times S^1 \\ \varphi_1 &: (z_1, z_2) = (z_2/z_1, z_1/|z_1|) \end{aligned}$$

and

$$\begin{aligned} \varphi_2 &: \pi^{-1}(C_2) \to C_2 \times S^1 \\ \varphi_2 &: (z_1, z_2) = (z_1/z_2, z_2/|z_2|). \end{aligned}$$

The transition map is

$$(z, u) \mapsto (1/z, \frac{z}{|z|} u)$$

which is of the correct form since $u \mapsto \frac{z}{|z|} \cdot u$ is a circle action.

Chapter 12

Lie Groups II

12.1 Definitions

Definition 12.1 *For a Lie group G and a fixed element $g \in G$, the maps L_g :$G \to G$ and $R_g : G \to G$ are defined by*

$$\begin{aligned} L_g x &= gx \text{ for } x \in G \\ R_g x &= xg \text{ for } x \in G \end{aligned}$$

*and are called **left translation** and **right translation** (by g) respectively.*

Definition 12.2 *A vector field $X \in \mathfrak{X}(G)$ is called **left invariant** if and only if $(L_g)_* X = X$ for all $g \in G$. A vector field $X \in \mathfrak{X}(G)$ is called **right invariant** if and only if $(R_g)_* X = X$ for all $g \in G$. The set of left invariant (resp. right invariant) vectors Fields is denoted $\mathfrak{L}(G)$ or $\mathfrak{X}^L(G)$ (resp. $\mathfrak{R}(G)$ or $\mathfrak{X}^R(G)$).*

Recall that by definition $(L_g)_* X = TL_g \circ X \circ L_g^{-1}$ and so left invariance means that $TL_g \circ X \circ L_g^{-1} = X$ or that $TL_g \circ X = X \circ L_g$. Thus $X \in \mathfrak{X}(G)$ is left invariant if and only if the following diagram commutes

$$\begin{array}{ccc} TG & \overset{TL_g}{\to} & TG \\ X \uparrow & & X \uparrow \\ G & \overset{L_g}{\to} & G \end{array}$$

for every $g \in G$. There is a similar diagram for right invariance.

Remark 12.1 *As we mentioned previously, L_{g*} is sometimes used to mean TL_g and so left invariance of X would then amount to the requirement that for any fixed $p \in G$ we have $L_{g*} X_p = X_{gp}$ for all $g \in G$.*

Lemma 12.1 *$\mathfrak{X}^L(G)$ is closed under the Lie bracket operation.*

Proof. Suppose that $X, Y \in \mathfrak{X}^L(G)$. Then by 7.8 we have

$$\begin{aligned}(L_g)_*[X,Y] &= (L_g)_*(\mathcal{L}_X Y) = \mathcal{L}_{L_{g*}X} L_{g*}Y \\ &= [L_{g*}X, L_{g*}Y] = [X,Y]\end{aligned}$$

■

Corollary 12.1 *$\mathfrak{X}^L(G)$ is an n-dimensional Lie algebra under the bracket of vector fields (see definition* **??***).*

Given a vector $v \in T_eG$ we can define a smooth left (resp. right) invariant vector field L^v (resp. R^v) such that $L^v(e) = v$ (resp. $R^v(e) = v$) by the simple prescription

$$\begin{aligned} L^v(g) &= TL_g \cdot v \\ (\text{resp. } R^v(g) &= \ TR_g \cdot v) \end{aligned}$$

and furthermore the map $v \mapsto L^v$ is a linear isomorphism from T_eG onto $\mathfrak{X}^L(G)$. The proof that this prescription gives invariant vector fields is an easy exercise but one worth doing. Now $\mathfrak{X}^L(G)$ is closed under bracket and forms a Lie algebra. The linear isomorphism just discovered shows that $\mathfrak{X}^L(G)$ is, in fact, a finite dimensional vector space and using the isomorphism we can transfer that Lie algebra structure to T_eG. This is the content of the following

Definition 12.3 *For a Lie group G, define the bracket of any two elements $v, w \in T_eG$ by*

$$[v, w] := [L^v, L^w](e).$$

With this bracket the vector space T_eG becomes a Lie algebra (see definition 7.2) and now have two Lie algebras $\mathfrak{X}^L(G)$ and T_eG. The Lie algebras thus defined are isomorphic by construction and the abstract Lie algebra isomorphic the either/both of them is often referred to as *the* Lie algebra of the Lie group G and denoted by $\mathfrak{Lie}(G)$ or $\mathfrak{g}$. Of course we are implying that $\mathfrak{Lie}(H)$ is denoted $\mathfrak{h}$ and $\mathfrak{Lie}(K)$ by $\mathfrak{k}$ etc. In some computations we will have to use a specific realization of $\mathfrak{g}$. Our default convention will be that $\mathfrak{g} = \mathfrak{Lie}(G) := T_eG$ with the bracket defined above and then occasionally identify this with the left invariant fields $\mathfrak{X}^L(G)$ under the vector field Lie bracket defined in definition **??**.

We shall not take up the detailed study of structure abstract Lie algebras until chapter 15 but we need to anticipate at least one notion:

Definition 12.4 *Given two Lie algebras over a field $\mathbb{F}$, say $(\mathfrak{a}, [,]_\mathfrak{a})$ and $(\mathfrak{b}, [,]_\mathfrak{b})$, an $\mathbb{F}$-linear map σ is called a* ***Lie algebra homomorphism*** *if and only if*

$$\sigma([v,w]_\mathfrak{a}) = [\sigma v, \sigma w]_\mathfrak{b}$$

for all $v, w \in \mathfrak{a}$. A Lie algebra isomorphism is defined in the obvious way. A Lie algebra ***isomorphism*** *$\mathfrak{g} \to \mathfrak{g}$ is called an* ***automorphism*** *of $\mathfrak{g}$.*

It is not hard to show that the set of all automorphisms of $\mathfrak{g}$, denoted $\mathrm{Aut}(\mathfrak{g})$, forms a Lie group (actually a Lie subgroup of $\mathrm{GL}(\mathfrak{g})$).

Define the maps $\omega_G : TG \to \mathfrak{g}$ (resp. $\omega_G^{right} : TG \to \mathfrak{g}$) by

$$\omega_G(X_g) = TL_{g^{-1}} \cdot X_g$$
$$(\text{resp. } \omega_G^{right}(X_g) = TR_{g^{-1}} \cdot X_g).$$

ω_G is a $\mathfrak{g}$ valued 1-form called the (left-) **Maurer-Cartan form**. We will call ω_G^{right} the right-Maurer-Cartan form but we will not be using it to the extent of ω_G.

Recall that if V is a finite dimensional vector space then each tangent space $T_x\mathrm{V}$ is naturally isomorphic to V. Now $G\mathrm{L}(n)$ is an open subset of the vector space of $n \times n$ real matrices $\mathbb{M}_{n\times n}$ and so we obtain natural vector space isomorphisms $T_g G\mathrm{L}(n) \cong \mathbb{M}_{n\times n}$ for all $g \in G\mathrm{L}(n)$. To move to the level of bundles we reconstitute these isomorphisms to get maps $T_g G\mathrm{L}(n) \to \{g\} \times \mathbb{M}_{n\times n}$ which we can then bundle together to get a trivialization $T\,G\mathrm{L}(n) \to G\mathrm{L}(n) \times \mathbb{M}_{n\times n}$. One could use this trivialization to identify $T\,G\mathrm{L}(n)$ with $G\mathrm{L}(n) \times \mathbb{M}_{n\times n}$ and this trivialization is just a special case of the general situation: When U is an open subset of a vector space V, we have a trivialization[1] $T\mathrm{V} \cong U \times \mathrm{V}$. Further on, we introduce two more trivializations of $T\,G\mathrm{L}(n) \cong G\mathrm{L}(n) \times \mathbb{M}_{n\times n}$ defined using the (left or right) Maurer-Cartan form. This will work for general Lie groups. Since these trivializations could also be used as identifying isomorphisms we had better introduce a bit of nonstandard terminology. Let us refer to the identification of $T\,G\mathrm{L}(n)$ with $G\mathrm{L}(n) \times \mathbb{M}_{n\times n}$, or more generally $T\mathrm{V}$ with $U \times \mathrm{V}$ as the canonical identification

Let us explicitly recall one way to describe the isomorphism $T_g G\mathrm{L}(n) \cong \mathbb{M}_{n\times n}$. If $v_g \in T_g G$ then there is some curve (of matrices) $c^g : t \mapsto c(t)$ such that $c^g(0) = g$ and $\dot{c}^g(0) = v_g \in T_g G$. By definition $\dot{c}^g(0) := T_0 c^g \cdot \frac{d}{dt}\big|_0$ which is based at g. If we just take the ordinary derivative we get the matrix that represents v_g:

$$\text{If } c(t) = \begin{bmatrix} g_1^1(t) & g_1^2(t) & \cdots \\ g_2^1(t) & \ddots & \\ \vdots & & g_n^n(t) \end{bmatrix}$$

then $\dot{c}(0) = v_g$ is identified with the matrix

$$a := \begin{bmatrix} \frac{d}{dt}\big|_{t=0} g_1^1 & \frac{d}{dt}\big|_{t=0} g_1^2 & \cdots \\ \frac{d}{dt}\big|_{t=0} g_2^1 & \ddots & \\ \vdots & & \frac{d}{dt}\big|_{t=0} g_n^n \end{bmatrix}.$$

As a particular case, we have a vector space isomorphism $\mathfrak{gl}(n) = T_I G\mathrm{L}(n) \cong \mathbb{M}_{n\times n}$ where I is the identity matrix in $G\mathrm{L}(n)$. This we want to use to identify $\mathfrak{gl}(n)$ with $\mathbb{M}_{n\times n}$. Now $\mathfrak{gl}(n) = T_I G\mathrm{L}(n)$ has a Lie algebra structure and we

[1] Actually, $U \times \mathrm{V}$ was our first *definition* of the tangent bundle of a vector space.

would like to transfer the Lie bracket from $\mathfrak{gl}(n)$ to $\mathbb{M}_{n\times n}$ is such a way that this isomorphism becomes a Lie algebra isomorphism. It turns out that the Lie bracket that we end up with for $\mathbb{M}_{n\times n}$ is the commutator bracket defined by $[A, B] := AB - BA$. This turns out to be so natural that we can safely identify the Lie algebra $\mathfrak{gl}(n)$ with $\mathbb{M}_{n\times n}$. Along these lines we will also be able to identify the Lie algebras of subgroups of $G\mathrm{L}(n)$ (or $G\mathrm{L}(n, \mathbb{C})$) with linear subspaces of $\mathbb{M}_{n\times n}$ (or $\mathbb{M}_{n\times n}(\mathbb{C})$).

Initially, the Lie algebra of is $G\mathrm{L}(n)$ is given as the space of left invariant vectors fields on $GL(n)$. The bracket is the bracket of vector fields. This bracket transferred to $T_I G\mathrm{L}(n)$ according to the isomorphism of $\mathfrak{X}(GL(n))$ with $T_I GL(n)$ is given in this case by $X \mapsto X(I)$. The situation is that we have two Lie algebra isomorphisms

$$\mathfrak{X}(\mathrm{GL}(n)) \cong T_I G\mathrm{L}(n) \cong \mathbb{M}_{n\times n}$$

and the origin of all of the Lie algebra structure is the one on $\mathfrak{X}(GL(n))$. The plan is then to figure out what is the left invariant vector field that corresponds to a given matrix from $\mathbb{M}_{n\times n}$. This gives a direct route between $\mathfrak{X}(GL(n))$ and $\mathbb{M}_{n\times n}$ allowing us to see what the correct bracket on $\mathbb{M}_{n\times n}$ should be. Note that a global coordinate system for $GL(n)$ is given by the maps x_l^k which are defined by the requirement that $x_l^k(A) = a_l^k$ whenever $A = (a_j^i)$. Thus any vector fields $X, Y \in \mathfrak{X}(GL(n))$ can be written

$$X = f_j^i \frac{\partial}{\partial x_j^i}$$
$$Y = g_j^i \frac{\partial}{\partial x_j^i}$$

for some functions f_j^i and g_j^i. Now let $(a_j^i) = A \in \mathbb{M}_{n\times n}$. The corresponding element of $T_I G\mathrm{L}(n)$ can be written $v^A = a_j^i \left.\frac{\partial}{\partial x_j^i}\right|_I$. Corresponding to v_A there is a left invariant vector field X^A which is given by $X^A(x) = T_I L_x \cdot v^A$ which in turn corresponds to the matrix $\mathrm{I}\left.\frac{d}{dt}\right|_0 xc(t) = xA$. Thus X^A is given by $X = f_j^i \frac{\partial}{\partial x_j^i}$ where $f_j^i(x) = xA = \left(x_k^i a_j^k\right)$. Similarly, let B be another matrix with corresponding left invariant vector field X^B. The bracket $[X^A, X^B]$ can now be computed as follows:

$$\begin{aligned}
[X^A, X^B] &= (f_j^i \frac{\partial g_l^k}{\partial x_j^i} - g_j^i \frac{\partial f_l^k}{\partial x_j^i}) \frac{\partial}{\partial x_l^k} \\
&= \left(x_r^i a_j^r \frac{\partial (x_s^k b_l^s)}{\partial x_j^i} - x_r^i b_j^r \frac{\partial (x_s^k a_l^s)}{\partial x_j^i} \right) \frac{\partial}{\partial x_l^k} \\
&= \left(x_r^k a_s^r b_l^s - x_r^k b_s^r a_l^s \right) \frac{\partial}{\partial x_l^k}
\end{aligned}$$

Evaluating at $I = (\delta_i^i)$ we have

$$\begin{aligned}
[X^A, X^B](I) &= \left(\delta_r^k a_s^r b_l^s - \delta_r^k b_s^r a_l^s\right) \left.\frac{\partial}{\partial x_l^k}\right|_I \\
&= \left(a_s^k b_l^s - b_s^k a_l^s\right) \left.\frac{\partial}{\partial x_l^k}\right|_I
\end{aligned}$$

which corresponds to the matrix $AB - BA$. Thus the proper Lie algebra structure on $\mathbb{M}_{n\times n}$ is given by the commutator $[A, B] = AB - BA$. In summary, we have

Proposition 12.1 *Under the identification of* $\mathfrak{gl}(n) = T_I G\mathrm{L}(n)$ *with* $\mathbb{M}_{n\times n}$ *the bracket is the commutator bracket*

$$[A, B] = AB - BA$$

Similarly, under the identification of $T_{\mathrm{id}}\mathrm{GL}(\mathrm{V})$ *with* $\mathrm{End}(\mathrm{V})$ *the bracket is*

$$[A, B] = A \circ B - B \circ A.$$

If $G \subset G\mathrm{L}(n)$ is some matrix group then $T_I G$ may be identified with a subspace of $\mathbb{M}_{n\times n}$ and it turns out that this linear subspace will be closed under the commutator bracket and so we actually have an identification of Lie algebras: $\mathfrak{g}$ is identified with a subspace of $\mathbb{M}_{n\times n}$. It is often the case that G is defined by some matrix equation(s). By differentiating this equation we find the defining equation(s) for $\mathfrak{g}$ (the subspace of $\mathbb{M}_{n\times n}$). We demonstrate this with a couple of examples.

Example 12.1 *Consider the orthogonal group* $\mathrm{O}(n) \subset G\mathrm{L}(n)$. *Given a curve of orthogonal matrices* $Q(t)$ *with* $Q(0) = 1$ *and* $\frac{d}{dt}\big|_{t=0} Q(0) = A$ *we compute by differentiating the defining equation* $I = Q^t Q$.

$$\begin{aligned}
0 &= \left.\frac{d}{dt}\right|_{t=0} Q^t Q \\
&= \left(\left.\frac{d}{dt}\right|_{t=0} Q\right)^t Q(0) + Q^t(0)\left(\left.\frac{d}{dt}\right|_{t=0} Q\right) \\
&= A^t + A
\end{aligned}$$

This means that we can identify $\mathfrak{o}(n)$ *with the space of skewsymmetric matrices. One can easily check that the commutator bracket of two such matrices is skewsymmetric as expected.*

We have considered matrix groups as subgroups of $G\mathrm{L}(n)$ but it often more convenient to consider subgroups of $G\mathrm{L}(n, \mathbb{C})$. Since $G\mathrm{L}(n, \mathbb{C})$ can be identified with a subgroup of $G\mathrm{L}(2n)$ this is only a slight change in viewpoint. The essential parts of our discussion go through for $G\mathrm{L}(n, \mathbb{C})$ without any essential change.

Example 12.2 *Consider the unitary group $U(n) \subset G\mathrm{L}(n, \mathbb{C})$. Given a curve of unitary matrices $Q(t)$ with $Q(0) = I$ and $\frac{d}{dt}\big|_{t=0} Q(0) = A$ we compute by differentiating the defining equation $I = \bar{Q}^t Q$.*

$$\begin{aligned} 0 &= \left.\frac{d}{dt}\right|_{t=0} \bar{Q}^t Q \\ &= \left(\left.\frac{d}{dt}\right|_{t=0} \bar{Q}\right)^t Q(0) + \bar{Q}^t(0)\left(\left.\frac{d}{dt}\right|_{t=0} Q\right) \\ &= \bar{A}^t + A \end{aligned}$$

This means that we can identify $\mathfrak{u}(n)$ with the space of skewhermitian matrices ($\bar{A}^t = -A$). One can easily check that $\mathfrak{u}(n)$ is closed under the commutator bracket.

As we have seen, $G\mathrm{L}(n)$ is an open set in the vector space and so its tangent bundle is trivial $TG\mathrm{L}(n) \cong G\mathrm{L}(n) \times \mathbb{M}_{n\times n}$. A general abstract Lie group is not an open subset of a vector space but we are still able to show that $TG\mathrm{L}(n)$ is trivial. There are two such trivializations obtained from the Maurer-Cartan forms. These are $\mathrm{triv}_L : TG \to G \times \mathfrak{g}$ and $\mathrm{triv}_R : TG \to G \times \mathfrak{g}$ and defined by

$$\begin{aligned} \mathrm{triv}_L(v_g) &= (g, \omega_G(v_g)) \\ \mathrm{triv}_R(v_g) &= (g, \omega_G^{right}(v_g)) \end{aligned}$$

for $v_g \in T_g G$. These maps are both vector bundle isomorphisms, that is, trivializations. Thus we have the following:

Proposition 12.2 *The tangent bundle of a Lie group is trivial: $TG \cong G \times \mathfrak{g}$.*

Proof. It is easy to check that triv_L and triv_R are vector bundle isomorphisms. ■

We will refer to triv_L and triv_R as the (left and right) **Maurer-Cartan trivializations**. How do these two trivializations compare? There is no special reason to prefer right multiplication. We could have used right invariant vector fields as our means of producing the Lie algebra and the whole theory would work 'on the other side' so to speak. What is the relation between left and right in this context? The bridge between left and right is the adjoint map.

Lemma 12.2 (Left-right lemma) *For any $v \in \mathfrak{g}$ the map $g \mapsto \mathrm{triv}_L^{-1}(g, v)$ is a left invariant vector field on G while $g \mapsto \mathrm{triv}_R^{-1}(g, v)$ is right invariant. Also, $\mathrm{triv}_R \circ \mathrm{triv}_L^{-1}(g, v) = (g, \mathrm{Ad}_g(v))$.*

Proof. The invariance is easy to check and is left as an exercise. Now the second statement is also easy:

$$\begin{aligned} &\mathrm{triv}_R \circ \mathrm{triv}_L^{-1}(g, v) \\ &= (g, TR_{g^{-1}} TL_g v) = (g, T(R_{g^{-1}} L_g) \cdot v) \\ &= (g, \mathrm{Ad}_g(v)). \end{aligned}$$

■

It is often convenient to actually identify the tangent bundle TG of a Lie group G with $G \times \mathfrak{g}$. Of course we must specify which of the two trivializations described above is being invoked. Unless indicated otherwise we shall use the "left version" described above: $v_g \mapsto (g, \omega_G(v_g)) = (g, TL_g^{-1}(v_g))$. It must be realized that we now have three natural ways to trivialize the tangent bundle of the general linear group. In fact, the usual one which we introduced earlier is actually the restriction to $TG\mathrm{L}(n)$ of the Maurer-Cartan trivialization of the abelian Lie group $(\mathbb{M}_{n\times n}, +)$.

In order to use the (left) Maurer-Cartan trivialization as an identification effectively, we need to find out how a few basic operations look when this identification is imposed.

The picture obtained from using the trivialization produced by the left Maurer-Cartan form:

1. The tangent map of left translation $TL_g : TG \to TG$ takes the form "TL_g" $: (x, v) \mapsto (gx, v)$. Indeed, the following diagram commutes:

$$\begin{array}{ccc} TG & \overset{TL_g}{\rightarrow} & TG \\ \downarrow & & \downarrow \\ G \times \mathfrak{g} & \overset{\text{“}TL_g\text{”}}{\rightarrow} & G \times \mathfrak{g} \end{array}$$

   where elementwise we have

$$\begin{array}{ccc} v_x & \overset{TL_g}{\rightarrow} & TL_g \cdot v_x \\ \downarrow & & \downarrow \\ \left[\begin{array}{c} (x, TL_x^{-1} v_x) \\ = (x, v) \end{array}\right] & \rightarrow & \left[\begin{array}{c} (gx, TL_{gx}^{-1} TL_g v_x) = (gx, TL_x^{-1} v_x) \\ = (gx, v) \end{array}\right] \end{array}$$

2. The tangent map of multiplication: This time we will invoke two identifications. First, group multiplication is a map $\mu : G \times G \to G$ and so on the tangent level we have a map $T(G \times G) \to G$. Now recall that we have a natural isomorphism $T(G \times G) \cong TG \times TG$ given by $T\pi_1 \times T\pi_2 : (v_{(x,y)}) \mapsto (T\pi_1 \cdot v_{(x,y)}, T\pi_2 \cdot v_{(x,y)})$. If we also identify TG with $G \times \mathfrak{g}$ then $TG \times TG \cong (G \times \mathfrak{g}) \times (G \times \mathfrak{g})$ we end up with the following "version" of $T\mu$:

$$\begin{aligned} \text{“}T\mu\text{”} &: (G \times \mathfrak{g}) \times (G \times \mathfrak{g}) \to G \times \mathfrak{g} \\ \text{“}T\mu\text{”} &: ((x, v), (y, w)) \mapsto (xy, TR_y v + TL_x w) \end{aligned}$$

   (See exercise 12.1).

3. The (left) Maurer-Cartan form is a map $\omega_G : TG \to T_eG = \mathfrak{g}$ and so there must be a "version", "ω_G", that uses the identification $TG \cong G \times \mathfrak{g}$. In fact, the map we seek is just projection:

$$\text{“}\omega_G\text{”} : (x, v) \mapsto v$$

4. The right Maurer-Cartan form is a little more complicated since we are using the isomorphism $TG \cong G \times \mathfrak{g}$ obtained from the *left* Maurer-Cartan form. The result follows from the "left-right lemma 12.2:

$$\text{“}\omega_G^{right}\text{”} : (x, v) \mapsto Ad_g(v)$$

   The adjoint map is nearly the same thing as the right Maurer-Cartan form once we decide to use the (left) trivialization $TG \cong G \times \mathfrak{g}$ as an identification.

5. A vector field $X \in \mathfrak{X}(G)$ should correspond to a section of the product bundle $G \times \mathfrak{g} \rightarrow G$ which must have the form $\overleftrightarrow{X} : x \mapsto (x, F^X(x))$ for some smooth $\mathfrak{g}$-valued function $f^X \in C^\infty(G; \mathfrak{g})$. It is an easy consequence of the definitions that $F^X(x) = \omega_G(X(x)) = TL_x^{-1} \cdot X(x)$. We ought to be able to identify how $\overleftrightarrow{X}$ acts on $C^\infty(G)$. We have $(\overleftrightarrow{X} f)(x) = (x, F^X(x))f = (TL_x \cdot F^X(x))f$

   A left invariant vector field becomes a constant section of $G \times \mathfrak{g}$. For example, if X is left invariant then the corresponding constant section is $x \mapsto (x, X(e))$.

Exercise 12.1 *Refer to 2. Show that the map "$T\mu$" defined so that the diagram below commutes is* $((x, v), (y, w)) \mapsto (xy, TR_y v + TL_x w)$.

$$\begin{array}{ccc} T(G \times G) & \overset{T\mu}{\rightarrow} & TG \\ \cong TG \times TG & & \\ \downarrow & & \downarrow \\ (G \times \mathfrak{g}) \times (G \times \mathfrak{g}) & \overset{\text{“}T\mu\text{”}}{\rightarrow} & G \times \mathfrak{g} \end{array}$$

Proposition 12.3 *Let $h : G_1 \rightarrow G_2$ be a Lie group homomorphism. The map $T_e h : \mathfrak{g}_1 \rightarrow \mathfrak{g}_2$ is a Lie algebra homomorphism called the* ***Lie differential*** *which is often denoted in this context by $dh : \mathfrak{g}_1 \rightarrow \mathfrak{g}_2$ or by $Lh : \mathfrak{g}_1 \rightarrow \mathfrak{g}_2$.*

Proof. For $v \in \mathfrak{g}_1$ and $x \in G$ we have

$$\begin{aligned} T_x h \cdot L^v(x) &= T_x h \cdot (T_e L_x \cdot v) \\ &= T_e(h \circ L_x) \cdot v \\ &= T_e(L_{h(x)} \circ h) \cdot v \\ &= T_e L_{h(x)}(T_e h \cdot v) \\ &= L^{dh(v)}(h(x)) \end{aligned}$$

so $L^v \backsim_h L^{dh(v)}$. Thus by 7.8 we have for any $v, w \in \mathfrak{g}_1$ that $L^{[v,w]} \backsim_h [L^{dh(v)}, L^{dh(w)}]$ or in other words, $[L^{dh(v)}, L^{dh(w)}] \circ h = Th \circ L^{[v,w]}$ which at e gives

$$[dh(v), dh(w)] = [v, w].$$

■

Theorem 12.1 *Invariant vector fields are complete. The integral curves through the identify element are the one-parameter subgroups.*

Proof. Let X be a left invariant vector field and $c : (a, b) \to G$ be the integral curve of X with $\dot{c}(0) = X(p)$. Let $a < t_1 < t_2 < b$ and choose an element $g \in G$ such that $gc(t_1) = c(t_2)$. Let $\Delta t = t_2 - t_1$ and define $\bar{c} : (a + \Delta t, b + \Delta t) \to G$ by $\bar{c}(t) = gc(t - \Delta t)$. Then we have

$$\begin{aligned}\bar{c}'(t) &= TL_g \cdot c'(t - \Delta t) = TL_g \cdot X(c(t - \Delta t)) \\ &= X(gc(t - \Delta t)) = X(\bar{c}(t))\end{aligned}$$

and so $\bar{c}$ is also an integral curve of X. Now on the intersection $(a + \Delta t, b)$ of the their domains, c and $\bar{c}$ are equal since they are both integral curve of the same field and since $\bar{c}(t_2) = gc(t_1) = c(t_2)$. Thus we can concatenate the curves to get a new integral curve defined on the larger domain $(a, b + \Delta t)$. Since this extension can be done again for a fixed Δt we see that c can be extended to (a, ∞). A similar argument gives that we can extend in the negative direction to get the needed extension of c to $(-\infty, \infty)$.

Next assume that c is the integral curve with $c(0) = e$. The proof that $c(s+t) = c(s)c(t)$ proceeds by considering $\gamma(t) = c(s)^{-1}c(s+t)$. Then $\gamma(0) = e$ and also

$$\begin{aligned}\gamma'(t) &= TL_{c(s)^{-1}} \cdot c'(s + t) = TL_{c(s)^{-1}} \cdot X(c(s + t)) \\ &= X(c(s)^{-1}c(s + t)) = X(\gamma(t)).\end{aligned}$$

By the uniqueness of integral curves we must have $c(s)^{-1}c(s + t) = c(t)$ which implies the result. Conversely, suppose $c : \mathbb{R} \to G$ is a one parameter subgroup and let $X_e = \dot{c}(0)$. There is a left invariant vector field X such that $X(e) = X_e$, namely, $X = L^{X_e}$. We must show that the integral curve through e of the field X is exactly c. But for this we only need that $\dot{c}(t) = X(c(t))$ for all t. Now $c(t + s) = c(t)c(s)$ or $c(t + s) = L_{c(t)}c(s)$. Thus

$$\dot{c}(t) = \left.\frac{d}{ds}\right|_0 c(t + s) = (T_{c(t)}L).\dot{c}(0) = X(c(t))$$

and we are done. ∎

Lemma 12.3 *Let $v \in \mathfrak{g} = T_eG$ and the corresponding left invariant field L^v. Then with $Fl^v := Fl^{L^v}$ we have that*

$$Fl^v(st) = Fl^{sv}(t) \tag{12.1}$$

A similar statement holds with R^v replacing L^v.

Proof. Let $u = st$. We have that $\left.\frac{d}{dt}\right|_{t=0} Fl^v(st) = \frac{du}{dt} \left.\frac{d}{du}\right|_{t=0} Fl^v(u)\frac{du}{dt} = sv$ and so by uniqueness $Fl^v(st) = Fl^{sv}(t)$. ∎

Definition 12.5 *For any $v \in \mathfrak{g} = T_eG$ we have the corresponding left invariant field L^v which has an integral curve through e that we denote by* $\exp(tv)$*. Thus the map $t \to \exp(tv)$ is a Lie group homomorphism from $\mathbb{R}$ into G that is a one-parameter subgroup. The map $v \mapsto \exp(1v) = \exp^G(v)$ is referred to as the* ***exponential map*** $\exp^G : \mathfrak{g} \to G$.

Lemma 12.4 *The map $\exp^G : \mathfrak{g} \to G$ is smooth.*

Proof. Consider the map $\mathbb{R} \times G \times \mathfrak{g} \to G \times \mathfrak{g}$ given by $(t, g, v) \mapsto (g \cdot \exp^G(tv), v)$. This map is easily seen to be the flow on $G \times \mathfrak{g}$ of the vector field $\widetilde{X} : (g, v) \mapsto (L^v(g), 0)$ and so is smooth. Now the restriction of this smooth flow to the submanifold $\{1\} \times \{e\} \times \mathfrak{g}$ is $(1, e, v) \mapsto (\exp^G(v), v)$ is also smooth, which clearly implies that $\exp^G(v)$ is smooth also. ∎

Theorem 12.2 *The tangent map of the exponential map $\exp^G : \mathfrak{g} \to G$ is the identity at the origin $0 \in T_eG = \mathfrak{g}$ and* exp *is a diffeomorphism of some neighborhood of the origin onto its image in G.*

Proof. By lemma 12.4 we know that $\exp^G : \mathfrak{g} \to G$ is a smooth map. Also, $\frac{d}{dt}\big|_0 \exp^G(tv) = v$ which means the tangent map is $v \mapsto v$. If the reader thinks through the definitions carefully, he or she will discover that we have here used the natural identification of $\mathfrak{g}$ with $T_0\mathfrak{g}$. ∎

Remark 12.2 *The "one-parameter subgroup" $\exp^G(tv)$ corresponding to a vector $v \in \mathfrak{g}$ is actually a homomorphism rather than a subgroup but the terminology is conventional.*

Proposition 12.4 *For a (Lie group) homomorphism $h : G_1 \to G_2$ the following diagram commutes:*

$$
\begin{array}{ccc}
\mathfrak{g}_1 & \xrightarrow{dh} & \mathfrak{g}_2 \\
\exp^{G_1} \downarrow & & \exp^{G_2} \downarrow \\
G_1 & \xrightarrow{h} & G_2
\end{array}
$$

Proof. The curve $t \mapsto h(\exp^{G_1}(tv))$ is clearly a one parameter subgroup. Also,

$$\left.\frac{d}{dt}\right|_0 h(\exp^{G_1}(tv)) = dh(v)$$

so by uniqueness of integral curves $h(\exp^{G_1}(tv)) = \exp^{G_2}(tdh(v))$. ∎

Remark 12.3 *We will sometimes not index the maps and shall just write* exp *for any Lie group.*

The reader may wonder what happened to right invariant vector fields and how do they relate to one parameter subgroups. The following theorem give various relationships.

Theorem 12.3 *For a smooth curve $c : \mathbb{R} \to G$ with $c(0) = e$ and $\dot{c}(0) = v$, the following are all equivalent:*

1. $c(t) = Fl_t^{L^v}(e)$
2. $c(t) = Fl_t^{R^v}(e)$
3. *c is a one parameter subgroup.*
4. $Fl_t^{L^v} = R_{c(t)}$
5. $Fl_t^{R^v} = L_{c(t)}$

Proof. By definition $Fl_t^{L^v}(e) = \exp(tv)$. We have already shown that 1 implies 3. The proof of 2 implies 3 would be analogous. We have also already shown that 3 implies 1.

Also, 4 implies 1 since then $Fl_t^{L^v}(e) = R_{c(t)}(e) = c(t)$. Now assuming 1 we have

$$\begin{aligned}
c(t) &= Fl_t^{L^v}(e) \\
\left.\frac{d}{dt}\right|_0 c(t) &= L^v(e) \\
\left.\frac{d}{dt}\right|_0 gc(t) &= \left.\frac{d}{dt}\right|_0 L_g(c(t)) \\
&= TL_g v = L^v(g) \text{ for any } g \\
\left.\frac{d}{dt}\right|_0 R_{c(t)} g &= L^v(g) \text{ for any } g \\
R_{c(t)} &= Fl_t^{L^v}
\end{aligned}$$

The rest is left to the reader. ∎

The tangent space at the identity of a Lie group G is a vector space and is hence a manifold. Thus exp is a smooth map between manifolds. As is usual we identify the tangent space $T_v(T_eG)$ at some $v \in T_eG$ with T_eG itself. The we have the following

Lemma 12.5 $T_e \exp = \mathrm{id} : T_eG \to T_eG$

Proof. $T_e \exp \cdot v = \left.\frac{d}{dt}\right|_0 \exp(tv) = v$. ∎

The Lie algebra of a Lie group and the group itself are closely related in many ways. One observation is the following:

Proposition 12.5 *If G is a connected Lie group then for any open neighborhood $V \subset \mathfrak{g}$ of 0 the group generated by $\exp(V)$ is all of G.*

sketch of proof. Since $T_e \exp = \mathrm{id}$ we have that exp is an open map near 0. The group generated by $\exp(V)$ is a subgroup containing an open neighborhood of e. The complement is also open. ∎

Now we prove a remarkable theorem that shows how an algebraic assumption can have implications in the differentiable category. First we need some notation.

Notation 12.1 *If S is any subset of a Lie group G then we define*

$$S^{-1} = \{s^{-1} : s \in S\}$$

and for any $x \in G$ we define

$$xS = \{xs : s \in S\}.$$

Theorem 12.4 *An abstract subgroup H of a Lie group G is a (regular) submanifold if and only if H is a closed set in G. If follows that H is a (regular) Lie subgroup of G.*

Proof. First suppose that H is a (regular) submanifold. Then H is locally closed. That is, every point $x \in H$ has an open neighborhood U such that $U \cap H$ is a relatively closed set in H. Let U be such a neighborhood of the identity element e. We seek to show that H is closed in G. Let $y \in \overline{H}$ and $x \in yU^{-1} \cap H$. Thus $x \in H$ and $y \in xU$. Now this means that $y \in \overline{H} \cap xU$, and thence $x^{-1}y \in \overline{H} \cap U = H \cap U$. So $y \in H$ and we have shown that H is closed.

Now conversely, let us suppose that H is a closed abstract subgroup of G. Since we can always use the diffeomorphism to translate any point to the identity it suffices to find a neighborhood U of e such that $U \cap H$ is a submanifold. The strategy is to find out what $\mathrm{Lie}(H) = \mathfrak{h}$ is likely to be and then exponentiate a neighborhood of $e \in \mathfrak{h}$.

First we will need to have an inner product on T_eG so choose any such. Then norms of vectors in T_eG makes sense. Choose a small neighborhood $\widetilde{U}$ of $0 \in T_eG = \mathfrak{g}$ on which exp is a diffeomorphism say $\exp : \widetilde{U} \to U$ with inverse denoted by $\log_U$. Define the set $\widetilde{H}$ in $\widetilde{U}$ by $\widetilde{H} = \log_U(H \cap U)$.

Claim 12.1 *If h_n is a sequence in $\widetilde{H}$ converging to zero and such that $u_n = h_n / |h_n|$ converges to $v \in \mathfrak{g}$ then $\exp(tv) \in H$ for all $t \in \mathbb{R}$.*

Proof of claim: Note that $th_n/|h_n| \to tv$ while $|h_n|$ converges to zero. But since $|h_n| \to 0$ we must be able to find a sequence $k(n) \in \mathbb{Z}$ such that $k(n)|h_n| \to t$. From this we have $\exp(k(n)h_n) = \exp(k(n)|h_n| \frac{h_n}{|h_n|}) \to \exp(tv)$. But by the properties of exp *proved previously we have $\exp(k(n)h_n) = (\exp(h_n))^{k(n)}$. But $\exp(h_n) \in H \cap U \subset H$ and so $(\exp(h_n))^{k(n)} \in H$. But since H is closed we have $\exp(tv) = \lim_{n\to\infty}(\exp(h_n))^{k(n)} \in H$.*

Claim 12.2 *The set W of all tv where v can be obtained as a limit $h_n/|h_n| \to v$ with $h_n \in \widetilde{H}$ is a vector space.*

Proof of claim: It is enough to show that if $h_n/|h_n| \to v$ and $h'_n/|h'_n| \to w$ with $h'_n, h_n \in \widetilde{H}$ then there is a sequence of elements h''_n from $\widetilde{H}$ with

$$h''_n/|h''_n| \to \frac{v+w}{|v+w|}.$$

This will follow from the observation that

$$h(t) = \log_U(\exp(tv)\exp(tw))$$

is in $\widetilde{H}$ *and by exercise 5.3 we have that*

$$\lim_{t \downarrow 0} h(t)/t = v + w$$

and so

$$\frac{h(t)/t}{|h(t)/t|} \to \frac{v+w}{|v+w|}.$$

The proof of the next claim will finish the proof of the theorem.

Claim 12.3 *Let W be the set from the last claim. Then* $\exp(W)$ *contains an open neighborhood of e in H. Let $W^\perp$ be the orthogonal compliment of W with respect to the inner product chosen above. Then we have $T_eG = W^\perp \oplus W$. It is not difficult to show that the map $\Sigma : W \oplus W^\perp \to G$ defined by*

$$v + w \mapsto \exp(v)\exp(w)$$

is a diffeomorphism in a neighborhood of the origin in T_eG. Now suppose that $\exp(W)$ *does not contain an open neighborhood of e in H. Then we can choose a sequence $(v_n, w_n) \in W \oplus W^\perp$ with $(v_n, w_n) \to 0$ and* $\exp(v_n)\exp(w_n) \in H$ *and yet $w_n \neq 0$. The space $W^\perp$ and the unit sphere in $W^\perp$ is compact so after passing to a subsequence we may assume that $w_n/|w_n| \to w \in W^\perp$ and of course $|w| = 1$. Since* $\exp(v_n) \in H$ *and H is at least an algebraic subgroup,* $\exp(v_n)\exp(w_n) \in H$*, it must be that* $\exp(w_n) \in H$ *also. But then by the definition of W we have that $w \in W$ which contradicts the fact that $|w| = 1$ and $w \in W^\perp$.*

■

12.2 Spinors and rotation

The matrix Lie group $SO(3)$ is the group of orientation preserving rotations of $\mathbb{R}^3$ acting by matrix multiplication on column vectors. The group $SU(2)$ is the group of complex 2×2 unitary matrices of determinant 1. We shall now expose an interesting relation between these groups. First recall the Pauli matrices:

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad , \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \quad \sigma_2 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} \quad , \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

The real vector space spanned by $\sigma_1, \sigma_2, \sigma_3$ is isomorphic to $\mathbb{R}^3$ and is the space of traceless Hermitian matrices. Let us temporarily denote the latter by $\widehat{\mathbb{R}^3}$. Thus we have a linear isomorphism $\mathbb{R}^3 \to \widehat{\mathbb{R}^3}$ given by $(x^1, x^2, x^3) \mapsto x^1\sigma_1 + x^2\sigma_2 + x^3\sigma_3$ which we abbreviate to $\vec{x} \mapsto \widehat{x}$. Now it is easy to check that $\det(\widehat{x})$ is just $-|\vec{x}|^2$. In fact, we may introduce an inner product on $\widehat{\mathbb{R}^3}$ by the formula $\langle \widehat{x}, \widehat{y} \rangle := -\frac{1}{2}tr(\widehat{x}\widehat{y})$ and then we have that $\vec{x} \mapsto \widehat{x}$ is an isometry. Next we notice that $SU(2)$ acts on $\widehat{\mathbb{R}^3}$ by $(g, \widehat{x}) \mapsto g\widehat{x}g^{-1} = g\widehat{x}g^*$ thus giving a representation ρ of $SU(2)$ in $\widehat{\mathbb{R}^3}$. It is easy to see that $\langle \rho(g)\widehat{x}, \rho(g)\widehat{y} \rangle = \langle \widehat{x}, \widehat{y} \rangle$ and so under the identification $\mathbb{R}^3 \leftrightarrow \widehat{\mathbb{R}^3}$ we see that $SU(2)$ act on $\mathbb{R}^3$ as an element of $O(3)$.

Exercise 12.2 *Show that in fact, the map $SU(2) \to O(3)$ has image $S\mathrm{O}(3)$ is actually a covering homomorphism onto $S\mathrm{O}(3)$ with kernel $\{\pm I\} \cong \mathbb{Z}_2$.*

Exercise 12.3 *Show that the algebra generated by the matrices $\sigma_0, i\sigma_1, i\sigma_2, i\sigma_3$ is isomorphic to the quaternion algebra and that the set of matrices $i\sigma_1, i\sigma_2, i\sigma_3$ span a real vector space which is equal as a set to the traceless skew Hermitian matrices $\mathfrak{su}(2)$.*

Let $I = i\sigma_1$, $J = i\sigma_2$ and $i\sigma_3 = K$. One can redo the above analysis using the isometry $\mathbb{R}^3 \to \mathfrak{su}(2)$ given by

$$
\begin{aligned}
(x^1, x^2, x^3) &\mapsto x^1 I + x^2 J + x^3 K \\
\vec{x} &\mapsto \widetilde{x}
\end{aligned}
$$

where this time $\langle \widetilde{x}, \widetilde{y} \rangle := \frac{1}{2}tr(\widetilde{x}\widetilde{y}^*) = -\frac{1}{2}tr(\widetilde{x}\widetilde{y})$. Notice that $\mathfrak{su}(2) = span\{I, J, K\}$ is the Lie algebra of $SU(2)$ the action $(g, \widehat{x}) \mapsto g\widehat{x}g^{-1} = g\widehat{x}g^*$ is just the adjoint action to be defined in a more general setting below. Anticipating this, let us write $Ad(g) : \widehat{x} \mapsto g\widehat{x}g^*$. This gives the map $g \mapsto Ad(g)$; a Lie group homomorphism $SU(2) \to S\mathrm{O}(\mathfrak{su}(2), \langle , \rangle)$. Once again we get the same map $SU(2) \to \mathrm{O}(3)$ which is a Lie group homomorphism and has kernel $\{\pm I\} \cong \mathbb{Z}_2$. In fact, we have the following commutative diagram:

$$
\begin{array}{ccc}
SU(2) & = & SU(2) \\
\rho \downarrow & & Ad \downarrow \\
S\mathrm{O}(3) & \cong & S\mathrm{O}(\mathfrak{su}(2), \langle , \rangle)
\end{array}
$$

Exercise 12.4 *Check the details here!*

What is the differential of the map $\rho : SU(2) \to \mathrm{O}(3)$ at the identity? Let $g(t)$ be a curve in $SU(2)$ with $\frac{d}{dt}\big|_{t=0} g = g'$. We have $\frac{d}{dt}(g(t)Ag^*(t)) = (\frac{d}{dt}g(t))Ag^*(t) + g(t)A(\frac{d}{dt}g(t))^*$ and so the map $ad : g' \mapsto g'A + Ag'^* = [g', A]$

$$
\begin{aligned}
\frac{d}{dt}\langle g\widehat{x}, g\widehat{y} \rangle &= \frac{d}{dt}\frac{1}{2}tr(g\widetilde{x}(g\widetilde{y})^*) \\
&\frac{1}{2}tr([g', \widetilde{x}], (\widetilde{y})^*) + \frac{1}{2}tr(\widetilde{x}, ([g', \widetilde{y}])^*) \\
&== \frac{1}{2}tr([g', \widetilde{x}], (\widetilde{y})^*) - \frac{1}{2}tr(\widetilde{x}, [g', \widetilde{y}]) \\
&= \langle [g', \widetilde{x}], \widetilde{y} \rangle - \langle \widetilde{x}, [g', \widetilde{y}] \rangle \\
&= .\langle ad(g')\widetilde{x}, \widetilde{y} \rangle - \langle \widetilde{x}, ad(g')\widetilde{y} \rangle
\end{aligned}
$$

>From this is follows that the differential of the map $SU(2) \to \mathrm{O}(3)$ takes $\mathfrak{su}(2)$ isomorphically onto the space $\mathfrak{so}(3)$. We have

$$\begin{array}{ccc} \mathfrak{su}(2) & = & \mathfrak{su}(2) \\ d\rho \downarrow & & ad \downarrow \\ S\mathrm{O}(3) & \cong & \mathfrak{so}(\mathfrak{su}(2), \langle , \rangle) \end{array}$$

where $\mathfrak{so}(\mathfrak{su}(2), \langle , \rangle)$ denotes the linear maps $\mathfrak{su}(2) \to \mathfrak{su}(2)$ skew-symmetric with respect to the inner product $\langle \widetilde{x}, \widetilde{y} \rangle := \frac{1}{2} tr(\widetilde{x}\widetilde{y}^*)$.

Chapter 13

DeRham Cohomology

For a given manifold M of dimension n we have the sequence of maps

$$C^\infty(M) = \Omega^0(M) \xrightarrow{d} \Omega^1(M) \xrightarrow{d} \cdots \xrightarrow{d} \Omega^n(M) \xrightarrow{d} 0$$

and we have defined the de Rham cohomology groups (actually vector spaces) as the quotient

$$H^i(M) = \frac{Z^i(M)}{B^i(M)}$$

where $Z^i(M) := \ker(d : \Omega^i(M) \to \Omega^{i+1}(M))$ and $B^i(M) := \operatorname{Im}(d : \Omega^{i-1}(M) \to \Omega^i(M))$. The elements of $Z^i(M)$ are called **closed** i-forms or **cocycles** and the elements of $B^i(M)$ are called **exact** i-forms or **coboundaries.**

Let us immediately consider a simple situation that will help the reader see what these cohomologies are all about. Let $M = \mathbb{R}^2 - \{0\}$ and consider the 1-form

$$\vartheta := \frac{xdy - ydx}{x^2 + y^2}.$$

We got this 1-form by taking the exterior derivative of $\theta = \arctan(x)$. This function is not defined as a single valued smooth function on all of $M = \mathbb{R}^2 - \{0\}$ but it turns out that the result $\frac{ydx - xdy}{x^2+y^2}$ is well defined on *all* of M. One may also check that $d\vartheta = 0$ and so ϑ is closed. We have the following situation:

1. $\vartheta := \frac{xdy - ydx}{x^2+y^2}$ is a smooth 1-form M with $d\vartheta = 0$.

2. There is no function f defined on (all of) M such that $\vartheta = df$.

3. For any small ball $B(p, \varepsilon)$ in M there is a function $f \in C^\infty(B(p, \varepsilon))$ such that $\vartheta|_{B(p,\varepsilon)} = df$.

   (1) says that ϑ is globally well defined and closed while (2) says that ϑ is not exact. (3) says that ϑ is what we might call locally exact. What prevents us from finding a (global) function with $\vartheta = df$? Could the same kind of situation occur if $M = \mathbb{R}^2$? The answer is no and this difference between $\mathbb{R}^2$ and $\mathbb{R}^2 - \{0\}$ is that $H^1(\mathbb{R}^2) = 0$ while $H^1(\mathbb{R}^2 - \{0\}) \neq 0$.

Exercise 13.1 *Verify (1) and (3) above.*

The reader may be aware that this example has something to do with path independence. In fact, if we could show that for a given 1-form α, the path integral $\int_c \alpha$ only depended on the beginning and ending points of the curve $c(0)$ then we could define a function $f(x) := \int_{x_0}^x \alpha$ where $\int_{x_0}^x \alpha$ is just the path integral for any path beginning at a fixed x_0 and ending at x. With this definition one can show that $df = \alpha$ and so α would be exact. In our example the form ϑ is not exact and so there must be a failure of path independence.

Exercise 13.2 *A smooth fixed endpoint homotopy between a path $c_0 : [x_0, x] \to M$ and $c_1 : [x_0, x] \to M$ is a one parameter family of paths h_s such that $h_0 = c_0$ and $h_1 = c_0$ and such that the map $H(s,t) := h_s(t)$ is smooth on $[0,1] \times [x_0, x]$. Show that if α is an exact 1-form then $\frac{d}{ds}\int_{h_s} \alpha = 0$.*

Since we have found a closed $1-$form on $\mathbb{R}^2\backslash\{0\}$ that is not exact we know that $H^1(\mathbb{R}^2\backslash\{0\}) \neq 0$. We are not yet in a position to determine $H^1(\mathbb{R}^2\backslash\{0\})$ completely. We will start out with even simpler spaces and eventually, develop the machinery to bootstrap our way up to more complicated situations.

First, let $M = \{p\}$. That is, M consists of a single point and is hence a 0-dimensional manifold. In this case,

$$\Omega^k(\{p\}) = Z^k(\{p\}) = \begin{cases} \mathbb{R} & \text{for } k = 0 \\ \\ 0 & \text{for } k > 0 \end{cases}$$

Furthermore, $B^k(\{p\}) = 0$ and so

$$H^k(\{p\}) = \begin{cases} \mathbb{R} & \text{for } k = 0 \\ \\ 0 & \text{for } k > 0 \end{cases}.$$

Next we consider the case $M = \mathbb{R}$. Here, $Z^0(\mathbb{R})$ is clearly just the constant functions and so is (isomorphic to) $\mathbb{R}$. On the other hand, $B^0(\mathbb{R}) = 0$ and so

$$H^0(\mathbb{R}) = \mathbb{R}.$$

Now since $d : \Omega^1(\mathbb{R}) \to \Omega^2(\mathbb{R}) = 0$ we see that $Z^1(\mathbb{R}) = \Omega^1(\mathbb{R})$. If $g(x)dx \in \Omega^1(\mathbb{R})$ then letting

$$f(x) := \int_0^x g(x)dx$$

we get $df = g(x)dx$. Thus, every $\Omega^1(\mathbb{R})$ is exact; $B^1(\mathbb{R}) = \Omega^1(\mathbb{R})$. We are led to

$$H^1(\mathbb{R}) = 0.$$

From this modest beginning we will be able to compute the de Rham cohomology for a large class of manifolds. Our first goal is to compute $H^k(\mathbb{R})$ for all k. In order to accomplish this we will need a good bit of preparation. The methods are largely algebraic and so will need to introduce a small portion of "homological algebra".

Definition 13.1 *Let R be a commutative ring. A **differential** $R-$**complex** is a direct sum of modules $C = \bigoplus_{k\in\mathbb{Z}} C^k$ together with a linear map $d : C \to C$ such that $d \circ d = 0$ and such that $d(C^k) \subset C^{k+1}$. Thus we have a sequence of linear maps*

$$\cdots C^{k-1} \xrightarrow{d} C^k \xrightarrow{d} C^{k+1}$$

where we have denoted the restrictions $d_k = d|\, C^k$ all simply by the single letter d.

Let $A = \bigoplus_{k\in\mathbb{Z}} A^k$ and $B = \bigoplus_{k\in\mathbb{Z}} B^k$ be differential complexes. A map $f : A \to B$ is called a **chain map** if f is a (degree 0) graded map such that $d \circ f = f \circ g$. In other words, if we let $f|\, A^k := f_k$ then we require that $f_k(A^k) \subset B^k$ and that the following diagram commutes for all k:

$$\begin{array}{ccccccc} \xrightarrow{d} & A^{k-1} & \xrightarrow{d} & A^k & \xrightarrow{d} & A^{k+1} & \xrightarrow{d} \\ & f_{k-1}\downarrow & & f_k\downarrow & & f_{k+1}\downarrow & \\ \xrightarrow{d} & B^{k-1} & \xrightarrow{d} & B^k & \xrightarrow{d} & B^{k+1} & \xrightarrow{d} \end{array}.$$

Notice that if $f : A \to B$ is a chain map then $\ker(f)$ and $\mathrm{im}(f)$ are complexes with $\ker(f) = \bigoplus_{k\in\mathbb{Z}} \ker(f_k)$ and $\mathrm{im}(f) = \bigoplus_{k\in\mathbb{Z}} \mathrm{im}(f_k)$. Thus the notion of **exact sequence of chain maps** may be defined in the obvious way.

Definition 13.2 *The k-th cohomology of the complex $C = \bigoplus_{k\in\mathbb{Z}} C^k$ is*

$$H^k(C) := \frac{\ker(\,d|\,C^k)}{\mathrm{im}(\,d|\,C^{k-1})}$$

The elements of $\ker(\,d|\,C^k)$ *(also denoted $Z^k(C)$) are called **cocycles** while the elements of* $\mathrm{im}(\,d|\,C^{k-1})$ *(also denoted $B^k(C)$) are called **coboundaries**.*

We already have an example since by letting $\Omega^k(M) := 0$ for $k < 0$ we have a differential complex $d : \Omega(M) \to \Omega(M)$ where d is the exterior derivative. In this case, $H^k(\Omega(M)) = H^k(M)$ by definition.

Remark 13.1 *In fact, $\Omega(M)$ is an algebra under the exterior product (recall that $\wedge : \Omega^k(M) \times \Omega^l(M) \to \Omega^{l+k}(M)$). This algebra structure actually remains active at the level of cohomology: If $\alpha \in Z^k(M)$ and $\beta \in Z^l(M)$ then for any $\alpha', \beta' \in \Omega^{k-1}(M)$ and any $\beta' \in \Omega^{l-1}(M)$ we have*

$$\begin{aligned}(\alpha + d\alpha') \wedge \beta &= \alpha \wedge \beta + d\alpha' \wedge \beta \\ &= \alpha \wedge \beta + d(\alpha' \wedge \beta) - (-1)^{k-1}\alpha' \wedge d\beta \\ &= \alpha \wedge \beta + d(\alpha' \wedge \beta)\end{aligned}$$

and similarly $\alpha \wedge (\beta + d\beta') = \alpha \wedge \beta + d(\alpha \wedge \beta')$. Thus we may define a product $H^k(M) \times H^l(M) \to H^{k+l}(M)$ by $[\alpha] \wedge [\beta] := [\alpha \wedge \beta]$.

Returning to the general algebraic case, if $f : A \to B$ is a chain map then it is easy to see that there is a natural (degree 0) graded map $f^* : H \to H$ defined by

$$f^*([x]) := [f(x)] \text{ for } x \in C^k.$$

Definition 13.3 *An exact sequence of chain maps of the form*

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

*is called a **short exact sequence**.*

Associated to every short exact sequence of chain maps there is a long exact sequence of cohomology groups:

$$\begin{tikzcd}
H^{k+1}(A) \arrow[r, "f^*"] & H^{k+1}(B) \arrow[r, "g^*"] & H^{k+1}(C) \\
H^k(A) \arrow[r, "f^*"] & H^k(B) \arrow[r, "g^*"] & H^k(C) \arrow[ull, "d^*"']
\end{tikzcd}$$

The maps f^* and g^* are the maps induced by f and g and the "coboundary map" $d^* : H^k(C) \to H^{k+1}(A)$ is defined as follows: Let $c \in Z^k(C) \subset C^k$ represent the class $[c] \in H^k(C)$ so that $dc = 0$. Starting with this we hope to end up with a well defined element of $H^{k+1}(A)$ but we must refer to the following diagram to explain how we arrive at a choice of representative of our class $d^*([c])$:

$$\begin{array}{ccccccccc}
0 \longrightarrow & A^{k+1} & \xrightarrow{f} & B^{k+1} & \xrightarrow{g} & C^{k+1} & \longrightarrow 0 \\
 & d\uparrow & & d\uparrow & & d\uparrow & \\
0 \longrightarrow & A^k & \xrightarrow{f} & B^k & \xrightarrow{g} & C^k & \longrightarrow 0
\end{array}$$

By the surjectivity of g there is an $b \in B^k$ with $g(b) = c$. Also, since $g(db) = d(g(b)) = dc = 0$, it must be that $db = f(a)$ for some $a \in A^{k+1}$. The scheme of the process is

$$c \dashrightarrow b \dashrightarrow a.$$

Certainly $f(da) = d(f(a)) = ddb = 0$ and so since f is 1-1 we must have $da = 0$ which means that $a \in Z^{k+1}(C)$. We would like to define $d^*([c])$ to be $[a]$ but we must show that this is well defined. Suppose that we repeat this process starting with $c' = c + dc_{k-1}$ for some $c_{k-1} \in C^{k-1}$. In our first step we find $b' \in B^k$

with $g(b') = c'$ and then a' with $f(a') = db'$. We wish to show that $[a] = [a']$. We have $g(b - b') = c - c = 0$ and so there is an $a_k \in A^k$ with $f(a_k) = b - b'$. By commutativity we have

$$\begin{aligned} f(d(a_k)) &= d(f(a_k)) = d(b - b') \\ &= db - db' = f(a) - f(a') = f(a - a') \end{aligned}$$

and then since f is 1-1 we have $d(a_k) = a - a'$ which means that $[a] = [a']$. Thus our definition $\delta([c]) := [a]$ is independent of the choices. We leave it to the reader to check (if there is any doubt) that d^* so defined is linear.

We now return to the de Rham cohomology. If $f : M \to N$ is a C^∞ map then we have $f^* : \Omega(N) \to \Omega(M)$. Since pull-back commutes with exterior differentiation and preserves the degree of differential forms, f^* is a chain map. Thus we have the induced map on the cohomology that we will also denote by f^* :

$$\begin{aligned} &f^* : H^*(M) \to H^*(M) \\ &f^* : [\alpha] \mapsto [f^*\alpha] \end{aligned}$$

where we have used $H^*(M)$ to denote the direct sum $\bigoplus_i H^i(M)$. Notice that $f \mapsto f^*$ together with $M \mapsto H^*(M)$ is a contravariant functor which means that for $f : M \to N$ and $g : N \to P$ we have

$$(g \circ f)^* = f^* \circ g^*.$$

In particular if $\iota_U : U \to M$ is inclusion of an open set U then $\iota_U^*\alpha$ is the same as restriction of the form α to U. If $[\alpha] \in H^*(M)$ then $f^*([\alpha]) \in H^*(U)$;

$$f^* : H^*(M) \to H^*(U).$$

13.1 The Meyer Vietoris Sequence

Suppose that $M = U_0 \cup U_1$ for open sets U. Let $U_0 \sqcup U_1$ denote the disjoint union of U and V. We then have inclusions $\iota_1 : U_1 \to M$ and $\iota_2 : U_2 \to M$ as well as the inclusions

$$\partial_0 : U_0 \cap U_1 \to U_1 \hookrightarrow U_0 \sqcup U_1$$

and

$$\partial_1 : U_0 \cap U_1 \to U_0 \hookrightarrow U_0 \sqcup U_1$$

that we indicate (following [Bott and Tu]) by writing

$$M \overset{\iota_0}{\underset{\iota_1}{\leftleftarrows}} U_0 \sqcup U_1 \overset{\partial_0}{\underset{\partial_1}{\leftleftarrows}} U_0 \cap U_1 \ .$$

This gives rise to the **Mayer-Vietoris short exact sequence**

$$0 \to \Omega(M) \xrightarrow{\iota^*} \Omega(U_0) \oplus \Omega(U_1) \xrightarrow{\partial^*} \Omega(U_0 \cap U_1) \to 0$$

where $\iota(\omega) := (\iota_0^*\omega, \iota_1^*\omega)$ and $\partial^*(\alpha, \beta) := (\partial_0^*(\beta) - \partial_1^*(\alpha))$. Notice that $\iota_0^*\omega \in \Omega(U_0)$ while $\iota_1^*\omega \in \Omega(U_1)$. Also, $\partial_0^*(\beta) = \beta|_{U_0 \cap U_1}$ and $\partial_1^*(\alpha) = \alpha|_{U_0 \cap U_1}$ and so live in $\Omega(U_0 \cap U_1)$.

Let us show that this sequence is exact. First if $\iota(\omega) := (\iota_1^*\omega, \iota_0^*\omega) = (0, 0)$ then $\omega|_{U_0} = \omega|_{U_1} = 0$ and so $\omega = 0$ on $M = U_0 \cup U_1$ thus ι^* is 1-1 and exactness at $\Omega(M)$ is demonstrated.

Next, if $\eta \in \Omega(U_0 \cap U_1)$ then we take a smooth partition of unity $\{\rho_0, \rho_1\}$ subordinate to the cover $\{U_0, U_1\}$ and then let $\omega := (-(\rho_1\eta)^{U_0}, (\rho_0\eta)^{U_1})$ where we have extended $\rho_1|_{U_0 \cap U_1}\eta$ by zero to a smooth function $(\rho_1\eta)^{U_0}$ on U_0 and $\rho_0|_{U_0 \cap U_1}\eta$ to a function $(\rho_0\eta)^{U_1}$ on U_1 (think about this). Now we have

$$\begin{aligned}
&\partial^*(-(\rho_1\eta)^{U_0}, (\rho_0\eta)^{U_1}) \\
&= ((\rho_0\eta)^{U_1}|_{U_0 \cap U_1} + (\rho_1\eta)^{U_0}|_{U_0 \cap U_1}) \\
&= \rho_0\eta|_{U_0 \cap U_1} + \rho_1\eta|_{U_0 \cap U_1} \\
&= (\rho_0 + \rho_1)\eta = \eta.
\end{aligned}$$

Perhaps the notation is too pedantic. If we let the restrictions and extensions by zero take care of themselves, so to speak, then the idea is expressed by saying that ∂^* maps $(-\rho_1\eta, \rho_0\eta) \in \Omega(U_0) \oplus \Omega(U_1)$ to $\rho_0\eta - (-\rho_1\eta) = \eta \in \Omega(U_0 \cap U_1)$. Thus we see that ∂^* is surjective.

It is easy to see that $\partial^* \circ \iota^* = 0$ so that $\operatorname{im}(\partial^*) \subset \ker(\iota^*)$. Finally, let $(\alpha, \beta) \in \Omega(U_0) \oplus \Omega(U_1)$ and suppose that $\partial^*(\alpha, \beta) = (0, 0)$. This translates to $\alpha|_{U_0 \cap U_1} = \beta|_{U_0 \cap U_1}$ which means that there is a form $\omega \in \Omega(U_0 \cup U_1) = \Omega(M)$ such that ω coincides with α on U_0 and with β on U_0. Thus

$$\begin{aligned}
\iota^*\omega &= (\iota_0^*\omega, \iota_1^*\omega) \\
&= (\alpha, \beta)
\end{aligned}$$

so that $\ker(\iota^*) \subset \operatorname{im}(\partial^*)$ that together with the reverse inclusion gives $\operatorname{im}(\partial^*) = \ker(\iota^*)$.

Following the general algebraic pattern, the Mayer-Vietoris short exact sequence gives rise to the **Mayer-Vietoris long exact sequence**:

$$\begin{tikzcd}
H^{k+1}(M) \arrow[r] & H^{k+1}(U) \oplus H^{k+1}(V) \arrow[r, "\partial"] & H^{k+1}(U \cap V) \\
H^k(M) \arrow[r] & H^k(U) \oplus H^k(V) \arrow[r, "\partial"] & H^k(U \cap V) \arrow[ull, bend right, "d^*"']
\end{tikzcd}$$

Since our description of the coboundary map in the algebraic case was rather abstract we will do well to take a closer look at d^* in the present context. Referring to the diagram below, $\omega \in \Omega^k(U \cap V)$ represents a cohomology class $[\omega] \in H^k(U \cap V)$ so that in particular $d\omega = 0$.

$$\begin{array}{ccccccccc} & & \uparrow & & \uparrow & & \uparrow & & \\ 0 & \to & \Omega^{k+1}(M) & \to & \Omega^{k+1}(U) \oplus \Omega^{k+1}(V) & \to & \Omega^{k+1}(U \cap V) & \to 0 \\ & & d\uparrow & & d\uparrow & & d\uparrow & & \\ 0 & \to & \Omega^{k}(M) & \to & \Omega^{k}(U) \oplus \Omega^{k}(V) & \to & \Omega^{k}(U \cap V) & \to 0 \end{array}$$

By exactness of the rows we can find a form $\xi \in \Omega^k(U) \oplus \Omega^k(V)$ which maps to ω. In fact, we may take $\xi = (-\rho_V \omega, \rho_U \omega)$ as before. Now since $d\omega = 0$ and the diagram commutes, we see that $d\xi$ must map to 0 in $\Omega^{k+1}(U \cap V)$. This just tells us that $-d\,(\rho_V \omega)$ and $d\,(\rho_U \omega)$ agree on the intersection $U \cap V$. Thus there is a well defined form in $\Omega^{k+1}(M)$ which maps to $d\xi$. This global form η is given by

$$\eta = \begin{cases} -d(\rho_V \omega) & \text{on } U \\ \\ d(\rho_U \omega) & \text{on } V \end{cases}$$

and then by definition $d^*[\omega] = [\eta] \in H^{k+1}(M)$.

Exercise 13.3 *Let the circle S^1 be parameterized by angle θ in the usual way. Let U be that part of a circle with $-\pi/4 < \theta < \pi/4$ and let V be given by $3\pi/4 < \theta < 5\pi/4$.*

(a) Show that $H(U) \cong H(V) \cong \mathbb{R}$

(b) Show that the "difference map" $H(U) \oplus H(V) \xrightarrow{\partial} H(U \cap V)$ has 1-dimensional image.

Now we come to an important lemma which provides the leverage needed to compute the cohomology of some higher dimensional manifolds based on that of lower dimensional manifolds. We start we a smooth manifold M and also form the product manifold $M \times \mathbb{R}$. We then have the obvious projection $pr_1 : M \times \mathbb{R} \to M$ and the "zero section" $s_0 : M \to M \times \mathbb{R}$ given by $x \mapsto (x, 0)$. We can apply the cohomology functor to this pair of maps:

$$\begin{array}{ccc} M \times \mathbb{R} & & \Omega(M \times \mathbb{R}) \\ pr_1 \downarrow\uparrow s_0 & \rightsquigarrow & pr_1^* \downarrow\uparrow s_0^* \\ M & & \Omega(M) \end{array}$$

The maps pr_1^* and s_0^* induce maps on cohomology:

$$\begin{array}{c} H^*(M \times \mathbb{R}) \\ pr_1^* \downarrow\uparrow s_0^* \\ H^*(M) \end{array}$$

Theorem 13.1 (Poincaré Lemma) *Given M and the maps defined above we have that $pr_1^* : H^*(M \times \mathbb{R}) \to H^*(M)$ and $s_0^* : H^*(M) \to H^*(M \times \mathbb{R})$ are mutual inverses. In particular,*

$$H^*(M \times \mathbb{R}) \cong H^*(M).$$

Proof. The main idea of the proof is the use of a so called homotopy operator which in the present case is a degree -1 map $K : \Omega(M \times \mathbb{R}) \to \Omega(M \times \mathbb{R})$ with the property that $id_M - pr_1^* \circ s_0^* = \pm(d \circ K - K \circ d)$. The point is that such a map must map closed forms to closed forms and exact forms to exact forms. This on the level of cohomology $\pm(d \circ K - K \circ d)$ and hence $id_M - pr_1^* \circ s_0^*$ must be the zero map so that in fact $id = pr_1^* \circ s_0^*$ on $H^*(M)$.

Our task is to construct K. First notice that a function $\phi(x, t)$ on $M \times \mathbb{R}$ which happens to be constant with respect to the second variable must be of the form $pr_1^* f$ for some $f \in C^\infty(M)$. Similarly, for $\alpha \in \Omega^k(M)$ we think of the form $pr_1^*\alpha$ as not depending on $t \in \mathbb{R}$. For any $\omega \in \Omega^k(M \times \mathbb{R})$ we can find a pair of functions $f_1(x, t)$ and $f_2(x, t)$ such that

$$\omega = f_1(x, t) pr_1^*\alpha + f_2(x, t) pr_1^*\beta \wedge dt$$

for some forms $\alpha, \beta \in \Omega^k(M)$. This decomposition is unique in the sense that if $f_1(x, t)pr_1^*\alpha + f_2(x, t)pr_1^*\beta \wedge dt = 0$ then $f_1(x, t)pr_1^*\alpha = 0$ and $f_2(x, t)pr_1^*\beta \wedge dt = 0$. Using this decomposition we have a well defined map o

$$\Omega^k(M \times \mathbb{R}) \ni \omega \mapsto \int_0^t f_2(x, \tau) d\tau \times pr_1^*\beta \in \Omega^{k-1}(M \times \mathbb{R})$$

This map is our proposed homotopy operator K.

Let us now check that K has the required properties. It is clear from what we have said that we may check the action of K separately on forms of the type and $f_1(x, t)pr_1^*\alpha$ and the type $f_2(x, t)pr_1^*\beta \wedge dt$.

Case I (type $f_1(x, t)pr_1^*\alpha$).If $\omega = f_1(x, t)pr_1^*\alpha$ then $K\omega = 0$ and so $(d \circ K - K \circ d)\omega = -K(d\omega)$ and

$$\begin{aligned}
K(d\omega) &= K(d(f_1(x, t)pr_1^*\alpha)) \\
&= K(df_1(x, t) \wedge pr_1^*\alpha + f_1(x, t)pr_1^* d\alpha) \\
&= K(df_1(x, t) \wedge pr_1^*\alpha) = \pm K(\frac{\partial f_1}{\partial t}(x, t)pr_1^*\alpha \wedge dt) \\
&= \pm \int_0^t \frac{\partial f_1}{\partial t}(x, \tau) d\tau \times pr_1^*\alpha = \pm\{f_1(x, t) - f(x, 0)\} pr_1^*\alpha
\end{aligned}$$

On the other hand,

$$\begin{aligned}
(id_M - pr_1^* \circ s_0^*)\omega &= \\
(id_M - pr_1^* \circ s_0^*) f_1(x, t) pr_1^*\alpha &= f_1(x, t)pr_1^*\alpha - f_1(x, 0)pr_1^*\alpha \\
&= \pm\{f_1(x, t) - f(x, 0)\} pr_1^*\alpha
\end{aligned}$$

as above. So in the present case we get $(d \circ K - K \circ d)\omega = \pm(id_M - pr_1^* \circ s_0^*)\omega$.

Case II. (type $\omega = f(x,t)dt \wedge pr_1^*\beta$) In this case

$$\begin{aligned}
&(d \circ K - K \circ d)\omega \\
&= dK\{f(x,t)dt \wedge pr_1^*\beta\} - Kd\{f(x,t)dt \wedge pr_1^*\beta\} \\
&= d\{\int_0^t f(x,\tau)d\tau \times pr_1^*\beta\} - Kd(f(x,t)dt \wedge pr_1^*\beta) + 0 \\
&= d\{\int_0^t f(x,\tau)d\tau \times pr_1^*\beta\} - K\left\{\frac{\partial f}{\partial t}(x,t)dt \wedge pr_1^*\beta + f(x,t)pr_1^*d\beta\right\} \\
&= f(x,t)dt \wedge pr_1^*\beta + \int_0^t f(x,\tau)d\tau \times pr_1^*d\beta - pr_1^*d\beta \times \int_0^t f(x,\tau)d\tau - \int_0^t \frac{\partial f}{\partial t}dt \times pr_1^*\beta \\
&= f(x,t)pr_1^*\beta \wedge dt = \omega
\end{aligned}$$

On the other hand, we also have $(id_M - pr_1^* \circ s_0^*)\omega = \omega$ since $s_0^* dt = 0$.

This concludes the proof. ■

Corollary 13.1

$$H^k(point) = H^k(\mathbb{R}^n) = \begin{cases} \mathbb{R} & if\ k = 0 \\ 0 & otherwise \end{cases}$$

Proof. One first verifies that the statement is true for $H^k(point)$. Then the remainder of the proof is a simple induction:

$$\begin{aligned}
H^k(point) &\cong H^k(point \times \mathbb{R}) = H^k(\mathbb{R}) \\
&\cong H^k(\mathbb{R} \times \mathbb{R}) = H^k(\mathbb{R}^2) \\
&\cong ..... \\
&\cong H^k(\mathbb{R}^{n-1} \times \mathbb{R}) = H^k(\mathbb{R}^n)
\end{aligned}$$

■

Corollary 13.2 (Homotopy invariance) *If $f : M \to N$ and $g : M \to N$ are homotopic then the induced maps $f^* : H^*(N) \to H^*(M)$ and $g^* : H^*(N) \to H^*(M)$ are equal.*

Proof. By extending the alleged homotopies in a trivial way we may assume that we have a map $F : M \times \mathbb{R} \to N$ such that

$$\begin{aligned}
F(x,t) &= f(x) \text{ for } t \geq 1 \\
F(x,t) &= g(x) \text{ for } t \leq 0.
\end{aligned}$$

If $s_1(x) := (x,1)$ and $s_0(x) := (x,0)$ then $f = F \circ s_1$ and $g = F \circ s_0$ and so

$$\begin{aligned}
f^* &= s_1^* \circ F^* \\
g^* &= s_0^* \circ F^*.
\end{aligned}$$

It is easy to check that s_1^* and s_0^* are one sided inverses of pr_1^*. But we have shown that pr_1^* is an isomorphism. It follows that $s_1^* = s_0^*$ and then from above we have $f^* = g^*$. ■

Homotopy plays a central role in algebraic topology and so this last result is very important. Recall that if there exist maps $f : M \to N$ and $g : N \to M$ such that both $f \circ g$ and $g \circ f$ are defined and homotopic to id_N and id_M respectively then f (or g) is called a homotopy equivalence and M and N are said to have the same homotopy type. In particular, if a topological space has the same homotopy type as a single point then we say that the space is contractible. If we are dealing with smooth manifolds we may take the maps to be smooth. In fact, any continuous map between two manifolds is continuously homotopic to a smooth map. We shall use this fact often without mention. The following corollary follows easily:

Corollary 13.3 *If M and N are smooth manifolds which are of the same homotopy type then $H^*(M) \cong H^*(N)$.*

Next consider the situation where A is a subset of M and $i : A \hookrightarrow M$ is the inclusion map. If there exist a map $r : M \to A$ such that $r \circ i = id_A$ then we say that r is a retraction of M onto A. If A is a submanifold of a smooth manifold M then if there is a retraction r of M onto A we may assume that r is smooth. If we can find a retraction r such that $i \circ r$ is homotopic to the identity id_M then we say that r is a deformation retraction. The following exercises show the usefulness of these ideas.

Exercise 13.4 *Let U_+ and U_- be open subsets of the sphere $S^n \subset \mathbb{R}^n$ given by*

$$\begin{aligned} U_+ &:= \{(x^i) \in S^n : -\varepsilon < x^{n+1} \leq 1 \\ U_- &:= \{(x^i) \in S^n : -1 \leq x^{n+1} < \varepsilon \end{aligned}$$

where $0 < \varepsilon < 1/2$. Show that there is a deformation retraction of $U_+ \cap U_-$ onto the equator $x^{n+1} = 0$ in S^n. Notice that the equator is a two point set in case $n = 0$.

Exercise 13.5 *Use the last exercise and the long Meyer-Vietoris sequence to show that*

$$H^k(S^n) = \begin{cases} \mathbb{R} & \text{if } k = 0 \text{ or } n \\ 0 & \text{otherwise} \end{cases}$$

13.1.1 Compactly Supported Cohomology

Let $\Omega_c(M)$ denote the algebra of compactly supported differential forms on a manifold M. Obviously, if M is compact then $\Omega_c(M) = \Omega(M)$ so our main interest here is the case where M is not compact. We now have a slightly different complex

$$\cdots \xrightarrow{d} \Omega_c^k(M) \xrightarrow{d} \Omega_c^{k+1}(M) \xrightarrow{d} \cdots$$

which has a corresponding cohomology $H_c^*(M)$ called the **de Rham cohomology with compact support**. Now if we look at the behavior of differential forms under the operation of pullback we immediately realize that the pull back of a differential form with compact support may not have compact support. In order to get desired functorial properties we consider the class of smooth maps called proper maps:

Definition 13.4 *A smooth map $f : P \to M$ is called a* ***proper map*** *if $f^{-1}(K)$ is compact whenever K is compact.*

It is easy to verify that the set of all smooth manifolds together with proper smooth maps is a category and the assignments $M \mapsto \Omega_c(M)$ and $f \mapsto \{\alpha \mapsto f^*\alpha\}$ is a contravariant functor. In plain language this means that if $f : P \to M$ is a proper map then $f^* : \Omega_c(M) \to \Omega_c(P)$ and for two such maps we have $(f \circ g)^* = g^* \circ f^*$ as before but the assumption that f and g are proper maps is essential.

We can also achieve functorial behavior by using the assignment $M \mapsto \Omega_c(M)$. The first thing we need is a new category (which is fortunately easy to describe). The category we have in mind has a objects the set of all open subsets of a fixed manifold M. The morphisms are inclusion maps $V \overset{j_{VU}}{\hookrightarrow} U$ which are only defined in case $V \subset U$. Now for any such inclusion j_{VU} we define a map $(j_{V,U})_* : \Omega_c(V) \to \Omega_c(U)$ according to the following simple prescription: For any $\alpha \in \Omega_c(V)$ let $(j_{V,U})_* \alpha$ be the form in $\Omega_c(U)$ which is equal to α at all points in V and equal to zero otherwise (this is referred to as extension by zero). Since the support of α is neatly inside the open set V we can see that the extension $(j_{V,U})_* \alpha$ is perfectly smooth.

We warn the reader that what we are about to describe is much simpler than the notation which describes it and a little doodling while reading might be helpful. If U and V are open subsets of M with nonempty intersection then we have inclusions $j_{V\cap U,U} : V \cap U \to U$ and $j_{V\cap U,V} : V \cap U \to V$ as well as the inclusions $j_{V,M} : V \to M$ and $j_{U,M} : U \to M$. Now if $U \sqcup V$ is the disjoint union of U and V, then we have inclusions $V \to U \sqcup V$ and $U \to U \sqcup V$ and after composing in the obvious way we get two different maps $j_1 : V \cap U \hookrightarrow U \sqcup V$ and $j_2 : V \cap U \hookrightarrow U \sqcup V$. Also, using $j_{V,M} : V \to M$ and $j_{U,M} : U \to M$ we get another obvious map $U \sqcup V \to M$ (which is not exactly an inclusion). Following [Bott and Tu] we denote this situation as

$$M \longleftarrow U \sqcup V \leftleftarrows V \cap U. \tag{13.1}$$

Now let us define a map $\delta : \Omega_c(V \cap U) \to \Omega_c(V) \oplus \Omega_c(U)$ by $\alpha \mapsto (-j_{V\cap U,U*}\alpha, j_{V\cap U,V*}\alpha)$ which we can see as arising from $U \sqcup V \leftleftarrows V \cap U$. If we also consider the map $\Omega_c(V) \oplus \Omega_c(U) \to \Omega_c(M)$ which sums: $(\alpha, \beta) \mapsto \alpha + \beta$ then we can associate to the sequence 13.1 above, the new sequence

$$0 \leftarrow \Omega_c(M) \overset{sum}{\longleftarrow} \Omega_c(V) \oplus \Omega_c(U) \overset{\delta}{\longleftarrow} \Omega_c(V \cap U) \leftarrow 0$$

This is the (short) Mayer-Vietoris sequence for differential forms with compact support.

Theorem 13.2 *The sequence 13.1.1 above is exact.*

The proof is not hard but in the interest of saving space we simply refer the reader to the book [Bott and Tu].

Corollary 13.4 *There is a long exact sequence*

$$\begin{tikzcd}
 & {} & \\
H_c^{k+1}(M) \arrow[r] & H_c^{k+1}(U) \oplus H_c^{k+1}(V) \arrow[r, "\partial"] & H_c^{k+1}(U \cap V) \arrow[ul, bend right, "d^*"'] \\
H_c^{k}(M) \arrow[r] & H_c^{k}(U) \oplus H_c^{k}(V) \arrow[r, "\partial"] & H_c^{k}(U \cap V) \arrow[ull, bend right, "d^*"'] \\
 & {} \arrow[ul, bend left, "d^*"] &
\end{tikzcd}$$

which is called the (long) Mayer-Vietoris sequence for cohomology with compact supports.

Chapter 14

Complex Manifolds

14.1 Some complex linear algebra

The set of all n-tuples of complex $\mathbb{C}^n$ numbers is a complex vector space and by choice of a basis, every complex vector space of finite dimension (over $\mathbb{C}$) is linearly isomorphic to $\mathbb{C}^n$ for some n. Now multiplication by $i := \sqrt{-1}$ is a complex linear map $\mathbb{C}^n \to \mathbb{C}^n$ and since $\mathbb{C}^n$ is also a real vector space $\mathbb{R}^{2n}$ under the identification

$$(x^1 + iy^1, ..., x^n + iy^n) \rightleftharpoons (x^1, y^1, ..., x^n, y^n)$$

we obtain multiplication by i as a real linear map $J_0 : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ given by the matrix

$$\begin{bmatrix} 0 & -1 & & & & \\ 1 & 0 & & & & \\ & & 0 & -1 & & \\ & & 1 & 0 & & \\ & & & & \ddots & \\ & & & & & \ddots \end{bmatrix}.$$

Conversely, if V is a real vector space of dimension $2n$ and there is a map $J : \mathrm{V} \to \mathrm{V}$ with $J^2 = -1$ then we can define the structure of a complex vector space on V by defining the scalar multiplication by complex numbers via the formula

$$(x + \mathrm{i}y)v := xv + yJv \text{ for } v \in \mathrm{V}.$$

Denote this complex vector space by V_J. Now if $e_1, ...e_n$ is a basis for V_J (over $\mathbb{C}$) then we claim that $e_1, ..., e_n, Je_1, ..., Je_n$ is a basis for V over $\mathbb{R}$. We only need to show that $e_1, ..., e_n, Je_1, ..., Je_n$ span. For this let $v \in \mathrm{V}$ then for some complex numbers $c^i = a^i + \mathrm{i}b^j$ we have $v = \sum c^i e_i = \sum (a^j + \mathrm{i}b^j)e_j = \sum a^j e_j + \sum b^j J e_j$ which is what we want.

Next we consider the complexification of V which is $\mathrm{V}_{\mathbb{C}} := \mathbb{C} \otimes \mathrm{V}$. Now any real basis $\{f_j\}$ of V is also a basis for $\mathrm{V}_{\mathbb{C}}$ if we identify f_j with $1 \otimes f_j$.

Furthermore, the linear map $J : \mathrm{V} \to \mathrm{V}$ extends to a complex linear map $J : \mathrm{V}_{\mathbb{C}} \to \mathrm{V}_{\mathbb{C}}$ and still satisfies $J^2 = -1$. Thus this extension has eigenvalues i and $-$i. Let $\mathrm{V}^{1,0}$ be the eigenspace for i and let $\mathrm{V}^{0,1}$ be the $-$i eigenspace. Of course we must have $\mathrm{V}_{\mathbb{C}} = \mathrm{V}^{1,0} \oplus \mathrm{V}^{0,1}$. The reader may check that the set of vectors $\{e_1 - \mathrm{i}Je_1, ..., e_n - \mathrm{i}Je_n\}$ span $\mathrm{V}^{1,0}$ while $\{e_1 + \mathrm{i}Je_1, ..., e_n + \mathrm{i}Je_n\}$ span $\mathrm{V}^{0,1}$. Thus we have a convenient basis for $\mathrm{V}_{\mathbb{C}} = \mathrm{V}^{1,0} \oplus \mathrm{V}^{0,1}$.

Lemma 14.1 *There is a natural complex linear isomorphism* $\mathrm{V}_J \cong \mathrm{V}^{1,0}$ *given by* $e_i \mapsto e_i - \mathrm{i}Je_i$. *Furthermore, the conjugation map on* $\mathrm{V}_{\mathbb{C}}$ *interchanges the spaces* $\mathrm{V}^{1,0}$ *and* $\mathrm{V}^{0,1}$.

Let us apply these considerations to the simple case of the complex plane $\mathbb{C}$. The realification is $\mathbb{R}^2$ and the map J is

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

If we identify the tangent space of $\mathbb{R}^{2n}$ at 0 with $\mathbb{R}^{2n}$ itself then $\{ \frac{\partial}{\partial x^i}\big|_0, \frac{\partial}{\partial y^i}\big|_0 \}_{1\le i\le n}$ is a basis for $\mathbb{R}^{2n}$. A complex basis for $\mathbb{R}^2_J \cong \mathbb{C}$ is $e_1 = \frac{\partial}{\partial x}\big|_0$ and so $\frac{\partial}{\partial x}\big|_0, J\frac{\partial}{\partial x}\big|_0$ is a basis for $\mathbb{R}^2$. This is clear anyway since $J\frac{\partial}{\partial x}\big|_0 = \frac{\partial}{\partial y}\big|_0$. Now the complexification of $\mathbb{R}^2$ is $\mathbb{R}^2_{\mathbb{C}}$ which has basis consisting of $e_1 - \mathrm{i}Je_1 = \frac{\partial}{\partial x}\big|_0 - \mathrm{i}\frac{\partial}{\partial y}\big|_0$ and $e_1 + \mathrm{i}Je_1 = \frac{\partial}{\partial x}\big|_0 + \mathrm{i}\frac{\partial}{\partial y}\big|_0$. These are usually denoted by $\frac{\partial}{\partial z}\big|_0$ and $\frac{\partial}{\partial \bar{z}}\big|_0$. More generally, we see that if $\mathbb{C}^n$ is realified to $\mathbb{R}^{2n}$ which is then complexified to $\mathbb{R}^{2n}_{\mathbb{C}} := \mathbb{C} \otimes \mathbb{R}^{2n}$ then a basis for $\mathbb{R}^{2n}_{\mathbb{C}}$ is given by

$$\{ \frac{\partial}{\partial z^1}\bigg|_0, .., \frac{\partial}{\partial z^n}\bigg|_0, \frac{\partial}{\partial \bar{z}^1}\bigg|_0 ..., \frac{\partial}{\partial \bar{z}^n}\bigg|_0 \}$$

where

$$2\,\frac{\partial}{\partial z^i}\bigg|_0 := \frac{\partial}{\partial x^i}\bigg|_0 - \mathrm{i}\,\frac{\partial}{\partial y^i}\bigg|_0$$

and

$$2\,\frac{\partial}{\partial \bar{z}^i}\bigg|_0 := \frac{\partial}{\partial x^i}\bigg|_0 + \mathrm{i}\,\frac{\partial}{\partial y^i}\bigg|_0 .$$

Now if we consider the tangent bundle $U \times \mathbb{R}^{2n}$ of an open set $U \subset \mathbb{R}^{2n}$ then we have the vector fields $\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^i}$. We can complexify the tangent bundle of $U \times \mathbb{R}^{2n}$ to get $U \times \mathbb{R}^{2n}_{\mathbb{C}}$ and then following the ideas above we have that the fields $\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^i}$ also span each complexified tangent space $T_pU := \{p\} \times \mathbb{R}^{2n}_{\mathbb{C}}$. On the other hand, so do the fields $\{\frac{\partial}{\partial z^1}, .., \frac{\partial}{\partial z^n}, \frac{\partial}{\partial \bar{z}^1}, .., \frac{\partial}{\partial \bar{z}^n}\}$. Now if $\mathbb{R}^{2n}$ had some complex structure, say $\mathbb{C}^n \cong (\mathbb{R}^{2n}, J_0)$, then J_0 defines a bundle map $J_0 : T_pU \to T_pU$ given by $(p, v) \mapsto (p, J_0 v)$. This can be extended to a complex bundle map $J_0 : TU_{\mathbb{C}} = \mathbb{C} \otimes TU \to TU_{\mathbb{C}} = \mathbb{C} \otimes TU$ and we get a bundle decomposition

$$TU_{\mathbb{C}} = T^{1.0}U \oplus T^{0.1}U$$

where $\frac{\partial}{\partial z^1}, .., \frac{\partial}{\partial z^n}$ spans $T^{1.0}U$ at each point and $\frac{\partial}{\partial \bar{z}^1}, .., \frac{\partial}{\partial \bar{z}^n}$ spans $T^{0.1}U$.

Now the symbols $\frac{\partial}{\partial z^1}$ etc., already have meaning as differential operators. Let us now show that this view is at least consistent with what we have done above. For a smooth complex valued function $f : U \subset \mathbb{C}^n \to \mathbb{C}$ we have for $p = (z_1, ..., z_n) \in U$

$$\begin{aligned}
\left.\frac{\partial}{\partial z^i}\right|_p f &= \frac{1}{2}\left(\left.\frac{\partial}{\partial x^i}\right|_p f - \mathrm{i}\left.\frac{\partial}{\partial y^i}\right|_p f\right) \\
&= \frac{1}{2}\left(\left.\frac{\partial}{\partial x^i}\right|_p u - \mathrm{i}\left.\frac{\partial}{\partial y^i}\right|_p u - \left.\frac{\partial}{\partial x^i}\right|_p \mathrm{i}v - \mathrm{i}\left.\frac{\partial}{\partial y^i}\right|_p \mathrm{i}v\right) \\
&\frac{1}{2}\left(\left.\frac{\partial u}{\partial x^i}\right|_p + \left.\frac{\partial v}{\partial y^i}\right|_p\right) + \frac{\mathrm{i}}{2}\left(\left.\frac{\partial u}{\partial y^i}\right|_p - \left.\frac{\partial v}{\partial x^i}\right|_p\right).
\end{aligned}$$

and

$$\begin{aligned}
\left.\frac{\partial}{\partial \bar{z}^i}\right|_p f &= \frac{1}{2}\left(\left.\frac{\partial}{\partial x^i}\right|_p f + \mathrm{i}\left.\frac{\partial}{\partial y^i}\right|_p f\right) \\
&= \frac{1}{2}\left(\left.\frac{\partial}{\partial x^i}\right|_p u + \mathrm{i}\left.\frac{\partial}{\partial y^i}\right|_p u + \left.\frac{\partial}{\partial x^i}\right|_p \mathrm{i}v + \mathrm{i}\left.\frac{\partial}{\partial y^i}\right|_p \mathrm{i}v\right) \\
&= \frac{1}{2}\left(\left.\frac{\partial u}{\partial x^i}\right|_p - \left.\frac{\partial v}{\partial y^i}\right|_p\right) + \frac{\mathrm{i}}{2}\left(\left.\frac{\partial u}{\partial y^i}\right|_p + \left.\frac{\partial v}{\partial x^i}\right|_p\right).
\end{aligned}$$

Definition 14.1 *A function $f : U \subset \mathbb{C}^n \to \mathbb{C}$ is called **holomorphic** if*

$$\frac{\partial}{\partial \bar{z}^i} f \equiv 0 \qquad \text{(all } i\text{)}$$

*on U. A function f is called **antiholomorphic** if*

$$\frac{\partial}{\partial z^i} f \equiv 0 \qquad \text{(all } i\text{).}$$

Definition 14.2 *A map $f : U \subset \mathbb{C}^n \to \mathbb{C}^m$ given by functions $f_1, ..., f_m$ is called **holomorphic** (resp. **antiholomorphic**) if each component function $f_1, ..., f_m$ is holomorphic (resp. antiholomorphic).*

Now if $f : U \subset \mathbb{C}^n \to \mathbb{C}$ is holomorphic then by definition $\left.\frac{\partial}{\partial \bar{z}^i}\right|_p f \equiv 0$ for all $p \in U$ and so we have the **Cauchy-Riemann equations**

$$\begin{aligned}
\frac{\partial u}{\partial x^i} &= \frac{\partial v}{\partial y^i} \\
\frac{\partial v}{\partial x^i} &= -\frac{\partial u}{\partial y^i}
\end{aligned} \qquad \text{(Cauchy-Riemann)}$$

and from this we see that for holomorphic f

$$\frac{\partial f}{\partial z^i} = \frac{\partial u}{\partial x^i} + \mathrm{i}\frac{\partial v}{\partial x^i} = \frac{\partial f}{\partial x^i}$$

which means that as derivations on the sheaf $\mathcal{O}$ of locally defined holomorphic functions on $\mathbb{C}^n$, the operators $\frac{\partial}{\partial z^i}$ and $\frac{\partial}{\partial x^i}$ are equal. This corresponds to the complex isomorphism $T^{1.0}U \cong TU, J_0$ which comes from the isomorphism in lemma **??**. In fact, if one looks at a function $f : \mathbb{R}^{2n} \to \mathbb{C}$ as a differentiable map of real manifolds then with J_0 giving the isomorphism $\mathbb{R}^{2n} \cong \mathbb{C}^n$, our map f is holomorphic if and only if

$$Tf \circ J_0 = J_0 \circ Tf$$

or in other words

$$\begin{pmatrix} \frac{\partial u}{\partial x^1} & \frac{\partial u}{\partial y^1} & \\ \frac{\partial v}{\partial x^1} & \frac{\partial v}{\partial y^1} & \\ & & \ddots \end{pmatrix} \begin{pmatrix} 0 & -1 & \\ 1 & 0 & \\ & & \ddots \end{pmatrix} = \begin{pmatrix} 0 & -1 & \\ 1 & 0 & \\ & & \ddots \end{pmatrix} \begin{pmatrix} \frac{\partial u}{\partial x^1} & \frac{\partial u}{\partial y^1} & \\ \frac{\partial v}{\partial x^1} & \frac{\partial v}{\partial y^1} & \\ & & \ddots \end{pmatrix}.$$

This last matrix equation is just the Cauchy-Riemann equations again.

14.2 Complex structure

Definition 14.3 *A manifold M is said to be an* ***almost complex manifold*** *if there is a smooth bundle map $J : TM \to TM$, called an* ***almost complex structure*** *, having the property that $J^2 = -1$.*

Definition 14.4 *A* ***complex manifold*** *M is a manifold modeled on $\mathbb{C}^n$ for some n, together with an atlas for M such that the transition functions are all holomorphic maps. The charts from this atlas are called* ***holomorphic charts****. We also use the phrase "****holomorphic coordinates****".*

Example 14.1 *Let $S^2(1/2) = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = 1/4\}$ be given coordinates $\psi^+ : (x_1, x_2, x_3) \mapsto \frac{1}{1-x_3}(x_1 + \mathrm{i}x_2) \in \mathbb{C}$ on $U^+ := \{(x_1, x_2, x_3) \in S^2 : 1 - x_3 \neq 0\}$ and $\psi^- : (x_1, x_2, x_3) \mapsto \frac{1}{1+x_3}(x_1 + \mathrm{i}x_2) \in \mathbb{C}$ on $U^- := \{(x_1, x_2, x_3) \in S^2 : 1 + x_3 \neq 0\}$. The chart overlap map or transition function is $\psi^- \circ \psi^+(z) = 1/z$. Since on $\psi^+U^+ \cap \psi^-U^-$ the map $z \mapsto 1/z$ is a biholomorphism we see that $S^2(1/2)$ can be given the structure of a complex 1-manifold.*

Another way to get the same complex 1-manifold is given by taking two copies of the complex plane, say $\mathbb{C}_z$ with coordinate z and $\mathbb{C}_w$ with coordinate w and *then identify $\mathbb{C}_z$ with $\mathbb{C}_w - \{0\}$ via the map $w = 1/z$. This complex surface is of course topologically a sphere and is also the 1 point compactification of the complex plane. As the reader will no doubt already be aware, this complex 1-manifold is called the **Riemann sphere**.*

Example 14.2 *Let $P_n(\mathbb{C})$ be the set of all complex lines through the origin in $\mathbb{C}^{n+1}$, which is to say, the set of all equivalence classes of nonzero elements of $\mathbb{C}^{n+1}$ under the equivalence relation*

$$(z^1, ..., z^{n+1}) \sim \lambda(z^1, ..., z^{n+1}) \text{ for } \lambda \in \mathbb{C}$$

For each i with $1 \leq i \leq n+1$ define the set

$$U_i := \{[z^1, ..., z^{n+1}] \in P_n(\mathbb{C}) : z^i \neq 0\}$$

and corresponding map $\psi_i : U_i \to \mathbb{C}^n$ by

$$\psi_i([z^1, ..., z^{n+1}]) = \frac{1}{z^i}(z^1, ..., \widehat{z^i}, ..., z^{n+1}) \in \mathbb{C}^n.$$

*One can check that these maps provide a holomorphic atlas for $P_n(\mathbb{C})$ which is therefore a complex manifold (**complex projective n-space**).*

Example 14.3 *Let $M_{m\times n}(\mathbb{C})$ be the space of $m \times n$ complex matrices. This is clearly a complex manifold since we can always "line up" the entries to get a map $M_{m\times n}(\mathbb{C}) \to \mathbb{C}^{mn}$ and so as complex manifolds $M_{m\times n}(\mathbb{C}) \cong \mathbb{C}^{mn}$. A little less trivially we have the complex general linear group $GL(n, \mathbb{C})$ which is an open subset of $M_{m\times n}(\mathbb{C})$ and so is an n^2 dimensional complex manifold.*

Example 14.4 (Grassmann manifold) *To describe this important example we start with the set $(M_{n\times k}(\mathbb{C}))_*$ of $n \times k$ matrices with rank $k < n$ (maximal rank). The columns of each matrix from $(M_{n\times k}(\mathbb{C}))_*$ span a k-dimensional subspace of $\mathbb{C}^n$. Define two matrices from $(M_{n\times k}(\mathbb{C}))_*$ to be equivalent if they span the same k-dimensional subspace. Thus the set $G(k, n)$ of equivalence classes is in one to one correspondence with the set of complex k dimensional subspaces of $\mathbb{C}^n$. Now let U be the set of all $[A] \in G(k, n)$ such that A has its first k rows linearly independent. This property is independent of the representative A of the equivalence class $[A]$ and so U is a well defined set. This last fact is easily proven by a Gaussian reduction argument. Now every element $[A] \in U \subset G(k, n)$ is an equivalence class that has a unique member A_0 of the form*

$$\begin{pmatrix} I_{k\times k} \\ Z \end{pmatrix}.$$

Thus we have a map on U defined by $\Psi : [A] \mapsto Z \in M_{n-k\times k}(\mathbb{C}) \cong \mathbb{C}^{k(n-k)}$. We wish to cover $G(k, n)$ with sets U_σ similar to U and defined similar maps. Let $\sigma_{i_1...i_k}$ be the shuffle permutation that puts the k columns indexed by $i_1, ..., i_k$ into the positions $1, ..., k$ without changing the relative order of the remaining *columns. Now consider the set* $U_{i_1...i_k}$ *of all* $[A] \in G(k,n)$ *such that any representative* A *has its* k *rows indexed by* $i_1, ..., i_k$ *linearly independent. The permutation induces an obvious 1-1 onto map* $\widetilde{\sigma_{i_1...i_k}}$ *from* $U_{i_1...i_k}$ *onto* $U = U_{1...k}$*. We now have maps* $\Psi_{i_1...i_k} : U_{i_1...i_k} \to M_{n-k\times k}(\mathbb{C}) \cong \mathbb{C}^{k(n-k)}$ *given by composition* $\Psi_{i_1...i_k} := \Psi \circ \widetilde{\sigma_{i_1...i_k}}$*. These maps form an atlas* $\{\Psi_{i_1...i_k}, U_{i_1...i_k}\}$ *for* $G(k,n)$ *that turns out to be a holomorphic atlas (biholomorphic transition maps) and so gives* $G(k,n)$ *the structure of a complex manifold called the* ***Grassmann manifold*** *of complex k-planes in* $\mathbb{C}^n$*.*

Definition 14.5 *A complex 1-manifold (so real dimension is 2) is called a* ***Riemann surface****.*

If S is a subset of a complex manifold M such that near each $p_0 \in S$ there exists a holomorphic chart $U, \psi = (z^1, ..., z^n)$ such that $0 \in S \cap U$ if and only if $z^{k+1}(p) = \cdots = z^n(p) = 0$ then the coordinates $z^1, ..., z^k$ restricted to $U \cap S$ give a chart on the set S and the set of all such charts gives S the structure of a complex manifold. In this case we call S a **complex submanifold** of M.

Definition 14.6 *In the same way as we defined differentiability for real manifolds we define the notion of a* ***holomorphic map*** *(resp.* ***antiholomorphic map****) from one complex manifold to another. Note however, that we must use holomorphic charts for the definition.*

The proof of the following lemma is straightforward.

Lemma 14.2 *Let* $\psi : U \to \mathbb{C}^n$ *be a holomorphic chart with* $p \in U$*. Then writing* $\psi = (z^1, ..., z^n)$ *and* $z^k = x^k + \mathrm{i}y^k$ *we have that the map* $J_p : T_pM \to T_pM$ *defined by*

$$J_p \left.\frac{\partial}{\partial x^i}\right|_p = \left.\frac{\partial}{\partial y^i}\right|_p$$
$$J_p \left.\frac{\partial}{\partial y^i}\right|_p = -\left.\frac{\partial}{\partial x^i}\right|_p$$

is well defined independent of the choice of coordinates.

The maps J_p combine to give a bundle map $J : TM \to TM$ and so an almost complex structure on M called the almost complex structure induced by the holomorphic atlas.

Definition 14.7 *An almost complex structure* J *on* M *is said to be* ***integrable*** *if there is an holomorphic atlas giving the map* J *as the induced almost complex structure. That is, if there is a family of admissible charts* $\psi_\alpha : U_\alpha \to \mathbb{R}^{2n}$ *such that after identifying* $\mathbb{R}^{2n}$ *with* $\mathbb{C}^n$ *the charts form a holomorphic atlas with* J *the induced almost complex structure. In this case, we call* J *a* ***complex structure****.*

14.3 Complex Tangent Structures

Let $\mathcal{F}_p(\mathbb{C})$ denote the algebra of germs of *complex* valued smooth functions at p on a complex n-manifold M thought of as a smooth real $2n-$manifold with real tangent bundle TM. Let $\mathrm{Der}_p(\mathcal{F})$ be the space of derivations this algebra. It is not hard to see that this space is isomorphic to the complexified tangent space $T_pM_{\mathbb{C}} = \mathbb{C}\otimes T_pM$. The (complex) algebra of germs of holomorphic functions at a point p in a complex manifold is denoted $\mathcal{O}_p$ and the set of derivations of this algebra denoted $\mathrm{Der}_p(\mathcal{O})$. We also have the algebra of germs of antiholomorphic functions at p which is $\overline{\mathcal{O}}_p$ and also $\mathrm{Der}_p(\overline{\mathcal{O}})$, the derivations of this algebra.

If $\psi : U \to \mathbb{C}^n$ is a holomorphic chart then writing $\psi = (z^1, ..., z^n)$ and $z^k = x^k + \mathrm{i}y^k$ we have the differential operators at $p \in U$:

$$\left\{ \left.\frac{\partial}{\partial z^i}\right|_p , \left.\frac{\partial}{\partial \bar{z}^i}\right|_p \right\}$$

(now transferred to the manifold). To be pedantic about it, we now denote the coordinates on $\mathbb{C}^n$ by $w_i = u_i + \mathrm{i}v_i$ and then

$$\left.\frac{\partial}{\partial z^i}\right|_p f := \left.\frac{\partial f \circ \psi^{-1}}{\partial w^i}\right|_{\psi(p)}$$

$$\left.\frac{\partial}{\partial \bar{z}^i}\right|_p f := \left.\frac{\partial f \circ \psi^{-1}}{\partial \bar{w}^i}\right|_{\psi(p)}$$

Thought of as derivations, these span $\mathrm{Der}_p(\mathcal{F})$ but we have also seen that they span the complexified tangent space at p. In fact, we have the following:

$$\begin{aligned}
T_pM_{\mathbb{C}} &= \mathrm{span}_{\mathbb{C}} \left\{ \left.\frac{\partial}{\partial z^i}\right|_p , \left.\frac{\partial}{\partial \bar{z}^i}\right|_p \right\} = \mathrm{Der}_p(\mathcal{F}) \\
T_pM^{1,0} &= \mathrm{span}_{\mathbb{C}} \left\{ \left.\frac{\partial}{\partial z^i}\right|_p \right\} \\
&= \{v \in \mathrm{Der}_p(\mathcal{F}) : vf = 0 \text{ for all } f \in \overline{\mathcal{O}}_p\} \\
T_pM^{0,1} &= \mathrm{span}_{\mathbb{C}} \left\{ \left.\frac{\partial}{\partial \bar{z}^i}\right|_p \right\} \\
&= \{v \in \mathrm{Der}_p(\mathcal{F}) : vf = 0 \text{ for all } f \in \mathcal{O}_p\}
\end{aligned}$$

and of course

$$T_pM = \mathrm{span}_{\mathbb{R}} \left\{ \left.\frac{\partial}{\partial x^i}\right|_p , \left.\frac{\partial}{\partial y^i}\right|_p \right\}.$$

The reader should go back and check that the above statements are consistent with our definitions as long as we view the $\left.\frac{\partial}{\partial z^i}\right|_p, \left.\frac{\partial}{\partial \bar{z}^i}\right|_p$ not only as the algebraic objects constructed above but also as derivations. Also, the definitions of

$T_pM^{1,0}$ and $T_pM^{0,1}$ are independent of the holomorphic coordinates since we also have

$$T_p^{1,0}M = \ker\{J_p : T_pM \to T_pM\}$$

14.4 The holomorphic tangent map.

We leave it to the reader to verify that the constructions that we have at each tangent space globalize to give natural vector bundles $TM_{\mathbb{C}}, TM^{1,0}$ and $TM^{0,1}$ (all with M as base space).

Let M and N be complex manifolds and let $f : M \to N$ be a smooth map. The tangent map extends to a map of the complexified bundles $Tf : TM_{\mathbb{C}} \to TN_{\mathbb{C}}$. Now $TM_{\mathbb{C}} = TM^{1,0} \oplus TM^{0,1}$ and similarly $TM_{\mathbb{C}} = TN^{1,0} \oplus TN^{0,1}$. If f is holomorphic then $Tf(T_p^{1,0}M) \subset T_{f(p)}^{1,0}N$. In fact, it is easily verified that the statement that $Tf(T_pM^{1,0}) \subset T_{f(p)}N^{1,0}$ is equivalent to the statement that the Cauchy-Riemann equations are satisfied by the local representative of F in any holomorphic chart. As a result we have

Proposition 14.1 $Tf(T_pM^{1,0}) \subset T_{f(p)}N^{1,0}$ *if and only if* f *is a holomorphic map.*

The map given by the restriction $T_pf : T_pM^{1,0} \to T_{fp}N^{1,0}$ is called the **holomorphic tangent** map at p. Of course, these maps concatenate to give a bundle map

14.5 Dual spaces

Let M, J be a complex manifold. The dual of $T_pM_{\mathbb{C}}$ is $T_p^*M_{\mathbb{C}} = \mathbb{C} \otimes T_p^*M$. Now the map J has a dual bundle map $J^* : T^*M_{\mathbb{C}} \to T^*M_{\mathbb{C}}$ that must also satisfy $J^* \circ J^* = -1$ and so we have the at each $p \in M$ the decomposition by eigenspaces

$$T_p^*M_{\mathbb{C}} = T_p^*M^{1,0} \oplus T_p^*M^{0,1}$$

corresponding to the eigenvalues $\pm$i.

Definition 14.8 *The space* $T_p^*M^{1,0}$ *is called the space of holomorphic covectors at* p *while* $T_p^*M^{0,1}$ *is the space of antiholomorphic covectors at* p.

We now choose a holomorphic chart $\psi : U \to \mathbb{C}^n$ at p. Writing $\psi = (z^1, ..., z^n)$ and $z^k = x^k + \mathrm{i}y^k$ we have the 1-forms

$$dz^k = dx^k + \mathrm{i}dy^k$$

and

$$d\bar{z}^k = dx^k - \mathrm{i}dy^k.$$

Equivalently, the pointwise definitions are $dz^k\big|_p = dx^k\big|_p + \mathrm{i}\ dy^k\big|_p$ and $d\bar{z}^k\big|_p = dx^k\big|_p - \mathrm{i}\ dy^k\big|_p$. Notice that we have the expected relations:

$$\begin{aligned}
dz^k(\frac{\partial}{\partial z^i}) &= (dx^k + \mathrm{i}dy^k)(\frac{1}{2}\frac{\partial}{\partial x^i} - \mathrm{i}\frac{1}{2}\frac{\partial}{\partial y^i}) \\
&= \frac{1}{2}\delta^k_j + \frac{1}{2}\delta^k_j = \delta^k_j \\
dz^k(\frac{\partial}{\partial \bar{z}^i}) &= (dx^k + \mathrm{i}dy^k)(\frac{1}{2}\frac{\partial}{\partial x^i} + \mathrm{i}\frac{1}{2}\frac{\partial}{\partial y^i}) \\
&= 0 \\
&\text{and similarly} \\
d\bar{z}^k(\frac{\partial}{\partial \bar{z}^i}) &= \delta^k_j \text{ and } d\bar{z}^k(\frac{\partial}{\partial z^i}) = \delta^k_j.
\end{aligned}$$

Let us check the action of J^* on these forms:

$$\begin{aligned}
J^*(dz^k)(\frac{\partial}{\partial z^i}) &= J^*(dx^k + \mathrm{i}dy^k)(\frac{\partial}{\partial z^i}) \\
&= (dx^k + \mathrm{i}dy^k)(J\frac{\partial}{\partial z^i}) \\
&= \mathrm{i}(dx^k + \mathrm{i}dy^k)\frac{\partial}{\partial z^i} \\
&= \mathrm{i}dz^k(\frac{\partial}{\partial z^i})
\end{aligned}$$

and

$$\begin{aligned}
J^*(dz^k)(\frac{\partial}{\partial \bar{z}^i}) &= dz^k(J\frac{\partial}{\partial \bar{z}^i}) \\
&= -\mathrm{i}dz^k(\frac{\partial}{\partial \bar{z}^i}) = 0 = \\
&= \mathrm{i}dz^k(\frac{\partial}{\partial \bar{z}^i}).
\end{aligned}$$

Thus we conclude that $dz^k\big|_p \in T_p^*M^{1,0}$. A similar calculation shows that $d\bar{z}^k\big|_p \in T_p^*M^{0,1}$ and in fact

$$\begin{aligned}
T_p^*M^{1,0} &= \operatorname{span}\left\{ dz^k\big|_p : k = 1, ..., n\right\} \\
T_p^*M^{0,1} &= \operatorname{span}\left\{ d\bar{z}^k\big|_p : k = 1, ..., n\right\}
\end{aligned}$$

and $\{ dz^1\big|_p, ..., dz^n|_p, d\bar{z}^1\big|_p, ..., d\bar{z}^n|_p\}$ is a basis for $T_p^*M_{\mathbb{C}}$.

Remark 14.1 *If we don't specify base points then we are talking about fields (over some open set) that form a basis for each fiber separately. These are called frame fields (e.g.* $\frac{\partial}{\partial z^i}, \frac{\partial}{\partial \bar{z}^i}$*) or coframe fields (e.g.* $dz^k, d\bar{z}^k$*).*

14.6 Examples

Under construction ASP

14.7 The holomorphic inverse and implicit functions theorems.

Let $(z^1, ..., z^n)$ and $(w^1, ..., w^m)$ be local coordinates on complex manifolds M and N respectively. Consider a smooth map $f : M \to N$. We suppose that $p \in M$ is in the domain of $(z^1, ..., z^n)$ and that $q = f(p)$ is in the domain of the coordinates $(w^1, ..., w^m)$. Writing $z^i = x^i + \mathrm{i}y^i$ and $w^i = u^i + \mathrm{i}v^i$ we have the following Jacobian matrices:

1. If we consider the underlying real structures then we have the Jacobian given in terms of the frame $\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^i}$ and $\frac{\partial}{\partial u^i}, \frac{\partial}{\partial v^i}$

$$J_p(f) = \begin{bmatrix} \frac{\partial u^1}{\partial x^1}(p) & \frac{\partial u^1}{\partial y^1}(p) & \frac{\partial u^1}{\partial x^2}(p) & \frac{\partial u^1}{\partial y^2}(p) & \cdots \\ \frac{\partial v^1}{\partial x^1}(p) & \frac{\partial v^1}{\partial y^1}(p) & \frac{\partial v^1}{\partial x^2}(p) & \frac{\partial v^1}{\partial y^2}(p) & \cdots \\ \frac{\partial u^2}{\partial x^1}(p) & \frac{\partial u^2}{\partial y^1}(p) & \frac{\partial u^2}{\partial x^2}(p) & \frac{\partial u^2}{\partial y^2}(p) & \cdots \\ \frac{\partial v^2}{\partial x^1}(p) & \frac{\partial v^2}{\partial y^1}(p) & \frac{\partial v^2}{\partial x^2}(p) & \frac{\partial v^2}{\partial y^2}(p) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

2. With respect to the bases $\frac{\partial}{\partial z^i}, \frac{\partial}{\partial \bar{z}^i}$ and $\frac{\partial}{\partial w^i}, \frac{\partial}{\partial \bar{w}^i}$ we have

$$J_{p,\mathbb{C}}(f) = \begin{bmatrix} J_{11} & J_{12} & \cdots \\ J_{12} & J_{22} & \\ \vdots & & \end{bmatrix}$$

where the J_{ij} are blocks of the form

$$\begin{bmatrix} \frac{\partial w^i}{\partial z^j} & \frac{\partial w^i}{\partial \bar{z}^j} \\ \frac{\partial \bar{w}^i}{\partial z^j} & \frac{\partial \bar{w}^i}{\partial \bar{z}^j} \end{bmatrix}.$$

If f is holomorphic then these block reduce to the form

$$\begin{bmatrix} \frac{\partial w^i}{\partial z^j} & 0 \\ 0 & \frac{\partial \bar{w}^i}{\partial \bar{z}^j} \end{bmatrix}.$$

It is convenient to put the frame fields in the order $\frac{\partial}{\partial z^1}, ..., \frac{\partial}{\partial z^n}, \frac{\partial}{\partial \bar{z}^1}, ..., \frac{\partial}{\partial \bar{z}^n}$ and similarly for the $\frac{\partial}{\partial w^i}, \frac{\partial}{\partial \bar{w}^i}$. In this case we have for holomorphic f

$$\mathcal{J}_{p,\mathbb{C}}(f) = \begin{bmatrix} J^{1,0} & 0 \\ 0 & \overline{J^{1,0}} \end{bmatrix}$$

where

$$J^{1,0}(f) = \left[\frac{\partial w^i}{\partial z^j}\right]$$
$$\overline{J^{1,0}}(f) = \left[\frac{\partial \bar{w}^i}{\partial \bar{z}^j}\right].$$

We shall call a basis arising from a holomorphic coordinate system "separated" when arranged this way. Note that $J^{1,0}$ is just the Jacobian of the holomorphic tangent map $T^{1,0}f : T^{1,0}M \to T^{1,0}N$ with respect to this the **holomorphic frame** $\frac{\partial}{\partial z^1}, ..., \frac{\partial}{\partial z^n}$.

We can now formulate the following version of the inverse mapping theorem:

Theorem 14.1 *(1) Let U and V be open sets in $\mathbb{C}^n$ and suppose that the map $f : U \to V$ is holomorphic with $J^{1,0}(f)$ nonsingular at $p \in U$. Then there exists an open set $U_0 \subset U$ containing p such that $f|_{U_0} : U_0 \to f(U_0)$ is a 1-1 holomorphic map with holomorphic inverse. That is, $f|_{U_0}$ is **biholomorphic**.*

(2) Similarly, if $f : U \to V$ is a holomorphic map between open sets of complex manifolds M and N then if $T_p^{1,0}f : T_p^{1,0}M \to T_{fp}^{1,0}N$ is a linear isomorphism then f is a biholomorphic map when restricted to a possibly smaller open set containing p.

We also have a holomorphic version of the implicit mapping theorem.

Theorem 14.2 *(1) Let $f : U \subset \mathbb{C}^n \to V \subset \mathbb{C}^k$ and let the component functions of f be $f_1, ..., f_k$. If $J_p^{1,0}(f)$ has rank k then there are holomorphic functions $g^1, g^2, ..., g^k$ defined near $0 \in \mathbb{C}^{n-k}$ such that*

$$f(z^1, ..., z^n) = p$$
$$\Leftrightarrow$$
$$z^j = g^j(z^{k+1}, ..., z^n) \text{ for } j = 1, .., k$$

(2) If $f : M \to N$ is a holomorphic map of complex manifolds and if for fixed $q \in N$ we have that each $p \in f^{-1}(q)$ is regular in the sense that $T_p^{1,0}f : T_p^{1,0}M \to T_q^{1,0}N$ is surjective, then $S := f^{-1}(q)$ is a complex submanifold of (complex) dimension $n - k$.

Example 14.5 *The map $\varphi : \mathbb{C}^{n+1} \to \mathbb{C}$ given by $(z^1, ..., z^{n+1}) \mapsto (z^1)^2 + \cdots + (z^{n+1})^2$ has Jacobian at any $(z^1, ..., z^{n+1})$ given by*

$$\left[\begin{array}{cccc} 2z^1 & 2z^2 & \cdots & 2z^{n+1} \end{array}\right]$$

*which has rank 1 as long as $(z^1, ..., z^{n+1}) \neq 0$. Thus $\varphi^{-1}(1)$ is a complex submanifold of $\mathbb{C}^{n+1}$ having complex dimension n. **Warning**: This manifold is not the same as the sphere given by $\left|z^1\right|^2 + \cdots + \left|z^{n+1}\right|^2 = 1$ which is a* real *submanifold of $\mathbb{C}^{n+1} \cong \mathbb{R}^{2n+2}$ of real dimension $2n + 1$.*

Chapter 15

Lie Groups and Lie Algebras

15.1 Lie Algebras

Let $\mathbb{F}$ denote one of the fields $\mathbb{R}$ or $\mathbb{C}$. In definition 7.2 we defined a real Lie algebra $\mathfrak{g}$ as a real algebra with a skew symmetric (bilinear) product (the Lie bracket), usually denoted with a bracket $v, w \mapsto [v, w]$, such that the Jacobi identity holds

$$[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0 \text{ for all } x, y, z \in \mathfrak{g}. \qquad \text{(Jacobi Identity)}$$

We also have the notion of a complex Lie algebra defined analogously.

Remark 15.1 *We will assume that all the Lie algebras we study are finite dimensional unless otherwise indicated.*

Let V be a finite dimensional vector space and recall that $\mathfrak{gl}(\mathrm{V},\mathbb{F})$ is the set of all $\mathbb{F}-$linear maps $\mathrm{V} \to \mathrm{V}$. The space $\mathfrak{gl}(\mathrm{V},\mathbb{F})$ is also denoted $\mathrm{Hom}_{\mathbb{F}}(\mathrm{V}, \mathrm{V})$ or $L_{\mathbb{F}}(\mathrm{V}, \mathrm{V})$ although in the context of Lie algebras we take $\mathfrak{gl}(\mathrm{V},\mathbb{F})$ as the preferred notation. We give $\mathfrak{gl}(\mathrm{V},\mathbb{F})$ its natural Lie algebra structure where the bracket is just the commutator bracket

$$[A, B] := A \circ B - B \circ A.$$

If the field involved is either irrelevant or known from context we will just write $\mathfrak{gl}(\mathrm{V})$. Also, we often identify $\mathfrak{gl}(\mathbb{F}^n)$ with the matrix Lie algebra $\mathbb{M}_{nxn}(\mathbb{F})$ with the bracket $AB - BA$.

For a Lie algebra $\mathfrak{g}$ we can associate to every basis $v_1, ..., v_n$ for $\mathfrak{g}$ the **structure constants** c_{ij}^k which are defined by

$$[v_i, v_j] = \sum_k c_{ij}^k v_k.$$

It then follows from the skew symmetry of the Lie bracket and the Jacobi identity it follow that the structure constants satisfy

$$\begin{array}{ll} \text{i)} & c_{ij}^k = -c_{ji}^k \\ \text{ii)} & \sum_k c_{rs}^k c_{kt}^i + c_{st}^k c_{kr}^i + c_{tr}^k c_{ks}^i = 0 \end{array} \tag{15.1}$$

Given a real Lie algebra $\mathfrak{g}$ we can extend to a complex Lie algebra $\mathfrak{g}_{\mathbb{C}}$ by defining as $\mathfrak{g}_{\mathbb{C}}$ the complexification $\mathfrak{g}_{\mathbb{C}} := \mathfrak{g} \otimes_{\mathbb{R}} \mathbb{C}$ and then extending the bracket by requiring

$$[v \otimes 1, w \otimes 1] = [v, w] \otimes 1.$$

Then $\mathfrak{g}$ can be identified with its image under the embedding map $v \mapsto v \otimes 1$. In practice one often omits the symbol $\otimes$ and the aforementioned identification of $\mathfrak{g}$ as a subspace of $\mathfrak{g}_{\mathbb{C}}$ the complexification just amounts to allowing complex coefficients.

Notation 15.1 *Given two subsets S_1 and S_2 of a Lie algebra $\mathfrak{g}$ we let $[\ S_1, S_2]$ denote the linear span of the set defined by $\{[x, y] : x \in S_1$ and $y \in S_2\}$. Also, let $S_1 + S_2$ denote the vector space of all $x + y : x \in S_1$ and $y \in S_2$.*

It is easy to verify that the following relations hold:

1. $[S_1 + S_2, S] \subset [S_1, S] + [S_2, S]$
2. $[S_1, S_2] = [S_2, S_1]$
3. $[S, [S_1, S_2]] \subset [[S, S_1], S_2] + [S_1, [S, S_2]]$

where S_1, S_2 and S are subsets of a Lie algebra $\mathfrak{g}$.

Definition 15.1 *A vector subspace $\mathfrak{a} \subset \mathfrak{g}$ is called a* ***subalgebra*** *if $[\mathfrak{a}, \mathfrak{a}] \subset \mathfrak{a}$ and an* ***ideal*** *if $[\mathfrak{g}, \mathfrak{a}] \subset \mathfrak{a}$.*

If $\mathfrak{a}$ is a subalgebra of $\mathfrak{g}$ and $v_1, ..., v_k, v_{k+1}, ..., v_n$ is a basis for $\mathfrak{g}$ such that $v_1, ..., v_k$ is a basis for $\mathfrak{a}$ then with respect to this basis the structure constants are such that

$$c_{ij}^s = 0 \text{ for } i, j \leq k \text{ and } s > k.$$

If $\mathfrak{a}$ is also an ideal then we must have

$$c_{ij}^s = 0 \text{ for } i \leq k \text{ and } s > k \text{ with any } j.$$

Remark 15.2 *The numbers c_{ij}^s may be viewed as the components of a an element (a tensor) of $T_{1,1}^1(\mathfrak{g})$.*

Example 15.1 *Let $\mathfrak{su}(2)$ denote the set of all traceless and Hermitian 2×2 complex matrices. This is a Lie algebra under the commutator bracket ($AB - BA$). A commonly used basis for $\mathfrak{su}(2)$ is e_1, e_2, e_3 where*

$$e_1 = \frac{1}{2} \begin{bmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{bmatrix}, \quad e_2 = \frac{1}{2} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad e_2 = \frac{1}{2} \begin{bmatrix} -\mathrm{i} & 0 \\ 0 & \mathrm{i} \end{bmatrix}.$$

The commutation relations satisfied by these matrices are

$$[e_i, e_j] = \epsilon_{ijk} e_k$$

where ϵ_{ijk} is the totally antisymmetric symbol given by

$$\epsilon_{ijk} := \begin{cases} 0 & \text{if } (i,j,k) \text{ is not a permutation of } (1,2,3) \\ 1 & \text{if } (i,j,k) \text{ is an even permutation of } (1,2,3) \\ -1 & \text{if } (i,j,k) \text{ is an odd permutation of } (1,2,3) \end{cases}.$$

Thus in this case the structure constants are $c^k_{ij} = \epsilon_{ijk}$. In physics it is common to use the Pauli matrices defined by $\sigma_i := 2\mathrm{i}e_i$ in terms of which the commutation relations become $[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k$.

Example 15.2 *The Weyl basis for $\mathfrak{gl}(n,\mathbb{R})$ is given by the n^2 matrices e_{sr} defined by*

$$(e_{rs})_{ij} := \delta_{ri}\delta_{sj}.$$

Notice that we are now in a situation where "double indices" will be convenient. For instance, the commutation relations read

$$[e_{ij}, e_{kl}] = \delta_{jk}e_{il} - \delta_{il}e_{kj}$$

while the structure constants are

$$c^{ij}_{sm,kr} = \delta^i_s \delta_{mk} \delta^j_r - \delta^i_k \delta_{rs} \delta^j_m.$$

15.2 Classical complex Lie algebras

If $\mathfrak{g}$ is a real Lie algebra we have seen that the complexification $\mathfrak{g}_{\mathbb{C}}$ is naturally a complex Lie algebra. Is convenient to omit the tensor symbol and use the following convention: Every element of $\mathfrak{g}_{\mathbb{C}}$ may be written at $v + \mathrm{i}w$ for $v, w \in \mathfrak{g}$ and then

$$\begin{aligned}&[v_1 + \mathrm{i}w_1, v_2 + \mathrm{i}w_2] \\ &= [v_1, v_2] - [w_1, w_2] + \mathrm{i}([v_1, w_2] + [w_1, v_2]).\end{aligned}$$

We shall now define a series of complex Lie algebras sometimes denoted by A_n, B_n, C_n and D_n for every integer $n > 0$. First of all, notice that the complexification $\mathfrak{gl}(n,\mathbb{R})_{\mathbb{C}}$ of $\mathfrak{gl}(n,\mathbb{R})$ is really just $\mathfrak{gl}(n,\mathbb{C})$; the set of complex $n \times n$ matrices under the commutator bracket.

The algebra A_n The set of all traceless $n \times n$ matrices is denoted A_{n-1} and also by $\mathfrak{sl}(n,\mathbb{C})$.

We call the readers attention to the following general fact: If $b(.,.)$ is a bilinear form on a complex vector space V then the set of all $A \in \mathfrak{gl}(n,\mathbb{C})$ such that $b(Az, w) + b(z, Aw) = 0$ for every $z, w \in V$ is a Lie subalgebra of $\mathfrak{gl}(n, \mathbb{C})$. This follows from

$$\begin{aligned} b([A, B]z, w) &= b(ABz, w) - b(BAz, w) \\ &= -b(Bz, Aw) + b(Az, Bw) \\ &= b(z, BAw) - b(z, ABw) \\ &= b(z, [B, A]w). \end{aligned}$$

The algebras B_n and D_n Let $m = 2n + 1$ and let $b(.,.)$ be a nondegenerate symmetric bilinear form on an m dimensional complex vector space V. Without loss we may assume $V = \mathbb{C}^m$ and we may take $b(z, w) = \sum_{i=1}^{m} z_i w_i$. We define B_n to be $\mathfrak{o}(m, \mathbb{C})$ where

$$\mathfrak{o}(m, \mathbb{C}) := \{A \in \mathfrak{gl}(m, \mathbb{C}) : b(Az, w) + b(z, Aw) = 0\}.$$

Similarly, for $m = 2n$ we define D_n to be $\mathfrak{o}(m, \mathbb{C})$.

The algebra C_n The algebra associated to a skew-symmetric nondegenerate bilinear form which we may take to be $b(z, w) = \sum_{i=1}^{n} z_i w_{n+i} - \sum_{i=1}^{n} z_{n+i} w_i$ on $\mathbb{C}^{2n}$ we have the symplectic algebra

$$C_n = \mathfrak{sp}(n, \mathbb{C}) := \{A \in \mathfrak{gl}(m, \mathbb{C}) : b(Az, w) + b(z, Aw) = 0\}.$$

15.2.1 Basic Definitions

The expected theorems hold for homomorphisms; the image $\mathrm{img}(\sigma) := \sigma(\mathfrak{a})$ of a homomorphism $\sigma : \mathfrak{a} \to \mathfrak{b}$ is a subalgebra of $\mathfrak{b}$ and the kernel $\ker(\sigma)$ is an ideal of $\mathfrak{a}$.

Definition 15.2 *Let $\mathfrak{h}$ be an ideal in $\mathfrak{g}$. On the quotient vector space $\mathfrak{g}/\mathfrak{h}$ with quotient map π we can define a Lie bracket in the following way: For $\bar{v}, \bar{w} \in \mathfrak{g}/\mathfrak{h}$ choose $v, w \in \mathfrak{g}$ with $\pi(v) = \bar{v}$ and $\pi(w) = \bar{w}$ we define*

$$[\bar{v}, \bar{w}] := \pi([v, w]).$$

*We call $\mathfrak{g}/\mathfrak{h}$ with this bracket the **quotient Lie algebra**.*

Exercise 15.1 *Show that the bracket defined in the last definition is well defined.*

Given two linear subspaces $\mathfrak{a}$ and $\mathfrak{b}$ of $\mathfrak{g}$ the (not necessarily direct) sum $\mathfrak{a} + \mathfrak{b}$ is just the space of all elements in $\mathfrak{g}$ of the form $a + b$ where $a \in \mathfrak{a}$ and $b \in \mathfrak{b}$. It is not hard to see that if $\mathfrak{a}$ and $\mathfrak{b}$ are ideals in $\mathfrak{g}$ then so is $\mathfrak{a} + \mathfrak{b}$.

Exercise 15.2 *Show that for $\mathfrak{a}$ and $\mathfrak{b}$ ideals in $\mathfrak{g}$ we have a natural isomorphism $(\mathfrak{a} + \mathfrak{b})/\mathfrak{b} \cong \mathfrak{a}/(\mathfrak{a} \cap \mathfrak{b})$.*

If $\mathfrak{g}$ is a Lie algebra, $\mathfrak{s}$ a subset of $\mathfrak{g}$ then the **centralizer of $\mathfrak{s}$ in $\mathfrak{g}$** is $\mathfrak{z}(\mathfrak{s}) := \{v \in \mathfrak{g} : [v, \mathfrak{b}] = 0 \text{ for all } v \in \mathfrak{s}\}$. If $\mathfrak{a}$ is a (Lie) subalgebra of $\mathfrak{g}$ then the **normalizer** of $\mathfrak{a}$ in $\mathfrak{g}$ is $\mathfrak{n}(\mathfrak{a}) := \{v \in \mathfrak{g} : [v, \mathfrak{a}] \subset \mathfrak{a}\}$. One can check that $\mathfrak{n}(\mathfrak{a})$ is an ideal in $\mathfrak{g}$.

There is also a Lie algebra product. Namely, if $\mathfrak{a}$ and $\mathfrak{b}$ are Lie algebras, then we can define a Lie bracket on $\mathfrak{a} \times \mathfrak{b}$ by

$$[(a_1, a_2), (b_1, b_2)] := ([a_1, b_1], [a_2, b_2]).$$

With this bracket $\mathfrak{a} \times \mathfrak{b}$ is a Lie algebra called the Lie algebra **product** of $\mathfrak{a}$ and $\mathfrak{b}$. The subspaces $\mathfrak{a} \times \{0\}$ and $\{0\} \times \mathfrak{b}$ are ideals in $\mathfrak{a} \times \mathfrak{b}$ that are clearly isomorphic to $\mathfrak{a}$ and $\mathfrak{b}$ respectively.

Definition 15.3 *The* ***center*** $\mathfrak{z}(\mathfrak{g})$ *of a Lie algebra* $\mathfrak{g}$ *is the subspace* $\mathfrak{z}(\mathfrak{g}) := \{x \in \mathfrak{g} : [x, y] = 0 \text{ for all } y \in \mathfrak{g}\}$.

15.3 The Adjoint Representation

It is easy to see that $\mathfrak{x}(\mathfrak{g})$ is the kernel of the map $v \to \mathrm{ad}(v)$ where $\mathrm{ad}(v) \in \mathfrak{gl}(\mathfrak{g})$ is given by $\mathrm{ad}(v)(x) := [v, x]$.

Definition 15.4 *A derivation of a Lie algebra* $\mathfrak{g}$ *is a linear map* $D : \mathfrak{g} \to \mathfrak{g}$ *such that*

$$D[v, w] = [Dv, w] + [v, Dw]$$

for all $v, w \in \mathfrak{g}$.

For each $v \in \mathfrak{g}$ the map $\mathrm{ad}(v) : \mathfrak{g} \to \mathfrak{g}$ is actually a derivation of the Lie algebra $\mathfrak{g}$. Indeed, this is exactly the content of the Jacobi identity. Furthermore, it is not hard to check that the space of all derivations of a Lie algebra $\mathfrak{g}$ is a subalgebra of $\mathfrak{gl}(\mathfrak{g})$. In fact, if D_1 and D_2 are derivations of $\mathfrak{g}$ then so is the commutator $D_1 \circ D_2 - D_2 \circ D_1$. We denote this subalgebra of derivations by $Der(\mathfrak{g})$.

Definition 15.5 *A* ***Lie algebra representation*** ρ *of* $\mathfrak{g}$ *on a vector space* V *is a Lie algebra homomorphism* $\rho : \mathfrak{g} \to \mathfrak{gl}(\mathrm{V})$.

One can construct Lie algebra representations in various way from given representations. For example, if $\rho_i : \mathfrak{g} \to \mathfrak{gl}(\mathrm{V}_i)$ $(i = 1, .., k)$ are Lie algebra representations then $\oplus \rho_i : \mathfrak{g} \to \mathfrak{gl}(\oplus_i \mathrm{V}_i)$ defined by

$$(\oplus_i \rho_i)(x)(v_1 \oplus \cdots \oplus v_n) = \rho_1(x)v_1 \oplus \cdots \oplus \rho_1(x)v_n \tag{15.2}$$

for $x \in \mathfrak{g}$ is a Lie algebra representation called the **direct sum representation** of the ρ_i. Also, if one defines

$$\begin{aligned}(\otimes_i \rho_i)(x)(\otimes v_1 \otimes \cdots \otimes v_k) &:= \rho_1(x)v_1 \otimes v_2 \otimes \cdots \otimes v_k \\ &\quad + v_1 \otimes \rho_2(x)v_2 \otimes \cdots \otimes v_k + \cdots + v_1 \otimes v_2 \otimes \cdots \otimes \rho_k(x)v_k\end{aligned}$$

(and extend linearly) then $\otimes_i \rho_i$ is a representation $\otimes_i \rho_i : \mathfrak{g} \to \mathfrak{gl}(\otimes_i \mathrm{V}_i)$ is Lie algebra representation called a **tensor product representation**.

Lemma 15.1 $\mathrm{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$ *is a Lie algebra representation on* $\mathfrak{g}$. *The image of* ad *is contained in the Lie algebra* $\mathrm{Der}(\mathfrak{g})$ *of all derivation of the Lie algebra* $\mathfrak{g}$.

Proof. This follows from the Jacobi identity (as indicated above) and from the definition of ad. ■

Corollary 15.1 $\mathfrak{x}(\mathfrak{g})$ *is an ideal in* $\mathfrak{g}$.

The image $\mathrm{ad}(\mathfrak{g})$ of ad in $\mathrm{Der}(\mathfrak{g})$ is called the adjoint algebra.

Definition 15.6 *The Killing form for a Lie algebra* $\mathfrak{g}$ *is the bilinear form given by*

$$K(X, Y) = Tr(\mathrm{ad}(X) \circ \mathrm{ad}(Y))$$

Lemma 15.2 *For any Lie algebra automorphism* $\vartheta : \mathfrak{g} \to \mathfrak{g}$ *and any* $X \in \mathfrak{g}$ *we have* $\mathrm{ad}(\vartheta X) = \vartheta \mathrm{ad} X \vartheta^{-1}$

Proof. $\mathrm{ad}(\vartheta X)(Y) = [\vartheta X, Y] = [\vartheta X, \vartheta\vartheta^{-1}Y] = \vartheta[X, \vartheta^{-1}Y] = \vartheta \circ \mathrm{ad}X \circ \vartheta^{-1}(Y)$. ■

Clearly $K(X, Y)$ is symmetric in X and Y but we have more identities:

Proposition 15.1 *The Killing forms satisfies identities:*
1) $K([X, Y], Z) = K([Z, X], Y)$ *for all* $X, Y, Z \in \mathfrak{g}$
2) $K(\rho X, \rho Y) = K(X, Y)$ *for any Lie algebra automorphism* $\rho : \mathfrak{g} \to \mathfrak{g}$ *and any* $X, Y \in \mathfrak{g}$.

Proof. For (1) we calculate

$$\begin{aligned}
K([X, Y], Z) &= Tr(\mathrm{ad}([X, Y]) \circ \mathrm{ad}(Z)) \\
&= Tr([\mathrm{ad}X, \mathrm{ad}Y] \circ \mathrm{ad}(Z)) \\
&= Tr(\mathrm{ad}X \circ \mathrm{ad}Y \circ \mathrm{ad}Z - \mathrm{ad}Y \circ \mathrm{ad}X \circ \mathrm{ad}Z) \\
&= Tr(\mathrm{ad}Z \circ \mathrm{ad}X \circ \mathrm{ad}Y - \mathrm{ad}X \circ \mathrm{ad}Z \circ \mathrm{ad}Y) \\
&= Tr([\mathrm{ad}Z, \mathrm{ad}X] \circ \mathrm{ad}Y) \\
&= Tr(\mathrm{ad}[Z, X] \circ \mathrm{ad}Y) = K([Z, X], Y)
\end{aligned}$$

where we have used that $Tr(ABC)$ is invariant under cyclic permutations of A, B, C.

For (2) just observe that

$$\begin{aligned}
K(\rho X, \rho Y) &= Tr(\mathrm{ad}(\rho X) \circ \mathrm{ad}(\rho Y)) \\
&= Tr(\rho \mathrm{ad}(X)\rho^{-1}\rho \mathrm{ad}(Y)\rho^{-1}) \quad \text{(lemma 15.2)} \\
&= Tr(\rho \mathrm{ad}(X) \circ \mathrm{ad}(Y)\rho^{-1}) \\
&= Tr(\mathrm{ad}(X) \circ \mathrm{ad}(Y)) = K(X, Y).
\end{aligned}$$

■

Since $K(X,Y)$ is symmetric in X,Y and so there must be a basis $\{X_i\}_{1\leq i\leq n}$ of $\mathfrak{g}$ for which the matrix (k_{ij}) given by

$$k_{ij} := K(X_i, X_j)$$

is diagonal.

Lemma 15.3 *If $\mathfrak{a}$ is an ideal in $\mathfrak{g}$ then the Killing form of $\mathfrak{a}$ is just the Killing form of $\mathfrak{g}$ restricted to $\mathfrak{a}\times\mathfrak{a}$.*

Proof. Let $\{X_i\}_{1\leq i\leq n}$ be a basis of $\mathfrak{g}$ such that $\{X_i\}_{1\leq i\leq r}$ is a basis for $\mathfrak{a}$. Now since $[\mathfrak{a},\mathfrak{g}]\subset\mathfrak{a}$, the structure constants c^i_{jk} with respect to this basis must have the property that $c^k_{ij}=0$ for $i\leq r<k$ and all j. Thus for $1\leq i,j\leq r$ we have

$$\begin{aligned} K_{\mathfrak{a}}(X_i,X_j) &= Tr(\mathrm{ad}(X_i)\mathrm{ad}(X_j)) \\ &= \sum_{i,j=1}^{r} c^i_{ik}c^k_{ji} = \sum_{i,j=1}^{r} c^i_{ik}c^k_{ji} \\ &\qquad\qquad = K_{\mathfrak{g}}(X_i,X_j). \end{aligned}$$

■

15.4 The Universal Enveloping Algebra

In a Lie algebra $\mathfrak{g}$ the product $[.,.]$ is not associative except in the trivial case that $[.,.]\equiv 0$. On the other hand, associative algebras play an important role in the study of Lie algebras. For one thing, if $\mathfrak{A}$ is an associative algebra then we can introduce the commutator bracket on $\mathfrak{A}$ by

$$[A,B] := AB - BA$$

which gives $\mathfrak{A}$ the structure of Lie algebra. From the other direction, if we start with a Lie algebra $\mathfrak{g}$ then we can construct an associative algebra called the universal enveloping algebra of $\mathfrak{g}$. This is done, for instance, by first forming the full tensor algebra on $\mathfrak{g}$;

$$T(\mathfrak{g}) = \mathbb{F}\oplus\mathfrak{g}\oplus(\mathfrak{g}\otimes\mathfrak{g})\oplus\cdots\oplus\mathfrak{g}^{\otimes k}\oplus\cdots$$

and then dividing out by an appropriate ideal:

Definition 15.7 *Associated to every Lie algebra $\mathfrak{g}$ there is an associative algebra $U(\mathfrak{g})$ called the* ***universal enveloping algebra*** *defined by*

$$U(\mathfrak{g}) := T(\mathfrak{g})/J$$

where J is the ideal generated by elements in $T(\mathfrak{g})$ of the form $X\otimes Y - Y\otimes X - [X,Y]$.

There is the natural map of $\mathfrak{g}$ into $U(\mathfrak{g})$ given by the composition $\pi : \mathfrak{g} \hookrightarrow T(\mathfrak{g}) \to T(\mathfrak{g})/J = U(\mathfrak{g})$. For $v \in \mathfrak{g}$, let v^* denote the image of v under this canonical map.

Theorem 15.1 *Let* V *be a vector space over the field* $\mathbb{F}$. *For every* ρ *representation of* $\mathfrak{g}$ *on* V *there is a corresponding representation* ρ^* *of* $U(\mathfrak{g})$ *on* V *such that for all* $v \in \mathfrak{g}$ *we have*

$$\rho(v) = \rho^*(v^*).$$

This correspondence, $\rho \mapsto \rho^*$ *is a 1-1 correspondence.*

Proof. Given ρ, there is a natural representation $T(\rho)$ on $T(\mathfrak{g})$. The representation $T(\rho)$ vanishes on J since

$$T(\rho)(X \otimes Y - Y \otimes X - [X,Y]) = \rho(X)\rho(Y) - \rho(Y)\rho(X) - \rho([X,Y]) = 0$$

and so $T(\rho)$ descends to a representation ρ^* of $U(\mathfrak{g})$ on $\mathfrak{g}$ satisfying $\rho(v) = \rho^*(v^*)$. Conversely, if σ is a representation of $U(\mathfrak{g})$ on V then we put $\rho(X) = \sigma(X^*)$. The map $\rho(X)$ is linear and a representation since

$$\begin{aligned} \rho([X,Y]) &= \sigma([X,Y]^*) \\ &= \sigma(\pi(X \otimes Y - Y \otimes X)) \\ &= \sigma(X^*Y^* - Y^*X^*) \\ &= \rho(X)\rho(Y) - \rho(Y)\rho(X) \end{aligned}$$

for all $X, Y \in \mathfrak{g}$. ■

Now let $X_1, X_2, ..., X_n$ be a basis for $\mathfrak{g}$ and then form the monomials $X_{i_1}^* X_{i_2}^* \cdots X_{i_r}^*$ in $U(\mathfrak{g})$. The set of all such monomials for a fixed r span a subspace of $U(\mathfrak{g})$, which we denote by $U^r(\mathfrak{g})$. Let c_{ik}^j be the structure constants for the basis $X_1, X_2, ..., X_n$. Then under the map π the structure equations become

$$[X_i^*, X_j^*] = \sum_k c_{ij}^k X_k^*.$$

By using this relation we can replace the spanning set $\mathcal{M}_r = \{X_{i_1}^* X_{i_2}^* \cdots X_{i_r}^*\}$ for $U^r(\mathfrak{g})$ by spanning set $\mathcal{M}_{\leq r}$ for $U^r(\mathfrak{g})$ consisting of all monomials of the form $X_{i_1}^* X_{i_2}^* \cdots X_{i_m}^*$ where $i_1 \leq i_2 \leq \cdots \leq i_m$ and $m \leq r$. In fact one can then concatenate these spanning sets $\mathcal{M}_{\leq r}$ and turns out that these combine to form a basis for $U(\mathfrak{g})$. We state the result without proof:

Theorem 15.2 (Birchoff-Poincarè-Witt) *Let* $e_{i_1 \leq i_2 \leq \cdots \leq i_m} = X_{i_1}^* X_{i_2}^* \cdots X_{i_m}^*$ *where* $i_1 \leq i_2 \leq \cdots \leq i_m$. *The set of all such elements* $\{e_{i_1 \leq i_2 \leq \cdots \leq i_m}\}$ *for all* m *is a basis for* $U(\mathfrak{g})$ *and the set* $\{e_{i_1 \leq i_2 \leq \cdots \leq i_m}\}_{m \leq r}$ *is a basis for the subspace* $U^r(\mathfrak{g})$.

Lie algebras and Lie algebra representations play an important role in physics and mathematics, and as we shall see below, every Lie group has an associated Lie algebra that, to a surprisingly large extent, determines the structure of the

Lie group itself. Let us first explore some of the important abstract properties of Lie algebras. A notion that is useful for constructing Lie algebras with desired properties is that of the **free Lie algebra** $\mathfrak{f}_n$ which is defined to be the quotient of the free algebra on n symbols by the smallest ideal such that we end up with a Lie algebra. Every Lie algebra can be realized as a quotient of one of these free Lie algebras.

Definition 15.8 *The **descending central series** $\{\mathfrak{g}_{(k)}\}$ of a Lie algebra $\mathfrak{g}$ is defined inductively by letting $\mathfrak{g}_{(1)} = \mathfrak{g}$ and then $\mathfrak{g}_{(k+1)} = [\mathfrak{g}_{(k)}, \mathfrak{g}]$.*

The reason for the term "descending" is the that we have the chain of inclusions

$$\mathfrak{g}_{(1)} \supset \cdots \supset \mathfrak{g}_{(k)} \supset \mathfrak{g}_{(k+1)} \supset \cdots$$

From the definition of Lie algebra homomorphism we see that if $\sigma : \mathfrak{g} \to \mathfrak{h}$ is a Lie algebra homomorphism then $\sigma(\mathfrak{g}_{(k)}) \subset \mathfrak{h}_{(k)}$.

Exercise 15.3 (!) *Use the Jacobi identity to prove that for all positive integers i and j, we have $[\mathfrak{g}_{(i)}, \mathfrak{g}_{(i)}] \subset \mathfrak{g}_{(i+j)}$.*

Definition 15.9 *A Lie algebra $\mathfrak{g}$ is called **k-step nilpotent** if and only if $\mathfrak{g}_{(k+1)} = 0$ but $\mathfrak{g}_{(k)} \neq 0$.*

The most studied nontrivial examples are the Heisenberg algebras which are 2-step nilpotent. These are defined as follows:

Example 15.3 *The $2n+1$ dimensional **Heisenberg algebra** $\mathfrak{h}_n$ is the Lie algebra (defined up to isomorphism) with a basis $\{X_1, ..., X_n, Y_1, ..., Y_n, Z\}$ subject to the relations*

$$[X_j, Y_j] = Z$$

and all other brackets of elements from this basis being zero. A concrete realization of $\mathfrak{h}_n$ is given as the set of all $(n+2) \times (n+2)$ matrices of the form

$$\begin{bmatrix} 0 & x_1 & ... & x_n & z \\ 0 & 0 & ... & 0 & y_1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & ... & 0 & y_n \\ 0 & 0 & ... & 0 & 0 \end{bmatrix}$$

where x_i, y_i, z are all real numbers. The bracket is the commutator bracket as is usually the case for matrices. The basis is realized in the obvious way by putting a lone 1 in the various positions corresponding to the potentially nonzero entries. For example,

$$X_1 = \begin{bmatrix} 0 & 1 & ... & 0 & 0 \\ 0 & 0 & ... & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & ... & 0 & 0 \\ 0 & 0 & ... & 0 & 0 \end{bmatrix}$$

and

$$Z = \begin{bmatrix} 0 & 0 & ... & 0 & 1 \\ 0 & 0 & ... & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & ... & 0 & 0 \\ 0 & 0 & ... & 0 & 0 \end{bmatrix}.$$

Example 15.4 *The space of all upper triangular $n \times n$ matrices $\mathfrak{n}_n$ which turns out to be $n-1$ step nilpotent.*

We also have the **free k-step nilpotent Lie algebra** given by the quotient $\mathfrak{f}_{n,k} := \mathfrak{f}_n/(\mathfrak{f}_n)_k$ where $\mathfrak{f}_n$ is the free Lie algebra mentioned above. (notice the difference between $\mathfrak{f}_{n,k}$ and $(\mathfrak{f}_n)_k$.

Lemma 15.4 *Every finitely generated k-step nilpotent Lie algebra is isomorphic to a quotient of the free k-step nilpotent Lie algebra.*

Proof. Suppose that $\mathfrak{g}$ is k-step nilpotent and generated by elements $X_1, ..., X_n$. Let $F_1, ..., F_n$ be the generators of $\mathfrak{f}_n$ and define a map $h : \mathfrak{f}_n \to \mathfrak{g}$ by sending F_i to X_i and extending linearly. This map clearly factors through $\mathfrak{f}_{n,k}$ since $h((\mathfrak{f}_n)_k) = 0$. Then we have a homomorphism $(\mathfrak{f}_n)_k \to \mathfrak{g}$ that is clearly onto and so the result follows. ■

Definition 15.10 *Let $\mathfrak{g}$ be a Lie algebra. We define the commutator series $\{\mathfrak{g}^{(k)}\}$ by letting $\mathfrak{g}^{(1)} = \mathfrak{g}$ and then inductively $\mathfrak{g}^{(k)} = [\mathfrak{g}^{(k-1)}, \mathfrak{g}^{(k-1)}]$. If $\mathfrak{g}^{(k)} = 0$ for some positive integer k, then we call $\mathfrak{g}$ a **solvable** Lie algebra.*

Clearly, the statement $\mathfrak{g}^{(2)} = 0$ is equivalent to the statement that $\mathfrak{g}$ is abelian. Another simple observation is that $\mathfrak{g}^{(k)} \subset \mathfrak{g}_{(k)}$ so that nilpotency implies solvability.

Exercise 15.4 (!) *Every subalgebra and every quotient algebra of a solvable Lie algebra is solvable. In particular, the homomorphic image of a solvable Lie algebra is solvable. Conversely, if $\mathfrak{a}$ is a solvable ideal in $\mathfrak{g}$ and $\mathfrak{g}/\mathfrak{a}$ is solvable, then $\mathfrak{g}$ is solvable. Hint: Use that $(\mathfrak{g}/\mathfrak{a})^{(j)} = \mathfrak{g}^{(j)}/\mathfrak{a}$.*

It follows from this exercise that we have

Corollary 15.2 *Let $h : \mathfrak{g} \to \mathfrak{g}$ be a Lie algebra homomorphism. If $\operatorname{img}(h) := h(\mathfrak{g})$ and $\ker(h)$ are both solvable then $\mathfrak{g}$ is solvable. In particular, if $\operatorname{img}(\mathrm{ad}) := \mathrm{ad}(\mathfrak{g})$ is solvable then so is $\mathfrak{g}$.*

Lemma 15.5 *If $\mathfrak{a}$ is a nilpotent ideal in $\mathfrak{g}$ contained in the center $\mathfrak{z}(\mathfrak{g})$ and if $\mathfrak{g}/\mathfrak{a}$ is nilpotent then $\mathfrak{g}$ is nilpotent.*

Proof. First, the reader can verify that $(\mathfrak{g}/\mathfrak{a})_{(j)} = \mathfrak{g}_{(j)}/\mathfrak{a}$. Now if $\mathfrak{g}/\mathfrak{a}$ is nilpotent then $\mathfrak{g}_{(j)}/\mathfrak{a} = 0$ for some j and so $\mathfrak{g}_{(j)} \subset \mathfrak{a}$ and if this is the case then we have $\mathfrak{g}_{(j+1)} = [\mathfrak{g}, \mathfrak{g}_{(j)}] \subset [\mathfrak{g}, \mathfrak{a}] = 0$. (Here we have $[\mathfrak{g}, \mathfrak{a}] = 0$ since $\mathfrak{a} \subset \mathfrak{z}(\mathfrak{g})$.) Thus $\mathfrak{g}$ is nilpotent. ■

Trivially, the center $\mathfrak{z}(\mathfrak{g})$ of a Lie algebra a solvable ideal.

Corollary 15.3 *Let $h : \mathfrak{g} \to \mathfrak{g}$ be a Lie algebra homomorphism. If* $\operatorname{img}(\mathrm{ad}) := \mathrm{ad}(\mathfrak{g})$ *is nilpotent then* $\mathfrak{g}$ *is nilpotent.*

Proof. Just use the fact that $\ker(\mathrm{ad}) = \mathfrak{z}(\mathfrak{g})$. ■

Theorem 15.3 *The sum of any family of solvable ideals in $\mathfrak{g}$ is a solvable ideal. Furthermore, there is a unique maximal solvable ideal that is the sum of all solvable ideals in $\mathfrak{g}$.*

Sketch of proof. The proof is a maximality argument based on the following idea : If $\mathfrak{a}$ and $\mathfrak{b}$ are solvable then $\mathfrak{a} \cap \mathfrak{b}$ is an ideal in the solvable $\mathfrak{a}$ and so is solvable. It is easy to see that $\mathfrak{a} + \mathfrak{b}$ is an ideal. We have by exercise 15.2 $(\mathfrak{a} + \mathfrak{b})/\mathfrak{b} \cong \mathfrak{a}/(\mathfrak{a} \cap \mathfrak{b})$. Since $\mathfrak{a}/(\mathfrak{a} \cap \mathfrak{b})$ is a homomorphic image of $\mathfrak{a}$ we see that $\mathfrak{a}/(\mathfrak{a} \cap \mathfrak{b}) \cong (\mathfrak{a} + \mathfrak{b})/\mathfrak{b}$ is solvable. Thus by our previous result $\mathfrak{a} + \mathfrak{b}$ is solvable. ■

Definition 15.11 *The maximal solvable ideal in $\mathfrak{g}$ whose existence is guaranteed by the last theorem is called the* ***radical*** *of $\mathfrak{g}$ and is denoted $rad(\mathfrak{g})$*

Definition 15.12 *A Lie algebra $\mathfrak{g}$ is called* ***simple*** *if it contains no ideals other than $\{0\}$ and $\mathfrak{g}$. A Lie algebra $\mathfrak{g}$ is called* ***semisimple*** *if it contains no abelian ideals (other than $\{0\}$).*

Theorem 15.4 (Levi decomposition) *Every Lie algebra is the semi-direct sum of its radical and a semisimple Lie algebra.*

Define semi-direct sum before this.

15.5 The Adjoint Representation of a Lie group

Definition 15.13 *Fix an element $g \in G$. The map $C_g : G \to G$ defined by $C_g(x) = gxg^{-1}$ is called conjugation and the tangent map $T_eC_g : \mathfrak{g} \to \mathfrak{g}$ is denoted Ad_g and called the* ***adjoint map***.

Proposition 15.2 *$C_g : G \to G$ is a Lie group homomorphism.*

The proof is easy.

Proposition 15.3 *The map $C : g \mapsto C_g$ is a Lie group homomorphism $G \to \mathrm{Aut}(G)$.*

The image of the map C inside $\mathrm{Aut}(G)$ is a Lie subgroup called the group of inner automorphisms and is denoted by $\mathrm{Int}(G)$.

Using proposition 12.3 we get the following

Corollary 15.4 $\mathrm{Ad}_g : \mathfrak{g} \to \mathfrak{g}$ *is Lie algebra homomorphism.*

Proposition 15.4 *The map $\mathrm{Ad} : g \mapsto \mathrm{Ad}_g$ is a homomorphism $G \to GL(\mathfrak{g})$ which is called the* ***adjoint representation*** *of G.*

Proof. We have

$$\begin{aligned}\mathrm{Ad}(g_1g_2) &= T_eC_{g_1g_2} = T_e(C_{g_1} \circ C_{g_2}) \\ &= T_eC_{g_1} \circ T_eC_{g_2} = \mathrm{Ad}_{g_1} \circ \mathrm{Ad}_{g_2}\end{aligned}$$

which shows that Ad is a group homomorphism. The smoothness follows from the following lemma applied to the map $C : (g, x) \mapsto C_g(x)$. ■

Lemma 15.6 *Let $f : M \times N \to N$ be a smooth map and define the partial map at $x \in M$ by $f_x(y) = f(x, y)$. Suppose that for every $x \in M$ the point y_0 is fixed by f_x:*

$$f_x(y_0) = y_0 \text{ for all } x.$$

The the map $A_{y_0} : x \mapsto T_{y_0} f_x$ is a smooth map from M to $GL(T_{y_0}N)$.

Proof. It suffices to show that A_{y_0} composed with an arbitrary coordinate function from some atlas of charts on $GL(T_{y_0}N)$ is smooth. But $GL(T_{y_0}N)$ has an atlas consisting of a single chart. Namely, choose a basis $v_1, v_2, ..., v_n$ of $T_{y_0}N$ and let $v^1, v^2, ..., v^n$ the dual basis of $T^*_{y_0}N$, then $\chi^i_j : A \mapsto v^i(Av_j)$ is a typical coordinate function. Now we compose;

$$\begin{aligned}\chi^i_j \circ A_{y_0}(x) &= v^i(A_{y_0}(x)v_j) \\ &= v^i(T_{y_0}f_x \cdot v_j).\end{aligned}$$

Now it is enough to show that $T_{y_0}f_x \cdot v_j$ is smooth in x. But this is just the composition the smooth maps $M \to TM \times TN \cong T(M \times N) \to T(N)$ given by

$$\begin{aligned}&x \mapsto ((x, 0), (y_0, v_j)) \mapsto (\partial_1 f)(x, y_0) \cdot 0 + (\partial_2 f)(x, y_0) \cdot v_j \\ &\quad = T_{y_0}f_x \cdot v_j.\end{aligned}$$

(The reader might wish to review the discussion leading up to lemma 4.4). ■

Recall that for $v \in \mathfrak{g}$ we have the associated left invariant vector field L^v as well as the right invariant R^v. Using this notation we have

Lemma 15.7 *Let $v \in \mathfrak{g}$. Then $L^v(x) = R^{\mathrm{Ad}(x)v}$.*

Proof. $L^v(x) = T_e(L_x) \cdot v = T(R_x)T(R_{x^{-1}})T_e(L_x) \cdot v = T(R_x)T(R_{x^{-1}} \circ L_x) \cdot v = R^{\mathrm{Ad}(x)v}$. ■

We have already defined the group $\mathrm{Aut}(G)$ and the subgroup $\mathrm{Int}(G)$. We have also defined $\mathrm{Aut}(\mathfrak{g})$ which has the subgroup $\mathrm{Int}(\mathfrak{g}) := Ad(G)$.

We now go one step further and take the differential of Ad.

Definition 15.14 *For a Lie group G with Lie algebra $\mathfrak{g}$ define the* ***adjoint representation of*** *$\mathfrak{g}$, a map* $\mathrm{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$ *by*

$$\mathrm{ad} = T_e \,\mathrm{Ad}$$

The following proposition shows that the current definition of ad agrees with that given previously for abstract Lie algebras:

Proposition 15.5 $\mathrm{ad}(v)w = [v, w]$ *for all* $v, w \in \mathfrak{g}$.

Proof. Let $v^1, ..., v^n$ be a basis for $\mathfrak{g}$ so that $\mathrm{Ad}(x)w = \sum a_i(x)v^i$ for some functions a_i. Then we have

$$\begin{aligned}\mathrm{ad}(v)w &= T_e(\mathrm{Ad}(.)w)v \\ &= d(\sum a_i(.)v^i)v \\ &= \sum (\,da_i|_e\, v)v^i \\ &= \sum (L^v a_i)(e)v^i\end{aligned}$$

On the other hand, by lemma 15.7

$$\begin{aligned}L^w(x) = R^{\mathrm{Ad}(x)w} &= R(\sum a_i(x)v^i) \\ &= \sum a_i(x)R^{v^i}(x)\end{aligned}$$

Then we have

$$[L^v, L^w] = [L^v, \sum a_i()R^{v^i}()] = 0 + \sum L^v(a_i)R^{v^i}.$$

Finally, we have

$$\begin{aligned}[w, v] &= [L^w, L^v](e) \\ &= \sum L^v(a_i)(e)R^{v^i}(e) = \sum L^v(a_i)(e)v^i \\ &\mathrm{ad}(v)w.\end{aligned}$$

■

The map $\mathrm{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g}) = \mathrm{End}(T_eG)$ is given as the tangent map at the identity of Ad which is a Lie algebra homomorphism. Thus by proposition 12.3 we have obtain

Proposition 15.6 $\mathrm{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$ *is a Lie algebra homomorphism.*

Proof. This follows from our study of abstract Lie algebras and proposition 15.5. ■

Let's look at what this means. Recall that the Lie bracket for $\mathfrak{gl}(\mathfrak{g})$ is just $A \circ B - B \circ A$. Thus we have

$$\mathrm{ad}([v, w]) = [\mathrm{ad}(v), \mathrm{ad}(w)] = \mathrm{ad}(v) \circ \mathrm{ad}(w) - \mathrm{ad}(w) \circ \mathrm{ad}(v)$$

which when applied to a third vector z gives

$$[[v, w], z] = [v, [w, z]] - [w, [v, z]]$$

which is just a version of the Jacobi identity. Also notice that using the anti-symmetry of the bracket we get $[z,[v,w]] = [w,[z,v]] + [v,[z,w]]$ which in turn is the same as

$$\mathrm{ad}(z)([v,w]) = [\mathrm{ad}(z)v, w] + [v, \mathrm{ad}(z)w]$$

so $\mathrm{ad}(z)$ is actually a derivation of the Lie algebra $\mathfrak{g}$ as explained before.

Proposition 15.7 *The Lie algebra $Der(\mathfrak{g})$ of all derivation of $\mathfrak{g}$ is the Lie algebra of the group of automorphisms* $\mathrm{Aut}(\mathfrak{g})$*. The image* $\mathrm{ad}(\mathfrak{g}) \subset Der(\mathfrak{g})$ *is the Lie algebra of the set of all inner automorphisms* $\mathrm{Int}(\mathfrak{g})$*.*

$$\begin{array}{ccc} \mathrm{ad}(\mathfrak{g}) & \subset & \mathrm{Der}(\mathfrak{g}) \\ \downarrow & & \downarrow \\ \mathrm{Int}(\mathfrak{g}) & \subset & \mathrm{Aut}(\mathfrak{g}) \end{array}$$

Let $\mu : G \times G \to G$ be the multiplication map. Recall that the tangent space $T_{(g,h)}(G \times G)$ is identified with $T_gG \times T_hG$. Under this identification we have

$$T_{(g,h)}\mu(v,w) = T_hL_gw + T_gR_hv$$

where $v \in T_gG$ and $w \in T_hG$. The following diagrams exhibit the relations:

$$\begin{array}{ccccc} & & G \times G, (g,h) & & \\ & \overset{pr_1}{\swarrow} & \| & \overset{pr_2}{\searrow} & \\ G, g & \rightarrow & G \times G, (g,h) & \leftarrow & G, h \\ & \underset{R_h}{\searrow} & \downarrow \mu & \underset{L_g}{\swarrow} & \\ & & G, gh & & \end{array}$$

The horizontal maps are the insertions $g \mapsto (g,h)$ and $h \mapsto (g,h)$. Applying the tangent functor to the last diagram gives.

$$\begin{array}{ccccc} Tpr_1 & & T_{(g,h)}(G \times G) & & Tpr_2 \\ & \swarrow & \updownarrow & \searrow & \\ T_gG & \rightarrow & T_gG \times T_hG & \leftarrow & T_hG \\ & \searrow & \downarrow T\mu & \swarrow & \\ T_gR_h & & T_{gh}G & & T_hL_g \end{array}$$

We will omit the proof be the reader should examine the diagrams and try to construct a proof on that basis.

We have another pair of diagrams to consider. Let $\nu : G \to G$ be the inversion map $\nu : g \mapsto g^{-1}$. We have the following commutative diagrams:

$$\begin{array}{ccccc} R_{g^{-1}} & & G, g & & L_{g^{-1}} \\ & \swarrow & & \searrow & \\ G, e & & \downarrow \nu & & G, e \\ & \searrow & & \swarrow & \\ L_{g^{-1}} & & G, g^{-1} & & R_{g^{-1}} \end{array}$$

Applying the tangent functor we get

$$\begin{array}{ccccc} TR_{g^{-1}} & & T_gG & & TL_{g^{-1}} \\ & \swarrow & & \searrow & \\ T_eG & & \downarrow T\nu & & T_eG \\ & \searrow & & \swarrow & \\ TL_{g^{-1}} & & T_{g^{-1}}G & & TR_{g^{-1}} \end{array} .$$

The result we wish to express here is that $T_g\nu = TL_{g^{-1}} \circ TR_{g^{-1}} = TR_{g^{-1}} \circ TL_{g^{-1}}$. Again the proof follows readily from the diagrams.

Chapter 16

Lie Group Actions and Homogenous Spaces

Here we set out our conventions regarding (right and left) group actions and the notion of equivariance. There is plenty of room for confusion just from the issues of right as opposed to left if one doesn't make a few observations and set down the conventions carefully from the start. We will make the usual choices but we will note how these usual choices lead to annoyances like the mismatch of homomorphisms with anti-homomorphisms in proposition 16.1 below.

16.1 Our Choices

1. When we write $L_X Y = \frac{d}{dt}\big|_0 \left(Fl_t^X\right)^* Y$ we are implicitly using a right action of $\mathrm{Diff}(M)$ on $\mathfrak{X}(M)$[1]. Namely, $Y \mapsto f^* Y$.

2. We have chosen to make the bracket of vector fields be defined so that $[X, Y] = XY - YX$ rather than by $YX - XY$. This makes it true that $L_X Y = [X, Y]$ so the first choice seems to influence this second choice.

3. We have chosen to define the bracket in a Lie algebra $\mathfrak{g}$ of a Lie group G to be given by using the identifying linear map $\mathfrak{g} = T_e G \to \mathfrak{X}^L(M)$ where $\mathfrak{X}^L(M)$ is **left** invariant vector fields. What if we had used **right** invariant vector fields? Then we would have $[X_e, Y_e]_{new} = [X', Y']_e$ where $X'_g = TR_g \cdot X_e$ is the right invariant vector field:

$$\begin{aligned} R_h^* X'(g) &= TR_h^{-1} X'(gh) = TR_h^{-1} TR_{gh} \cdot X_e \\ &= TR_h^{-1} \circ T(R_h \circ R_g) \cdot X_e = TR_g \cdot X_e \\ &= X'(g) \end{aligned}$$

[1] At least when X is complete since otherwise Fl_t^X is only a diffeomorphism on relatively compact set open sets and even then only for small enough t).

But notice that Now on the other hand, consider the inversion map $\nu : G \to G$. We have $v \circ R_{g^{-1}} = L_g \circ v$ and also $T\nu = -\operatorname{id}$ at T_eG so

$$\begin{aligned}(\nu^* X')(g) &= T\nu \cdot X'(g^{-1}) = T\nu \cdot TR_{g^{-1}} \cdot X_e \\ &= T(L_g \circ v)X_e = TL_g Tv \cdot X_e \\ &= -TL_g X_e = -X(g)\end{aligned}$$

thus $\nu^*[X', Y'] = [\nu^* X', \nu^* Y'] = [-X, -Y] = [X, Y]$. Now at e we have $(\nu^*[X', Y'])(e) = Tv \circ [X', Y'] \circ \nu(e) = -[X', Y']_e$. So we have $[X, Y]_e = -[X', Y']_e$.

So this choice is different by a sign also.

The source of the problem may just be conventions but it is interesting to note that if we consider $\mathrm{Diff}(M)$ as an infinite dimensional Lie group then the vector fields of that manifold would be maps $\overleftrightarrow{X} : \mathrm{Diff}(M) \to \mathfrak{X}(M)$ such $\overleftrightarrow{X}(\phi)$ is a vector field in $\mathfrak{X}(M)$ such that $Fl_0^{\overleftrightarrow{X}(\phi)} = \phi$. In other words, a field for every diffeomorphism, a "field of fields" so to speak. Then in order to get the usual bracket in $\mathfrak{X}(M)$ we would have to use right invariant (fields of) fields (instead of the conventional left invariant choice) and evaluate them at the identity element of $\mathrm{Diff}(M)$ to get something in $T_{\mathrm{id}} \mathrm{Diff}(M) = \mathfrak{X}(M)$. This makes one wonder if right invariant vector fields would have been a better convention to start with. Indeed some authors do make that convention.

16.1.1 Left actions

Definition 16.1 *A left action of a Lie group G on a manifold M is a smooth map $\lambda : G \times M \to M$ such that $\lambda(g_1, \lambda(g_2, m)) = \lambda(g_1 g_2, m))$ for all $g_1, g_2 \in G$. Define the partial map $\lambda_g : M \to M$ by $\lambda_g(m) = \lambda(g, m)$ and then the requirement is that $\widetilde{\lambda} : g \mapsto \lambda_g$ is a group homomorphism $G \to \mathrm{Diff}(M)$. We often write $\lambda(g, m)$ as $g \cdot m$.*

Definition 16.2 *For a left group action as above, we have for every $v \in \mathfrak{g}$ we define a vector field $v^\lambda \in \mathfrak{X}(M)$ defined by*

$$v^\lambda(m) = \left.\frac{d}{dt}\right|_{t=0} \exp(tv) \cdot m$$

*which is called the **fundamental vector field** associated with the action λ.*

Notice that $v^\lambda(m) = T\lambda_{(e,m)} \cdot (v, 0)$.

Proposition 16.1 *Given left action $\lambda : G \times M \to M$ of a Lie group G on a manifold M, the map $\widetilde{\lambda} : g \mapsto \lambda_g$ is a group homomorphism $G \to \mathrm{Diff}(M)$ by definition.* ***Despite this, the map $X \mapsto X^\lambda$ is a Lie algebra anti-homomorphism*** *$\mathfrak{g} \to \mathfrak{X}(M)$:*

$$[v, w]^\lambda = -[v^\lambda, w^\lambda]_{\mathfrak{X}(M)}$$

which implies that the bracket for the Lie algebra $\mathfrak{diff}(M)$ *of* $\mathrm{Diff}(M)$ *(as an infinite dimensional Lie group) is in fact* $[X,Y]_{\mathfrak{diff}(M)} := -[X,Y]_{\mathfrak{X}(M)}$.

Proposition 16.2 *If G acts on itself from the* left *by multiplication* $L : G\times G \to G$ *then the fundamental vector fields are the* right *invariant vector fields!*

16.1.2 Right actions

Definition 16.3 *A right action of a Lie group G on a manifold M is a smooth map* $\rho : M \times G \to M$ *such that* $\rho(\rho(m, g_2), g_1) = \rho(m, g_2g_1))$ *for all* $g_1, g_2 \in G$. *Define the partial map* $\rho^g : M \to M$ *by* $\rho^g(m) = \rho(m, g)$. *Then the requirement is that* $\widetilde{\rho} : g \mapsto \rho^g$ *is a group* ***anti-homomorphism*** $G \to \mathrm{Diff}(M)$. *We often write* $\rho(m, g)$ *as* $m \cdot g$

Definition 16.4 *For a right group action as above, we have for every* $v \in \mathfrak{g}$ *a vector field* $v^\rho \in \mathfrak{X}(M)$ *defined by*

$$v^\rho(m) = \left.\frac{d}{dt}\right|_{t=0} m \cdot \exp(tv)$$

which is called the ***fundamental vector field*** *associated with the right action* ρ.

Notice that $v^\rho(m) = T\rho_{(m,e)} \cdot (0, v)$.

Proposition 16.3 *Given right action* $\rho : M \times G \to M$ *of a Lie group G on a manifold M, the map* $\widetilde{\rho} : g \mapsto \rho^g$ *is a group anti-homomorphism* $G \to \mathrm{Diff}(M)$ *by definition.* ***However, the map*** $X \mapsto X^\lambda$ ***is a true Lie algebra homomorphism*** $\mathfrak{g} \to \mathfrak{X}(M)$:

$$[v, w]^\rho = [v^\rho, w^\rho]_{\mathfrak{X}(M)}$$

this disagreement again implies that the Lie algebra $\mathfrak{diff}(M)$ *of* $\mathrm{Diff}(M)$ *(as an infinite dimensional Lie group) is in fact* $\mathfrak{X}(M)$, *but with the bracket* $[X,Y]_{\mathfrak{diff}(M)} := -[X,Y]_{\mathfrak{X}(M)}$.

Proposition 16.4 *If G acts on itself from the* right *by multiplication* $L : G \times G \to G$ *then the fundamental vector fields are the* left *invariant vector fields* $\mathfrak{X}_L(G)$*!*

Proof: Exercise.

16.1.3 Equivariance

Definition 16.5 *Given two left actions* $\lambda_1 : G \times M \to M$ *and* $\lambda_2 : G \times S \to S$ *we say that a map* $f : M \to N$ *is (left)* ***equivariant*** *(with respect to these actions) if*

$$f(g \cdot s) = g \cdot f(s)$$
$$i.e.$$
$$f(\lambda_1(g, s)) = \lambda_2(g, f(s))$$

with a similar definition for right actions.

Notice that if $\lambda : G \times M \to M$ is a left action then we have an **associated right action** $\lambda^{-1} : M \times G \to M$ given by

$$\lambda^{-1}(p, g) = \lambda(g^{-1}, p).$$

Similarly, to a right action $\rho : M \times G \to M$ there is an associated left action

$$\rho^{-1}(g, p) = \rho(p, g^{-1})$$

and then we make the follow conventions concerning equivariance when mixing right with left.

Definition 16.6 *Is is often the case that we have a right action on a manifold P (such as a principle bundle) and a left action on a manifold S. Then equivariance is defined by converting the right action to its associated left action. Thus we have the requirement*

$$f(s \cdot g^{-1}) = g \cdot f(s)$$

or we might do the reverse and define equivariance by

$$f(s \cdot g) = g^{-1} \cdot f(s)$$

16.1.4 The action of $\mathrm{Diff}(M)$ and map-related vector fields.

Given a diffeomorphism $\Phi : M \to N$ define $\Phi_\star : \Gamma(M, TM) \to \Gamma(N, TN)$ by

$$\Phi_* X = T\Phi \circ X \circ \Phi^{-1}$$

and $\Phi^* : \Gamma(M, TN) \to \Gamma(M, TM)$ by

$$\Phi^* X = T\Phi^{-1} \circ X \circ \Phi$$

If $M = N$, this gives a right and left pair of actions of the diffeomorphism group $\mathrm{Diff}(M)$ on the space of vector fields $\mathfrak{X}(M) = \Gamma(M, TM)$.

$$\begin{aligned} \mathrm{Diff}(M) \times \mathfrak{X}(M) &\to \mathfrak{X}(M) \\ (\Phi, X) &\mapsto \Phi_* X \end{aligned}$$

and

$$\begin{aligned} \mathfrak{X}(M) \times \mathrm{Diff}(M) &\to \mathfrak{X}(M) \\ (X, \Phi) &\mapsto \Phi^* X \end{aligned}$$

16.1.5 Lie derivative for equivariant bundles.

Definition 16.7 *An equivariant left action for a bundle $E \to M$ is a pair of actions $\gamma^E : G \times E \to E$ and : $G \times M \to M$ such that the diagram below commutes*

$$\begin{array}{llll} \gamma^E : & G \times E & \to & E \\ & \downarrow & & \downarrow \\ \gamma : & G \times M & \to & M \end{array}$$

In this case we can define an action on the sections $\Gamma(E)$ *via*

$$\gamma_g^*\mathbf{s} = (\gamma^E)^{-1} \circ \mathbf{s} \circ \gamma_g$$

and then we get a Lie derivative for $\mathbf{X} \in LG$

$$L_{\mathbf{X}}(\mathbf{s}) = \left.\frac{d}{dt}\right|_0 \gamma^*_{\exp t\mathbf{X}}\mathbf{s}$$

Chapter 17

Homogeneous Spaces and Klein Geometries.

We have already seen several homogeneous spaces.

1. Affine spaces: These are the spaces $\mathbf{A}^n$ which are just $\mathbb{R}^n$ acted on the left by the affine group $Aff(n)$ or the proper affine group $Aff^+(n)$.

2. Special affine space: This is again the space $\mathbf{A}^n$ but with the group now restricted to be the special affine group $SAff(n)$.

3. Euclidean spaces: These are the spaces $\mathbf{E}^n$ which are just $\mathbb{R}^n$ acted on the left by the group of Euclidean motions $\mathrm{Euc}(n)$ or the proper affine group $\mathrm{Euc}^+(n)$ (defined by requiring the linear part to be orientation preserving.

4. Projective spaces: These are the spaces $\mathbb{R}P^n = P(\mathbb{R}^{n+1})$ (consisting of lines through the origin in $\mathbb{R}^{n+1}$) acted on the left by the $\mathrm{PSL}(n)$.

In this chapter we will have actions acting both from the left and from the right and so we will make a distinction in the notation.

Definition 17.1 *The **orbit** of $x \in M$ under a right action by G is denoted $x \cdot G$ or xG and the set of orbits M/G partition M into equivalence classes. For left actions we write $G \cdot x$ and $G \setminus M$ for the orbit space.*

Example 17.1 *If H is a closed subgroup of G then H acts on G from the right by right multiplication.. The space of orbits G/H of this* right *action is just the set of **right cosets**. Similarly, we have the space of left cosets $H \setminus G$.*

Warning: We call gH a *right* coset because it comes from a right action but for at least half of the literature the convention is to call gH a left coset and G/H the space of left cosets.

We recall a few more definitions from chapter 2:

Definition 17.2 *A left (resp. right) action is said to be* ***effective*** *if $g \cdot p = x$ (resp. $x \cdot g = x$) for every $x \in M$ implies that $g = e$ and is said to be* ***free*** *if $g \cdot x = x$ (resp. $x \cdot g = x$) for even one $x \in M$ implies that $g = e$.*

Definition 17.3 *A left (resp. right) action is said to be a* ***transitive action*** *if there is only one orbit in the space of orbits.*

This single orbit would have to be M and transitivity means, by definition, that given pair $x, y \in M$, there is a $g \in G$ with $g \cdot x = y$ (resp. $x \cdot g = y$).

Theorem 17.1 *Let $\lambda : G \times M \to M$ be a left action and fix $x_0 \in M$. Let $H = H_{x_0}$ be the* ***isotropy subgroup*** *of x_0 defined by*

$$H = \{g \in G : g \cdot x_0 = x_0\}.$$

Then we have a natural bijection

$$G \cdot x_0 \cong G/H$$

given by $g \cdot x_0 \mapsto gH$. In particular, if the action is transitive then $G/H \cong M$ and x_0 maps to H.

Exercise 17.1 *Show that the action of $Aff^+(\mathbf{A}^2)$ on $\mathbf{A}^2$ is transitive and effective but not free.*

Exercise 17.2 *Fix a point x_0 (say $(0,0)$) in $\mathbf{A}^2$ and show that $H := \{g \in Aff^+(\mathbf{A}^2) : gx_0 = x_0\}$ is a closed subgroup of $Aff^+(\mathbf{A}^2)$ isomorphic to $\mathrm{SL}(2)$.*

Exercise 17.3 *Let $H \cong \mathrm{SL}(2)$ be as in the previous exercise. Show that there is a natural 1-1 correspondence between the cosets of $Aff^+(\mathbf{A}^2)/H$ and the points of $\mathbf{A}^2$.*

Exercise 17.4 *Show that the bijection of the previous example is a homeomorphism if we give $Aff^+(\mathbf{A}^2)/H$ its quotient topology.*

Exercise 17.5 *Let S^2 be the unit sphere considered a subset of $\mathbb{R}^3$ in the usual way. Let the group $SO(3)$ act on S^2 by rotation. This action is clearly continuous and transitive. Let $n = (0,0,1)$ be the "north pole". Show that if H is the (closed) subgroup of $SO(3)$ consisting of all $g \in \mathrm{SO}(3)$ such that $g \cdot n = n$ then $x = g \cdot n \mapsto g$ gives a well defined bijection $S^2 \cong \mathrm{SO}(3)/H$. Note that $H \cong \mathrm{SO}(2)$ so we may write $S^2 \cong \mathrm{SO}(3)/\mathrm{SO}(2)$.*

Let $\lambda : G \times M \to M$ be a left action and fix $x_0 \in M$. Denote the projection onto cosets by π and also write $r^{x_0} : g \longmapsto gx_0$. Then we have the following equivalence of maps

$$\begin{array}{ccc} G & = & G \\ \pi \downarrow & & \downarrow r^{x_0} \\ G/H & \cong & M \end{array}$$

Thus, in the transitive action situation from the last theorem, we may as well assume that $M = G/H$ and then we have the literal equality $r^{x_0} = \pi$ and the left action is just $l_g : g_0H \mapsto hg_0H$ (or $gx \mapsto hx$ where $x \in G/H$). Continuing for now to make a distinction between M and G/H we note that the isotropy H also acts on M or equivalently on G/H. We denote this action by $I^{x_0} : H \times M \to M$ where $I^{x_0} : h \mapsto hx$. The equivalent action on G/H is $(h, gH) \mapsto hgH$. Of course I^{x_0} is just the restriction of $l : G \times M \to M$ to $H \times M \subset G \times M$ and $I_h^{x_0} = l_h$ for any $h \in H$ ($I_h^{x_0} =: I^{x_0}(h)$).

Exercise 17.6 *Let $H_1 := G_{x_1}$ (isotropy of x_1) and $H_2 := G_{x_2}$ (isotropy of x_2) where $x_2 = gx_1$ for some $g \in G$. Show that there is a natural Lie group isomorphisms $H_1 \cong H_2$ and a natural diffeomorphism $G/H_1 \cong G/H_2$ which is an equivalence of actions.*

For each $h \in H$ the map $I_h^{x_0} : M \to M$ fixes the point x_0 and so the differential $T_{x_0}I_h^{x_0}$ maps $T_{x_0}M$ onto itself. Let us abbreviate $T_{x_0}I_h^{x_0}$ to $\iota_{x_0}(h)$. For each h we have a linear automorphism $\iota_{x_0}(h) \in GL(T_{x_0}M)$ and it is easy to check that $h \mapsto \iota_{x_0}(h)$ is a group representation on the space $T_{x_0}M$. This representation is called the **linear isotropy representation** (at x_0) and the group $I^{x_0}(H) \subset \mathrm{GL}(T_{x_0}M)$ is called the **linear isotropy subgroup**. On the other hand we for each $h \in H$ have another action $C_h : G \to G$ given by $g \longmapsto hgh^{-1}$ which fixes H and whose derivative is the adjoint map $Ad_h : \mathfrak{g} \to \mathfrak{g}$ and $h \longmapsto Ad_h$ is the adjoint representation defined earlier in the book. It is easy to see that the map $Ad_h : \mathfrak{g} \to \mathfrak{g}$ descends to a map $\widetilde{Ad_h} : \mathfrak{g}/\mathfrak{h} \to \mathfrak{g}/\mathfrak{h}$. We are going to show that there is a natural isomorphism $T_{x_0}M \cong \mathfrak{g}/\mathfrak{h}$ such that for each $h \in H$ the following diagram commutes:

$$\begin{array}{cccc} \widetilde{Ad_h}: & \mathfrak{g}/\mathfrak{h} \to & \mathfrak{g}/\mathfrak{h} \\ & \downarrow & \downarrow \\ I_{x_0}: & T_{x_0}M \to & T_{x_0}M \end{array} \tag{17.1}$$

One way to state the meaning of this result is to say that $h \mapsto \widetilde{Ad_h}$ is a representation of H on the vector space $\mathfrak{g}/\mathfrak{h}$ which is equivalent to the linear isotropy representation. The isomorphism $T_{x_0}M \cong \mathfrak{g}/\mathfrak{h}$ is given in the following very natural way: Let $\xi \in \mathfrak{g}$ and consider $T_e\pi(\xi) \in T_{x_0}M$. If $\varsigma \in \mathfrak{h}$ then

$$T_e\pi(\xi + \varsigma) = T_e\pi(\xi) + T_e\pi(\varsigma) = T_e\pi(\xi)$$

and so $\xi \mapsto T_e\pi(\xi)$ induces a map on $\mathfrak{g}/\mathfrak{h}$. Now if $T_e\pi(\xi) = 0 \in T_{x_0}M$ then as you are asked to show in exercise 17.7 below $\xi \in \mathfrak{h}$ which in turn means that the induced map $\mathfrak{g}/\mathfrak{h} \to T_{x_0}M$ has a trivial kernel. As usual this implies that the map is in fact an isomorphism since $\dim(\mathfrak{g}/\mathfrak{h}) = \dim(T_{x_0}M)$. Let us now see why the diagram 17.1 commutes. Let us take a the scenic root to the conclusion since it allows us to see the big picture a bit better. First the following diagram clearly commutes:

$$\begin{array}{ccc} \exp t\xi & \stackrel{C_h}{\to} & h(\exp t\xi)h^{-1} \\ \pi\downarrow & & \pi\downarrow \\ (\exp t\xi)H & \stackrel{\tau_h}{\to} & h(\exp t\xi)H \end{array}$$

which under the identification $M = G/H$ is just

$$\begin{array}{ccc} \exp t\xi & \rightarrow & h(\exp t\xi)h^{-1} \\ \pi\downarrow & & \pi\downarrow \\ (\exp t\xi)x_0 & \rightarrow & h(\exp t\xi)x_0 \end{array}.$$

Applying the tangent functor (looking at the differential) we get the commutative diagram

$$\begin{array}{ccc} \xi & \rightarrow & Ad_h\xi \\ \downarrow & & \downarrow \\ T_e\pi(\xi) & \stackrel{d\tau_h}{\rightarrow} & T_e\pi(Ad_h\xi) \end{array}$$

and in turn

$$\begin{array}{ccc} [\xi] & \mapsto & \widetilde{Ad_h}([\xi]) \\ \downarrow & & \downarrow \\ T_e\pi(\xi) & \mapsto & T_e\pi(Ad_h\xi) \end{array}.$$

This latter diagram is in fact the element by element version of 17.1.

Exercise 17.7 *Show that $T_e\pi(\xi) \in T_{x_0}M$ implies that $\xi \in \mathfrak{h}$.*

Consider again homogenous space $\mathbf{E}^n$ with the action of Euc(n). This example has a special property that is shared by many important homogeneous spaces called reductivity.

`Note: More needs to be added here about reductive homogeneous spaces.`

17.1 Geometry of figures in Euclidean space

We take as our first example the case of a Euclidean space. We intend to study figures in $\mathbf{E}^n$ but we first need to set up some machinery using differential forms and moving frames. Let e_1, e_2, e_3 be a moving frame on $\mathbf{E}^3$. Using the identification of $T_x\mathbf{E}^3$ with $\mathbb{R}^3$ we may think of each e_i as a function with values in $\mathbb{R}^3$. If x denotes the identity map then we interpret dx as the map $T\mathbf{E}^3 \to \mathbb{R}^3$ given by composing the identity map on the tangent bundle $T\mathbf{E}^3$ with the canonical projection map $T\mathbf{E}^3 = \mathbf{E}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$. If $\theta^1, \theta^2, \theta^3$ is the frame field dual to e_1, e_2, e_3 then we may write

$$dx = \sum e_i \theta^i. \tag{17.2}$$

Also, since we are interpreting each e_i as an $\mathbb{R}^3-$valued function we may take the componentwise exterior derivative to make sense of de_i. Then de_i is a vector valued differential form: If $e_i = f_1\mathbf{i} + f_2\mathbf{j} + f_3\mathbf{k}$ then $de_i = df_1\mathbf{i} + df_2\mathbf{j} + df_3\mathbf{k}$. We may write

$$de_j = \sum e_i \omega_j^i \tag{17.3}$$

for some set of $1-$forms ω_j^i which we arrange in a matrix $\omega = (\omega_j^i)$. If we take exterior derivative of equations 17.2 and 17.3 For the first one we calculate

$$\begin{aligned}
0 = ddx &= \sum_{i=1}^{n} e_i \theta^i \\
&= \sum_{i=1}^{n} de_i \wedge \theta^i + \sum_{i=1}^{n} e_i \wedge d\theta^i \\
&= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} e_j \omega_j^i \right) \wedge \theta^i + \sum_{i=1}^{n} e_i \wedge d\theta^i.
\end{aligned}$$

From this we get the first of the following two structure equations. The second one is obtained similarly from the result of differentiating 17.3.

$$\begin{aligned}
d\theta^i &= -\sum \omega_j^i \wedge \theta^j \\
d\omega_j^i &= -\sum \omega_k^i \wedge \omega_j^k
\end{aligned} \tag{17.4}$$

Furthermore, if we differentiate $e_i \cdot e_j = \delta_{ij}$ we find out that $\omega_j^i = -\omega_i^j$.

If we make certain innocent identifications and conventions we can relate the above structure equations to the group $\mathrm{Euc}(n)$ and its Lie algebra. We will identify $\mathbf{E}^n$ with the set of column vectors of the form

$$\begin{bmatrix} 1 \\ x \end{bmatrix} \text{ where } x \in \mathbb{R}^n$$

Then the group $\mathrm{Euc}(n)$ is presented as the set of all square matrices of the form

$$\begin{bmatrix} 1 & 0 \\ v & Q \end{bmatrix} \text{ where } Q \in \mathrm{O}(n) \text{ and } v \in \mathbb{R}^n.$$

The action $\mathrm{Euc}(n)\times\mathbf{E}^n \to \mathbf{E}^n$ is then simply given by matrix multiplication (see chapter 2). One may easily check that the matrix Lie algebra that we identify as the Lie algebra $\mathfrak{euc}(n)$ of $\mathrm{Euc}(n)$ consists of all matrices of the form

$$\begin{bmatrix} 0 & 0 \\ v & A \end{bmatrix} \text{ where } v \in \mathbb{R}^n \text{ and } A \in \mathfrak{so}(n) \text{ (antisymmetric matrices)}$$

The isotropy of the point $o := \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is easily seen to be the subgroup G_o consisting of all elements of the form

$$\begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} \text{ where } Q \in \mathrm{O}(n)$$

which is clearly isomorphic to $\mathrm{O}(n)$. This isotropy group G_o is just the group of rotations about the origin. The origin is not supposed to special and so we should point out that any point would work fine for what we are about to do. In fact, for any other point $x \sim \begin{bmatrix} 1 \\ x \end{bmatrix}$ we have an isomorphism $G_o \cong G_x$ given by $h \mapsto t_x h t_x$ where

$$t_x = \begin{bmatrix} 1 & 0 \\ x & I \end{bmatrix}.$$

(see exercise 17.6).

For each $x \in \mathbf{E}^n$ tangent space $T_x\mathbf{E}^n$ consists of pairs $x \times v$ where $v \in \mathbb{R}^n$ and so the dot product on $\mathbb{R}^n$ gives an obvious inner product on each $T_x\mathbf{E}^n$: For two tangent vectors in $T_x\mathbf{E}^n$, say $v_x = x \times v$ and $w_x = x \times w$ we have $\langle v_x, w_x\rangle = v \cdot w$.

Remark 17.1 *The existence of an inner product in each tangent space makes $\mathbf{E}^n$ a Riemannian manifold (a smoothness condition is also needed). Riemannian geometry (studied in chapter 20) is one possible generalization of Euclidean geometry (See figure 2 and the attendant discussion). Riemannian geometry represents an approach to geometry that is initially quite different in spirit from Klein's approach.*

Now think of each element of the frame $e_1, ..., e_n$ as a column vector of functions and form a matrix of functions e. Let x denote the "identity map" given as a column vector of functions $x = (x^1, x^2, x^3) = (x, y, z)$. Next we form the matrix of functions

$$\begin{bmatrix} 1 & 0 \\ x & e \end{bmatrix}$$

This just an element of $\mathrm{Euc}(n)$! Now we can see that the elements of $\mathrm{Euc}(n)$ are in a natural $1-1$ correspondents with the set of all frames on E^n and the matrix we have just introduce corresponds exactly to the moving frame $x \mapsto (e_1(x), ..., e_n(x))$. The differential of this matrix gives a matrix of one forms

$$\varpi = \begin{bmatrix} 0 & 0 \\ \theta & \omega \end{bmatrix}$$

and it is not hard to see that θ is the column consisting of the same θ^i as before and also that $\omega = (\omega^i_j)$. Also, notice that $x \mapsto \varpi(x) = \begin{bmatrix} 0 & 0 \\ \theta & \omega \end{bmatrix}$ takes values in the Lie algebra $\mathfrak{euc}(n)$. This looking like a very natural state of affairs. In fact, the structure equations are encoded as a singe matrix equation

$$d\varpi = \varpi \wedge \varpi.$$

The next amazingly cool fact it that if we pull back ϖ to $\mathrm{Euc}(n)$ via the projection $\pi : \begin{bmatrix} 0 & 0 \\ x & e \end{bmatrix} \mapsto x \in \mathbf{E}^n$ then we obtain the Maurer-Cartan form of the group $\mathrm{Euc}(n)$ and the equation $d\varpi = \varpi \wedge \varpi$ pulls back to the structure equations for $\mathrm{Euc}(n)$.

Chapter 18

Distributions and Frobenius' Theorem

18.1 Definitions

In this section we take M to be a C^∞ manifold modelled on a Banach space M. Roughly speaking, a smooth distribution is an assignment $\triangle$ of a subspace $\triangle_p \subset T_pM$ to each $p \in M$ such that for each fixed $p_0 \in M$ there is a family of smooth vector fields $X_1, ..., X_k$ defined on some neighborhood U_{p_0} of p_0 and such that $\triangle_x = span\{X_1(x), ..., X_k(x)\}$ for each $x \in U_{p_0}$. We call the distribution regular if we can always choose these vector fields to be linearly independent on each tangent space T_xM for $x \in U_{p_0}$ and each U_{p_0}. It follows that in this case k is locally constant. For a regular distribution k is called the rank of the distribution. A rank k regular distribution is the same thing as a rank k subbundle of the tangent bundle. We can also consider regular distributions of infinite rank by simply defining such to be a subbundle of the tangent bundle.

Definition 18.1 *A (smooth) regular distribution on a manifold M is a smooth vector subbundle of the tangent bundle TM.*

18.2 Integrability of Regular Distributions

By definition a regular distribution $\triangle$ is just another name for a subbundle $\triangle \subset TM$ of the tangent bundle and we write $\triangle_p \subset T_pM$ for the fiber of the subbundle at p. So what we have is a smooth assignment of a subspace $\triangle_p$ at every point. The subbundle definition guarantees that the spaces $\triangle_p$ all have the same dimension (if finite) in each connected component of M. This dimension is called the rank of the distribution. There is a more general notion of distribution that we call a singular distribution and that is defined in the same way except for the requirement of constancy of dimension. We shall study singular distributions later.

Definition 18.2 *Let X locally defined vector field. We say that X* ***lies in*** *the distribution $\triangle$ if $X(p) \in \triangle_p$ for each p in the domain of X. In this case, we write $X \in \triangle$ (a slight abuse of notation).*

Note that in the case of a regular distribution we can say that for X to lie in the distribution $\triangle$ means that X takes values in the subbundle $\triangle \subset TM$.

Definition 18.3 *We say that a locally defined differential j-form ω* ***vanishes on*** *$\triangle$ if for every choice of vector fields $X_1, ..., X_j$ defined on the domain of ω that lie in $\triangle$ the function $\omega(X_1, ..., X_j)$ is identically zero.*

For a regular distribution $\triangle$ consider the following two conditions.

Fro1 For every pair of locally defined vector fields X and Y with common domain that lie in the distribution $\triangle$ the bracket $[X, Y]$ also lies in the distribution.

Fro2 For each locally defined smooth 1-form ω that vanishes on $\triangle$ the 2-form $d\omega$ also vanishes on $\triangle$.

Lemma 18.1 *Conditions (1) and (2) above are equivalent.*

Proof. The proof that these two conditions are equivalent follows easily from the formula

$$d\omega(X, Y) = X(\omega(Y)) - Y\omega(X) - \omega([X, Y]).$$

Suppose that (1) holds. If ω vanishes on $\triangle$ and X, Y lie in $\triangle$ then the above formula becomes

$$d\omega(X, Y) = -\omega([X, Y])$$

which shows that $d\omega$ vanishes on $\triangle$ since $[X, Y] \in \triangle$ by condition (1). Conversely, suppose that (2) holds and that $X, Y \in \triangle$. Then $d\omega(X, Y) = -\omega([X, Y])$ again and a local argument using the Hahn-Banach theorem shows that $[X, Y] = 0$. ■

Definition 18.4 *If either of the two equivalent conditions introduced above holds for a distribution $\triangle$ then we say that $\triangle$ is* ***involutive****.*

Exercise 18.1 *Suppose that $\mathfrak{X}$ is a family of locally defined vector fields of M such that for each $p \in M$ and each local section X of the subbundle $\triangle$ defined in a neighborhood of p, there is a finite set of local fields $\{X_i\} \subset \mathfrak{X}$ such that $X = \sum a^i X_i$ on some possible smaller neighborhood of p. Show that if $\mathfrak{X}$ is closed under bracketing then $\triangle$ is involutive.*

There is a very natural way for distributions to arise. For instance, consider the punctured 3-space $M = \mathbb{R}^3 - \{0\}$. The level sets of the function $\varepsilon : (x, y, x) \mapsto x^2 + y^2 + x^2$ are spheres whose union is all of $\mathbb{R}^3 - \{0\}$. Now define a distribution by the rule that $\triangle_p$ is the tangent space at p to the sphere containing p. Dually, we can define this distribution to be the given by the rule

$$\triangle_p = \{v \in T_pM : d\varepsilon(v) = 0\}.$$

The main point is that each p contains a submanifold S such that $\triangle_x = T_xS$ for all $x \in S \cap U$ for some sufficiently small open set $U \subset M$. On the other hand, not all distributions arise in this way.

Definition 18.5 *A distribution $\triangle$ on M is called **integrable at** $p \in M$ there is a submanifold S_p containing p such that $\triangle_x = T_xS_p$ for all $x \in S$. (Warning: S_p is locally closed but not necessarily a closed subset and may only be defined very near p.) We call such submanifold a **local integral submanifold** of $\triangle$.*

Definition 18.6 *A regular distribution $\triangle$ on M is called **(completely) integrable** if for every $p \in M$ there is a (local) integral submanifold of $\triangle$ containing p.*

If one considers a distribution on a finite dimensional manifold there is a nice picture of the structure of an integrable distribution. Our analysis will eventually allow us to see that a regular distribution $\triangle$ of rank k on an n-manifold M is (completely) integrable if and only if there is a cover of M by charts ψ_a, U_a such that if $\psi_a = (y^1, ..., y^n)$ then for each $p \in U_a$ the submanifold $S_{\alpha,p}$ defined by $S_{\alpha,p} := \{x \in U_a : y^i(x) = y^i(p) \text{ for } k+1 \le i \le n\}$ has

$$\triangle_x = T_xS_{\alpha,p} \text{ for all } x \in S_p.$$

Some authors use this as the definition of integrable distribution but this definition would be inconvenient to generalize to the infinite dimensional case. A main goal of this section is to prove the theorem of Frobenius which says that a regular distribution is integrable if and only if it is involutive.

18.3 The local version Frobenius' theorem

Here we study regular distributions; also known as tangent subbundles. The presentation draws heavily on that given in [L1]. Since in the regular case a distribution is a subbundle of the tangent bundle it will be useful to consider such subbundles a little more carefully. Recall that if $E \to M$ is a subbundle of TM then $E \subset TM$ and there is an atlas of adapted VB-charts for TM; that is, charts $\phi : \tau_M^{-1}(U) \to U \times \mathsf{M} = U \times \mathsf{E} \times \mathsf{F}$ where $\mathsf{E} \times \mathsf{F}$ is a fixed splitting of M. Thus M is modelled on the split space $\mathsf{E} \times \mathsf{F} = \mathsf{M}$. Now for all local questions we may assume that in fact the tangent bundle is a trivial bundle of the form $(U_1 \times U_2) \times (\mathsf{E} \times \mathsf{F})$ where $U_1 \times U_2 \subset \mathsf{E} \times \mathsf{F}$. It is easy to see that our subbundle must now consist of a choice of subspace $\mathsf{E}_1(x, y)$ of $(\mathsf{E} \times \mathsf{F})$ for every $(x, y) \in U_1 \times U_2$. In fact, the inverse of our trivialization gives a map

$$\phi^{-1} : (U_1 \times U_2) \times (\mathsf{E} \times \mathsf{F}) \to (U_1 \times U_2) \times (\mathsf{E} \times \mathsf{F})$$

such that the image under ϕ^{-1} of $\{(x,y)\}\times\mathsf{E}\times\{0\}$ is exactly $\{(x,y)\}\times\mathsf{E}_1(x,y)$. The map ϕ^{-1} must have the form

$$\phi((x,y),v,w) = ((x,y), f_{(x,y)}(v,w), g_{(x,y)}(v,w))$$

for where $f_{(x,y)} : \mathsf{E}\times\mathsf{F}\to\mathsf{E}$ and $g_{(x,y)} : \mathsf{E}\times\mathsf{F}\to\mathsf{E}$ are linear maps depending smoothly on (x,y). Furthermore, for all (x,y) the map $f_{(x,y)}$ takes $\mathsf{E}\times\{0\}$ isomorphically onto $\{(x,y)\}\times\mathsf{E}_1(x,y)$. Now the composition

$$\kappa : (U_1\times U_2)\times\mathsf{E} \hookrightarrow (U_1\times U_2)\times(\mathsf{E}\times\{0\}) \overset{\phi^{-1}}{\to} (U_1\times U_2)\times(\mathsf{E}\times\mathsf{F})$$

maps $\{(x,y)\}\times\mathsf{E}$ isomorphically onto $\{(x,y)\}\times\mathsf{E}_1(x,y)$ and must have the form

$$\kappa(x,y,v) = (x,y,\lambda(x,y)\cdot v, \ell(x,y)\cdot v)$$

for some smooth maps $(x,y)\mapsto\lambda(x,y)\in L(\mathsf{E},\mathsf{E})$ and $(x,y)\mapsto\ell(x,y)\in L(\mathsf{E},\mathsf{F})$. By a suitable "rotation" of the space $\mathsf{E}\times\mathsf{F}$ for each (x,y) we may assume that $\lambda_{(x,y)} = \mathrm{id}_{\mathsf{E}}$. Now for fixed $v\in\mathsf{E}$ the map $X_v : (x,y)\mapsto(x,y,v,\ell_{(x,y)}v)$ is (a local representation of) a vector field with values in the subbundle E. The principal part is $\mathsf{X}_v(x,y) = (v,\ell_{(x,y)}\cdot v)$.

Now $\ell(x,y)\in L(\mathsf{E},\mathsf{F})$ and so $D\ell(x,y)\in L(\mathsf{E}\times\mathsf{F}, L(\mathsf{E},\mathsf{F}))$. In general for a smooth family of linear maps Λ_u and a smooth map $v:(x,y)\mapsto v(x,y)$ we have

$$D(\Lambda_u\cdot v)(w) = D\Lambda_u(w)\cdot v + \Lambda_u\cdot(Dv)(w)$$

and so in the case at hand

$$\begin{aligned}&D(\ell(x,y)\cdot v)(w_1,w_2)\\&= (D\ell(x,y)(w_1,w_2))\cdot v + \ell(x,y)\cdot(Dv)(w_1,w_2).\end{aligned}$$

For any two choices of smooth maps v_1 and v_2 as above we have

$$\begin{aligned}[\mathsf{X}_{v_1},\mathsf{X}_{v_2}]_{(x,y)} &= (D\mathsf{X}_{v_2})_{(x,y)}\mathsf{X}_{v_1}(x,y) - (D\mathsf{X}_{v_1})_{(x,y)}\mathsf{X}_{v_2}(x,y)\\
&= ((Dv_2)(v_1,\ell_{(x,y)}v_1) - (Dv_1)(v_2,\ell_{(x,y)}v_2), D\ell(x,y)(v_1,\ell_{(x,y)}v_1)\cdot v_2\\
&+ \ell(x,y)\cdot(Dv_2)(v_1,\ell_{(x,y)}v_1) - D\ell(x,y)(v_2,\ell_{(x,y)}v_2)\cdot v_1\\
&- \ell(x,y)\cdot(Dv_1)(v_2,\ell_{(x,y)}v_2))\\
&= (\xi, D\ell(x,y)(v_1,\ell_{(x,y)}v_1)\cdot v_2 - D\ell(x,y)(v_2,\ell_{(x,y)}v_2)\cdot v_1 + \ell(x,y)\cdot\xi).\end{aligned}$$

where $\xi = (Dv_2)(v_1,\ell_{(x,y)}v_1) - (Dv_1)(v_2,\ell_{(x,y)}v_2)$. Thus $[\mathsf{X}_{v_1},\mathsf{X}_{v_2}]_{(x,y)}$ is in the subbundle if and only if

$$D\ell(x,y)(v_1,\ell_{(x,y)}v_1)\cdot v_2 - D\ell(x,y)(v_2,\ell_{(x,y)}v_2)\cdot v_1.$$

We thus arrive at the following characterization of involutivity:

Lemma 18.2 *Let Δ be a subbundle of TM. For every $p\in M$ there is a tangent bundle chart containing T_pM of the form described above so that any vector field* *vector field taking values in the subbundle is represented as a map* $\mathsf{X}_v : U_1 \times U_2 \to \mathsf{E} \times \mathsf{F}$ *of the form* $(x, y) \mapsto (v(x, y), \ell_{(x,y)} v(x, y))$*. Then* Δ *is involutive (near* p*) if and only if for any two smooth maps* $v_1 : U_1 \times U_2 \to \mathsf{E}$ *and* $v_2 : U_1 \times U_2 \to \mathsf{E}$ *we have*

$$D\ell(x, y)(v_1, \ell_{(x,y)} v_1) \cdot v_2 - D\ell(x, y)(v_2, \ell_{(x,y)} v_2) \cdot v_1.$$

Theorem 18.1 *A regular distribution* Δ *on* M *is integrable if and only if it is involutive.*

Proof. First suppose that Δ is integrable. Let X and Y be local vector fields that lie in Δ. Pick a point x in the common domain of X and Y. Our choice of x being arbitrary we just need to show that $[X, Y](x) \in \Delta$. Let $S \subset M$ be a local integral submanifold of Δ containing the point x. The restrictions $X|_S$ and $Y|_S$ are related to X and Y by an inclusion map and so by the result on related vector fields we have that $[\, X|_S \,, Y|_S] = [X, Y]|_S$ on some neighborhood of x. Since S is a manifold and $[\, X|_S \,, Y|_S]$ a local vector field on S we see that $[X, Y]|_S\,(x) = [X, Y](x)$ is tangent to S and so $[X, Y](x) \in \Delta$. Suppose now that Δ is involutive. Since this is a local question we may assume that our tangent bundle is a trivial bundle $(U_1 \times U_2) \times (\mathsf{E} \times \mathsf{F})$ and by our previous lemma we know that for any two smooth maps $v_1 : U_1 \times U_2 \to \mathsf{E}$ and $v_2 : U_1 \times U_2 \to \mathsf{E}$ we have

$$D\ell(x, y)(v_1, \ell_{(x,y)} v_1) \cdot v_2 - D\ell(x, y)(v_2, \ell_{(x,y)} v_2) \cdot v_1.$$

Claim 18.1 *For any* $(x_0, y_0) \in U_1 \times U_2$ *there exists possibly smaller open product* $U_1' \times U_2' \subset U_1 \times U_2$ *containing* (x_0, y_0) *and a unique smooth map* $\alpha : U_1' \times U_2' \to U_2$ *such that* $\alpha(x_0, y) = y$ *for all* $y \in U_2'$ *and*

$$D_1 \alpha(x, y) = \ell(x, \alpha(x, y))$$

for all $(x, y) \in U_1' \times U_2'$.

Before we prove this claim we show how the result follows from it. For any $y \in U_2'$ we have the partial map $\alpha_y(x) := \alpha(x, y)$ and equation **??** above reads $D\alpha_y(x, y) = \ell(x, \alpha_y(x))$. Now if we define the map $\phi : U_1' \times U_2' \to U_1 \times U_2$ by $\phi(x, y) := (x, \alpha_y(x))$ then using this last version of equation **??** and the condition $\alpha(x_0, y) = y$ from the claim we see that

$$\begin{aligned} D_2 \alpha(x_0, y_0) &= D\alpha(x_0, .)(y_0) \\ &= D\,\mathrm{id}_{U_2'} = \mathrm{id}\,. \end{aligned}$$

Thus the Jacobian of ϕ at (x_0, y_0) has the block form

$$\begin{pmatrix} \mathrm{id} & 0 \\ * & \mathrm{id} \end{pmatrix}.$$

By the inverse function theorem ϕ is a local diffeomorphism in a neighborhood of (x_0, y_0). We also have that

$$\begin{aligned} (D_1 \phi)(x, y) \cdot (v, w) &= (v, D\alpha_y(x) \cdot w) \\ &= (v, \ell(x, \alpha_y(x)) \cdot v). \end{aligned}$$

Which represents an elements of the subbundle but is also of the form of tangents to the submanifolds that are the images of $U_1' \times \{y\}$ under the diffeomorphism ϕ for various choices of $y \in U_2$. This clearly saying that the subbundle is integrable.

Proof of the claim: By translation we may assume that $(x_0, y_0) = (0, 0)$. We use theorem H.2 from appendix B. With the notation of that theorem we let $f(t, x, y) := \ell(tz, y) \cdot z$ where $y \in U_2$ and z is an element of some ball $B(0, \epsilon)$ in E. Thus the theorem provides us with a smooth map $\beta : J_0 \times B(0, \epsilon) \times U_2$ satisfying $\beta(0, z, y) = y$ and

$$\frac{\partial}{\partial t}\beta(t, z, y) = \ell(tz, \beta(t, z, y)) \cdot z.$$

We will assume that $1 \in J$ since we can always arrange for this by making a change of variables of the type $t = as, z = x/a$ for a sufficiently small positive number a (we may do this at the expense of having to choose a smaller ϵ for the ball $B(0, \epsilon)$. We claim that if we define

$$\alpha(x, y) := \beta(1, x, y)$$

then for sufficiently small $|x|$ we have the required result. In fact we shall show that

$$D_2\beta(t, z, y) = t\ell(tz, \beta(t, z, y))$$

from which it follows that

$$D_1\alpha(x, y) = D_2\beta(1, x, y) = \ell(x, \alpha(x, y))$$

with the correct initial conditions (recall that we translated to (x_0, y_0)). Thus it remains to show that equation **??** holds. ¿From (3) of theorem H.2 we know that $D_2\beta(t, z, y)$ satisfies the following equation for any $v \in \mathsf{E}$:

$$\begin{aligned}\frac{\partial}{\partial t}D_2\beta(t, z, y) = t\frac{\partial}{\partial t}\ell(tz, \beta(t, z, y)) \cdot v \cdot z \\ + D_2\ell(tz, \beta(t, z, y)) \cdot D_2\beta(t, z, y) \cdot v \cdot z \\ + \ell(tz, \beta(t, z, y)) \cdot v.\end{aligned}$$

Now we fix everything but t and define a function of one variable:

$$\Phi(t) := D_2\beta(t, z, y) \cdot v - t\ell(tz, \beta(t, z, y).$$

Clearly, $\Phi(0) = 0$. Now we use two fixed vectors v,z and construct the fields $\mathsf{X}_v(x, y) = (v, \ell_{(x,y)} \cdot v)$ and $\mathsf{X}_z(x, y) = (z, \ell_{(x,y)} \cdot z)$. In this special case, the equation of lemma 18.2 becomes

$$D\ell(x, y)(v, \ell_{(x,y)}v) \cdot z - D\ell(x, y)(z, \ell_{(x,y)}z) \cdot v.$$

Now with this in mind we compute $\frac{d}{dt}\Phi(t)$:

$$
\begin{aligned}
\frac{d}{dt}\Phi(t) &= \frac{\partial}{\partial t}(D_2\beta(t,z,y)\cdot v - t\ell(tz,\beta(t,z,y)) \\
&= \frac{\partial}{\partial t}D_2\beta(t,z,y)\cdot v - t\frac{d}{dt}\ell(tz,\beta(t,z,y)) - \ell(tz,\beta(t,z,y)) \\
&= \frac{\partial}{\partial t}D_2\beta(t,z,y)\cdot v - t\{D_1\ell(tz,\beta(t,z,y))\cdot z \\
&+ D_2\ell(tz,\beta(t,z,y))\cdot\frac{\partial}{\partial t}\beta(t,z,y) - \ell(tz,\beta(t,z,y)) \\
&= D_2\ell(tz,\beta(t,z,y))\cdot\{D_2\beta(t,z,y)\cdot v - t\ell(tz,\beta(t,z,y)\}\cdot z \text{ (use 18.3)} \\
&= D_2\ell(tz,\beta(t,z,y))\cdot\Phi(t)\cdot z.
\end{aligned}
$$

So we arrive at $\frac{d}{dt}\Phi(t) = D_2\ell(tz,\beta(t,z,y))\cdot\Phi(t)\cdot z$ with initial condition $\Phi(0)=0$ which implies that $\Phi(t)\equiv 0$. This latter identity is none other than $D_2\beta(t,z,y)\cdot v = t\ell(tz,\beta(t,z,y))$. ∎

It will be useful to introduce the notion of a codistribution and then explore the dual relationship existing between distributions and codistributions.

Definition 18.7 *A (regular) codistribution Ω on a manifold M is a subbundle of the cotangent bundle. Thus a smooth assignment of a subspace $\Omega_x \subset T_x^*M$ for every $x \in M$. If $\dim \Omega_x = l < \infty$ we call this a rank l co-distribution.*

Using the definition of vector bundle chart adapted to a subbundle it is not hard to show , as indicated in the first paragraph of this section, that a (smooth) distribution of rank $k < \infty$ can be described in the following way:

Claim 18.2 *For a smooth distribution $\triangle$ of rank on M we have that for every $p \in M$ there exists a family of smooth vector fields $X_1, ..., X_k$ defined near p such that $\triangle_x = \operatorname{span}\{X_1(x), ..., X_k(x)\}$ for all x near p.*

Similarly, we have

Claim 18.3 *For a smooth codistribution Ω of rank k on M we have that for every $p \in M$ there exists a family of smooth 1-forms fields $\omega_1, ..., \omega_k$ defined near p such that $\Omega_x = \operatorname{span}\{\omega_1(x), ..., \omega_k(x)\}$ for all x near p.*

On the other hand we can use a codistribution to define a distribution and visa-versa. For example, for a regular codistribution Ω on M we can define a distribution $\triangle^{\perp\Omega}$ by

$$\triangle_x^{\perp\Omega} := \{v \in T_xM : \omega_x(v) = 0 \text{ for all } \omega_x \in \Omega_x\}.$$

Similarly, if $\triangle$ is a regular distribution on M then we can define a codistribution $\Omega^{\perp\triangle}$ by

$$\Omega_x^{\perp\triangle} := \{\omega_x \in T_x^*M : \omega_x(v) = 0 \text{ for all } v \in \triangle_x\}.$$

Notice that if $\triangle_1 \subset \triangle_2$ then $\triangle_2^{\perp\Omega} \subset \triangle_1^{\perp\Omega}$ and $(\triangle_1 \cap \triangle_2)^{\perp\Omega} = \triangle_1^{\perp\Omega} + \triangle_2^{\perp\Omega}$ etc.

18.4 Foliations

Definition 18.8 *Let M be a smooth manifold modeled on* M *and assume that* $\mathsf{M} = \mathsf{E} \times \mathsf{F}$. *A* ***foliation*** $\mathcal{F}_M$ *of M (or on M) is a partition of M into a family of disjoint subsets connected $\{\mathcal{L}_\alpha\}_{\alpha \in A}$ such that for every $p \in M$, there is a chart centered at p of the form $\varphi : U \to V \times W \subset \mathsf{E} \times \mathsf{F}$ with the property that for each $\mathcal{L}_\alpha$ the connected components $(U \cap \mathcal{L}_\alpha)_\beta$ of $U \cap \mathcal{L}_\alpha$ are given by*

$$\varphi((U \cap \mathcal{L}_\alpha)_\beta) = V \times \{c_{\alpha,\beta}\}$$

where $c_{\alpha,\beta} \in W \subset \mathsf{F}$ are constants. These charts are called distinguished charts for the foliation or ***foliation charts****. The connected sets $\mathcal{L}_\alpha$ are called the* ***leaves*** *of the foliation while for a given chart as above the connected components $(U \cap \mathcal{L}_\alpha)_\beta$ are called* ***plaques****.*

Recall that the connected components $(U \cap \mathcal{L}_\alpha)_\beta$ of $U \cap \mathcal{L}_\alpha$ are of the form $C_x(U \cap \mathcal{L}_\alpha)$ for some $x \in \mathcal{L}_\alpha$. An important point is that a fixed leaf $\mathcal{L}_\alpha$ may intersect a given chart domain U in many, even an infinite number of disjoint connected pieces no matter how small U is taken to be. In fact, it may be that $C_x(U \cap \mathcal{L}_\alpha)$ is dense in U. On the other hand, each $\mathcal{L}_\alpha$ is connected by definition. The usual first example of this behavior is given by the irrationally foliated torus. Here we take $M = T^2 := S^1 \times S^1$ and let the leaves be given as the image of the immersions $\iota_a : t \mapsto (e^{\mathrm{i}at}, e^{\mathrm{i}t})$ where a is a real numbers. If a is irrational then the image $\iota_a(\mathbb{R})$ is a (connected) dense subset of $S^1 \times S^1$. On the other hand, even in this case there are an infinite number of distinct leaves.

It may seem that a foliated manifold is just a special manifold but from one point of view a foliation is a generalization of a manifold. For instance, we can think of a manifold M as foliation where the points are the leaves. This is called the **discrete foliation** on M. At the other extreme a manifold may be thought of as a foliation with a single leaf $\mathcal{L} = M$ (the **trivial foliation**). We also have handy many examples of 1-dimensional foliations since given any global flow the orbits (maximal integral curves) are the leaves of a foliation. We also have the following special cases:

Example 18.1 *On a product manifold say $M \times N$ we have two complementary foliations:*

$$\{\{p\} \times N\}_{p \in M}$$

and

$$\{M \times \{q\}\}_{q \in N}$$

Example 18.2 *Given any submersion $f : M \to N$ the level sets $\{f^{-1}(q)\}_{q \in N}$ form the leaves of a foliation. The reader will soon realize that any foliation is given locally by submersions. The global picture for a general foliation can be very different from what can occur with a single submersion.*

Example 18.3 *The fibers of any vector bundle foliate the total space.*

Example 18.4 *(Reeb foliation) Consider the strip in the plane given by* $\{(x, y) : |x| \leq 1\}$*. For* $a \in \mathbb{R} \cup \{\pm\infty\}$ *we form leaves* $\mathcal{L}_a$ *as follows:*

$$\mathcal{L}_a := \{(x, a + f(x)) : |x| \leq 1\} \text{ for } a \in \mathbb{R}$$
$$\mathcal{L}_{\pm\infty} := \{(\pm 1, y) : |y| \leq 1\}$$

where $f(x) := \exp\left(\frac{x^2}{1-x^2}\right) - 1$*. By rotating this symmetric foliation about the* y *axis we obtain a foliation of the solid cylinder. This foliation is such that translation of the solid cylinder* C *in the* y *direction maps leaves diffeomorphically onto leaves and so we may let* $\mathbb{Z}$ *act on* C *by* $(x, y, z) \mapsto (x, y + n, z)$ *and then* $C/\mathbb{Z}$ *is a solid torus with a foliation called the Reeb foliation.*

Example 18.5 *The one point compactification of* $\mathbb{R}^3$ *is homeomorphic to* $S^3 \subset \mathbb{R}^4$*. Thus* $S^3 - \{p\} \cong \mathbb{R}^3$ *and so there is a copy of the Reeb foliated solid torus inside* S^3*. The complement of a solid torus in* S^3 *is another solid torus. It is possible to put another Reeb foliation on this complement and thus foliate all of* S^3*. The only compact leaf is the torus that is the common boundary of the two complementary solid tori.*

Exercise 18.2 *Show that the set of all* $v \in TM$ *such that* $v = T\varphi^{-1}(v, 0)$ *for some* $v \in \mathsf{E}$ *and some foliated chart* φ *is a (smooth) subbundle of* TM *that is also equal to* $\{v \in TM : v \text{ is tangent to a leaf}\}$.

Definition 18.9 *The* ***tangent bundle of a foliation*** $\mathcal{F}_M$ *with structure pseudogroup* $\mathcal{G}_{\mathsf{M},\mathsf{F}}$ *is the subbundle* $T\mathcal{F}_M$ *of* TM *defined by*

$$\begin{aligned} T\mathcal{F}_M &:= \{v \in TM : v \text{ is tangent to a leaf}\} \\ &= \{v \in TM : v = T\varphi^{-1}(v, 0) \text{ for some } v \in \mathsf{E} \text{ and some foliated chart } \varphi\} \end{aligned}$$

18.5 The Global Frobenius Theorem

The first step is to show that the (local) integral submanifolds of an integrable regular distribution can be glued together to form maximal integral submanifolds. These will form the leaves of a distribution.

Exercise 18.3 *If* Δ *is an integrable regular distribution of* TM*, then for any two local integral submanifolds* S_1 *and* S_2 *of* Δ *that both contain a point* x_0*, there is an open neighborhood* U *of* x_0 *such that*

$$S_1 \cap U = S_2 \cap U$$

Theorem 18.2 *If* Δ *is a subbundle of* TM *(i.e. a regular distribution) then the following are equivalent:*
1) Δ *is involutive.*
2) Δ *is integrable.*
3) There is a foliation $\mathcal{F}_M$ *on* M *such that* $T\mathcal{F}_M = \Delta$*.*

Proof. The equivalence of (1) and (2) is the local Frobenius theorem already proven. Also, the fact that (3) implies (2) is follows from 18.2. Finally, assume that (2) holds so that Δ is integrable. Recall that each (local) integral submanifold is an immersed submanifold which carries a submanifold topology generated by the connected components of the intersections of the integral submanifolds with chart domains. Any integral submanifold S has a smooth structure given by restricting charts U, ψ on M to connected components of $S \cap U$ (not on all of $S \cap U$!). Recall that a local integral submanifold is a regular submanifold (we are not talking about maximal immersed integral submanifolds!). Thus we may take U small enough that $S \cap U$ is connected. Now if we take two (local) integral submanifolds S_1 and S_2 of Δ and any point $x_0 \in S_1 \cap S_2$ (assuming this is nonempty) then a small enough chart U, ψ with $x_0 \in U$ induces a chart $U \cap S_1, \psi|_{U \cap S_1}$ on S_1 and a chart $C_{x_0}(U \cap S_2), \psi|_{C_{x_0}(U \cap S_2)}$ on S_2. But as we know $S_1 \cap U = S_2 \cap U$ and the overlap is smooth. Thus the union $S_1 \cup S_2$ is a smooth manifold with charts given by $U \cap (S_1 \cup S_2), \psi|_{U \cap (S_1 \cup S_2)}$ and the overlap maps are $U \cap (S_1 \cap S_2), \psi|_{U \cap (S_1 \cap S_2)}$. We may extend to a maximal connected integral submanifold using Zorn's lemma be we can see the existence more directly. Let $\mathcal{L}_a(x_0)$ be the set of all points that can be connected to x_0 by a smooth path $c : [0, 1] \to M$ with the property that for any $t_0 \in [0, 1]$, the image $c(t)$ is contained inside a (local) integral submanifold for all t sufficiently near t_0. Using what we have seen about gluing intersecting integral submanifold together and the local uniqueness of such integral submanifolds we see that $\mathcal{L}_a(x_0)$ is a smooth connected immersed integral submanifold that must be the maximal connected integral submanifold containing x_0. Now since x_0 was arbitrary there is a maximal connected integral submanifold containing any point of M. By construction we have that the foliation $\mathcal{L}$ given by the union of all these leaves satisfies (3). ∎

There is an analogy between the notion of a foliation on a manifold and a differentiable structure on a manifold.

From this point of view we think of a foliation as being given by a maximal **foliation atlas** that is defined to be a cover of M by foliated charts. The compatibility condition on such charts is that when the domains of two foliation charts, say $\varphi_1 : U_1 \to V_1 \times W_1$ and $\varphi_2 : U_2 \to V_2 \times W_2$, then the overlap map has the form

$$\varphi_2 \circ \varphi_1^{-1}(x, y) = (f(x, y), g(y)).$$

A plaque in a chart $\varphi_1 : U_1 \to V_1 \times W_1$ is a connected component of a set of the form $\varphi_1^{-1}\{(x, y) : y = \text{constant}\}$.

18.6 Singular Distributions

Lemma 18.3 *Let $X_1, ..., X_n$ be vector fields defined in a neighborhood of $x \in M$ such that $X_1(x), ..., X_n(x)$ are a basis for T_xM and such that $[X_i, X_j] = 0$ in a neighborhood of x. Then there is an open chart $U, \psi = (y^1, ..., y^n)$ containing x such that $X_i|_U = \frac{\partial}{\partial y^i}$.*

Proof. For a sufficiently small ball $B(0,\epsilon) \subset \mathbb{R}^n$ and $t = (t_1, ..., t_n) \in B(0,\epsilon)$ we define

$$f(t_1, ..., t_n) := Fl_{t_1}^{X_1} \circ \cdots \circ Fl_{t_n}^{X_n}(x).$$

By theorem 7.9 the order that we compose the flows does not change the value of $f(t_1, ..., t_n)$. Thus

$$\begin{aligned}
&\frac{\partial}{\partial t_i} f(t_1, ..., t_n) \\
&= \frac{\partial}{\partial t_i} Fl_{t_1}^{X_1} \circ \cdots \circ Fl_{t_n}^{X_n}(x) \\
&= \frac{\partial}{\partial t_i} Fl_{t_i}^{X_i} \circ Fl_{t_1}^{X_1} \circ \cdots \circ Fl_{t_n}^{X_n}(x) \text{ (put the } i\text{-th flow first)} \\
&X_i(Fl_{t_1}^{X_1} \circ \cdots \circ Fl_{t_n}^{X_n}(x)).
\end{aligned}$$

Evaluating at $t = 0$ shows that $T_0 f$ is nonsingular and so $(t_1, ..., t_n) \mapsto f(t_1, ..., t_n)$ is a diffeomorphism on some small open set containing 0. The inverse of this map is the coordinate chart we are looking for (check this!). ∎

Definition 18.10 *Let $\mathfrak{X}_{loc}(M)$ denote the set of all sections of the presheaf $\mathfrak{X}_M$. That is*

$$\mathfrak{X}_{loc}(M) := \bigcup_{open\ U \subset M} \mathfrak{X}_M(U).$$

Also, for a distribution Δ let $\mathfrak{X}_\Delta(M)$ denote the subset of $\mathfrak{X}_{loc}(M)$ consisting of local fields X with the property that $X(x) \in \Delta_x$ for every x in the domain of X.

Definition 18.11 *We say that a subset of local vector fields $\mathcal{X} \subset \mathfrak{X}_\Delta(M)$ **spans** a distribution Δ if for each $x \in M$ the subspace Δ_x is spanned by $\{X(x) : X \in \mathcal{X}\}$.*

If Δ is a smooth distribution (and this is all we shall consider) then $\mathfrak{X}_\Delta(M)$ spans Δ. On the other hand, as long as we make the convention that the empty set spans the set $\{0\}$ for every vector space we are considering, then any $\mathcal{X} \subset \mathfrak{X}_\Delta(M)$ spans some smooth distribution which we denote by $\Delta(\mathcal{X})$.

Definition 18.12 *An immersed integral submanifold of a distribution Δ is an injective immersion $\iota : S \to M$ such that $T_s\iota(T_sS) = \Delta_{\iota(s)}$ for all $s \in S$. An immersed integral submanifold is called **maximal** its image is not properly contained in the image of any other immersed integral submanifold.*

Since an immersed integral submanifold is an injective map we can think of S as a subset of M. In fact, it will also turn out that an immersed integral submanifold is automatically smoothly universal so that the image $\iota(S)$ is an initial submanifold. Thus in the end, we may as well assume that $S \subset M$ and that $\iota : S \to M$ is the inclusion map. Let us now specialize to the finite dimensional case. Note however that we do *not* assume that the rank of the distribution is constant.

Now we proceed with our analysis. If $\iota : S \to M$ is an immersed integral submanifold and of a distribution $\triangle$ then if $X \in \mathfrak{X}_\Delta(M)$ we can make sense of $\iota^* X$ as a local vector field on S. To see this let U be the domain of X and take $s \in S$ with $\iota(s) \in U$. Now $X(\iota(s)) \in T_s\iota(T_sS)$ we can define

$$\iota^* X(s) := (T_s\iota)^{-1} X(\iota(s)).$$

$\iota^* X(s)$ is defined on some open set in S and is easily seen to be smooth by considering the local properties of immersions. Also, by construction $\iota^* X$ is ι related to X.

Next we consider what happens if we have two immersed integral submanifolds $\iota_1 : S_1 \to M$ and $\iota_2 : S_2 \to M$ such that $\iota_1(S_1) \cap \iota_2(S_2) \neq \emptyset$. By proposition 7.1 we have

$$\iota_i \circ \mathrm{Fl}_t^{\iota_i^* X} = \mathrm{Fl}_t^X \circ \iota_i \text{ for } i = 1, 2.$$

Now if $x_0 \in \iota_1(S_1) \cap \iota_2(S_2)$ then we choose s_1 and s_2 such that $\iota_1(s_1) = \iota_2(s_2) = x_0$ and pick local vector fields $X_1, ..., X_k$ such that $(X_1(x_0), ..., X_k(x_0))$ is a basis for $\triangle_{x_0}$. For $i = 1$ and 2 we define

$$f_i(t^1, ..., t^k) := (\mathrm{Fl}_{t^1}^{\iota_i^* X_1} \circ \cdots \circ \mathrm{Fl}_{t^k}^{\iota_i^* X_k})$$

and since $\frac{\partial}{\partial t^j}\big|_0 f_i = \iota_i^* X_j$ for $i = 1, 2$ and $j = 1, ..., k$ we conclude that f_i, $i = 1, 2$ are diffeomorphisms when suitable restricted to a neighborhood of $0 \in \mathbb{R}^k$. Now we compute:

$$\begin{aligned}
(\iota_2^{-1} \circ \iota_1 \circ f_1)(t^1, ..., t^k) &= (\iota_2^{-1} \circ \iota_1 \circ \mathrm{Fl}_{t^1}^{\iota_1^* X_1} \circ \cdots \circ \mathrm{Fl}_{t^k}^{\iota_1^* X_k})(x_1) \\
&= (\iota_2^{-1} \mathrm{Fl}_{t^1}^{X_1} \circ \cdots \circ \mathrm{Fl}_{t^k}^{X_k} \circ \iota_1)(x_1) \\
&= (\mathrm{Fl}_{t^1}^{\iota_2^* X_1} \circ \cdots \circ \mathrm{Fl}_{t^k}^{\iota_2^* X_k} \circ \iota_2^{-1} \circ \iota_1)(x_1) \\
&= f_2(t^1, ..., t^k).
\end{aligned}$$

Now we can see that $\iota_2^{-1} \circ \iota_1$ is a diffeomorphism. This allows us to glue together the all the integral manifolds that pass through a fixed x in M to obtain a unique maximal integral submanifold through x. We have prove the following result:

Proposition 18.1 *For a smooth distribution Δ on M and any $x \in M$ there is a unique maximal integral manifold L_x containing x called the **leaf** through x.*

Definition 18.13 *Let $\mathcal{X} \subset \mathfrak{X}_{loc}(M)$. We call X a **stable** family of local vector fields if for any $X, Y \in \mathcal{X}$ we have*

$$(\mathrm{Fl}_t^X)^* Y \in \mathcal{X}$$

whenever $(\mathrm{Fl}_t^X)^* Y$ is defined. Given an arbitrary subset of local fields $\mathcal{X} \subset \mathfrak{X}_{loc}(M)$ let $\mathcal{S}(\mathcal{X})$ denote the set of all local fields of the form

$$(\mathrm{Fl}_{t^1}^{X_1} \circ \mathrm{Fl}_{t^2}^{X_2} \circ \cdots \circ \mathrm{Fl}_{t^t}^{X_k})^* Y$$

where $X_i, Y \in \mathcal{X}$ and where $t = (t^1, ..., t^k)$ varies over all k-tuples such that the above expression is defined.

Exercise 18.4 *Show that $\mathcal{S}(\mathcal{X})$ is the smallest stable family of local vector fields containing $\mathcal{X}$.*

Definition 18.14 *If a diffeomorphism ϕ of a manifold M with a distribution Δ is such that $T_x\phi(\Delta_x) \subset \Delta_{\phi(x)}$ for all $x \in M$ then we call ϕ an* ***automorphism of*** *Δ. If $\phi : U \to \phi(U)$ is such that $T_x\phi(\Delta_x) \subset \Delta_{\phi(x)}$ for all $x \in U$ we call ϕ a* ***local automorphism of*** *Δ.*

Definition 18.15 *If $X \in \mathfrak{X}_{loc}(M)$ is such that $T_x\mathrm{Fl}_t^X(\Delta_x) \subset \Delta_{\mathrm{Fl}_t^X(x)}$ we call X a (local)* ***infinitesimal automorphism*** *of Δ. The set of all such is denoted $\mathrm{aut}_{loc}(\Delta)$.*

Example 18.6 *Convince yourself that $\mathrm{aut}_{loc}(\Delta)$ is stable.*

For the next theorem recall the definition of $\mathfrak{X}_\Delta$.

Theorem 18.3 *Let Δ be a smooth singular distribution on M. Then the following are equivalent:*

1) Δ is integrable.

2) $\mathfrak{X}_\Delta$ is stable.

3) $\mathrm{aut}_{loc}(\Delta) \cap \mathfrak{X}_\Delta$ spans Δ.

4) There exists a family $\mathcal{X} \subset \mathfrak{X}_{loc}(M)$ such that $\mathcal{S}(\mathcal{X})$ spans Δ.

Proof. Assume (1) and let $X \in \mathfrak{X}_\Delta$. If $\mathcal{L}_x$ is the leaf through $x \in M$ then by proposition 7.1

$$\mathrm{Fl}_{-t}^X \circ \iota = \iota \circ \mathrm{Fl}_{-t}^{\iota^* X}$$

where $\iota : \mathcal{L}_x \hookrightarrow M$ is inclusion. Thus

$$\begin{aligned} T_x(\mathrm{Fl}_{-t}^X)(\Delta_x) &= T(\mathrm{Fl}_{-t}^X) \cdot T_x\iota \cdot (T_x\mathcal{L}_x) \\ &= T(\iota \circ \mathrm{Fl}_{-t}^{\iota^* X}) \cdot (T_x\mathcal{L}_x) \\ &= T\iota T_x(\mathrm{Fl}_{-t}^{\iota^* X}) \cdot (T_x\mathcal{L}_x) \\ &= T\iota T_{\mathrm{Fl}_{-t}^{\iota^* X}(x)}\mathcal{L}_x = \Delta_{\mathrm{Fl}_{-t}^{\iota^* X}(x)}. \end{aligned}$$

Now if Y is in $\mathfrak{X}_\Delta$ then at an arbitrary x we have $Y(x) \in \Delta_x$ and so the above shows that $((\mathrm{Fl}_t^X)^*Y)(x) \in \Delta$ so $(\mathrm{Fl}_t^X)^*Y)$ is in $\mathfrak{X}_\Delta$. We conclude that $\mathfrak{X}_\Delta$ is stable and have shown that $(1) \Rightarrow (2)$.

Next, if (2) hold then $\mathfrak{X}_\Delta \subset \mathrm{aut}_{loc}(\Delta)$ and so we have (3).

If (3) holds then we let $\mathcal{X} := \mathrm{aut}_{loc}(\Delta) \cap \mathfrak{X}_\Delta$. Then for $Y, Y \in \mathcal{X}$ we have $(\mathrm{Fl}_t^X)^*Y \in \mathfrak{X}_\Delta$ and so $\mathcal{X} \subset \mathcal{S}(\mathcal{X}) \subset \mathfrak{X}_\Delta$. from this we see that since $\mathcal{X}$ and $\mathfrak{X}_\Delta$ both span Δ so does $\mathcal{S}(\mathcal{X})$.

Finally, we show that (4) implies (1). Let $x \in M$. Since $\mathcal{S}(\mathcal{X})$ spans the distribution and is also stable by construction we have

$$T(\mathrm{Fl}_t^X)\Delta_x = \Delta_{\mathrm{Fl}_t^X(x)}$$

for all fields X from $\mathcal{S}(\mathcal{X})$. Let the dimension Δ_x be k and choose fields $X_1, ..., X_k \in \mathcal{S}(\mathcal{X})$ such that $X_1(x), ..., X_k(x)$ is a basis for Δ_x. Define a map $f :: \mathbb{R}^k \to M$ by

$$f(t^1, ..., t^n) := (\mathrm{Fl}_{t^1}^{X_1} \mathrm{Fl}_{t^2}^{X_2} \circ \cdots \circ \mathrm{Fl}_{t^k}^{X_k})(x)$$

which is defined (and smooth) near $0 \in \mathbb{R}^k$. As in lemma 18.3 we know that the rank of f at 0 is k and the image of a small enough open neighborhood of 0 is a submanifold. In fact, this image, say $S = f(U)$ is an integral submanifold of Δ through x. To see this just notice that the $T_x S$ is spanned by $\frac{\partial f}{\partial t^j}(0)$ for $j = 1, 2, ..., k$ and

$$\begin{aligned}
\frac{\partial f}{\partial t^j}(0) &= \left.\frac{\partial}{\partial t^j}\right|_0 (\mathrm{Fl}_{t^1}^{X_1} \mathrm{Fl}_{t^2}^{X_2} \circ \cdots \circ \mathrm{Fl}_{t^k}^{X_k})(x) \\
&= T(\mathrm{Fl}_{t^1}^{X_1} \mathrm{Fl}_{t^2}^{X_2} \circ \cdots \circ \mathrm{Fl}_{t^{j-1}}^{X_{j-1}}) X_j ((\mathrm{Fl}_{t^j}^{X_j} \mathrm{Fl}_{t^{j+1}}^{X_{j+1}} \circ \cdots \circ \mathrm{Fl}_{t^k}^{X_k})(x)) \\
&= ((\mathrm{Fl}_{-t^1}^{X_1})^* (\mathrm{Fl}_{-t^2}^{X_2})^* \circ \cdots \circ (\mathrm{Fl}_{-t^{j-1}}^{X_{j-1}})^* X_j)(f(t^1, ..., t^n)).
\end{aligned}$$

But $\mathcal{S}(\mathcal{X})$ is stable so each $\frac{\partial f}{\partial t^j}(0)$ lies in $\Delta_{f(t)}$. From the construction of f and remembering **??** we see that $\mathrm{span}\{\frac{\partial f}{\partial t^j}(0)\} = T_{f(t)} S = \Delta_{f(t)}$ and we are done. ■

18.7 Problem Set

1. Let H be the Heisenberg group consisting of all matrices of the form

$$A = \begin{bmatrix} 1 & x_{12} & x_{13} \\ 0 & 1 & x_{23} \\ 0 & 0 & 1 \end{bmatrix}.$$

   That is, the upper-diagonal 3×3 real matrices with 1's on the diagonal. Let V_{12}, V_{13}, V_{23} be the left-invariant vector-fields on H that have values at the identity $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$, respectively. Let $\Delta_{\{V_{12}, V_{13}\}}$ and $\Delta_{\{V_{12}, V_{23}\}}$ be the $2-$dimensional distributions generated by the indicated pairs of vector fields. Show that $\Delta_{\{V_{12}, V_{13}\}}$ is integrable and $\Delta_{\{V_{12}, V_{23}\}}$ is not.

Chapter 19

Connections and Covariant Derivatives

19.1 Definitions

The notion of a "connection" on a vector bundle and the closely related notion of "covariant derivative" can be approached in so many different ways that we shall only be able to cover a small portion of the subject. "covariant derivative" and "connection" are sometimes treated as synonymous. In fact, a covariant derivative is sometimes called a Kozsul connection (or even just a connection; the terms are conflated). At first, we will make a distinction but eventually we will use the word connection to refer to both concepts. In any case, the central idea is that of measuring the rate of change of fields on bundles in the direction of a vector or vector field on the base manifold.

A covariant derivative (Koszul connection) can either be defined as a map $\nabla : \mathfrak{X}(M) \times \Gamma(M, E) \to \Gamma(M, E)$ with certain properties from which one deduces a well defined map $\nabla : TM \times \Gamma(M, E) \to TM$ with nice properties or the other way around. We also hope that the covariant derivative is natural with respect to restrictions to open sets and so a sheaf theoretic definition could be given. For finite dimensional manifolds the several approaches are essentially equivalent.

Definition 19.1 (I) *A **Koszul connection** or **covariant derivative** on a $C^\infty-$vector bundle $E \to M$ is a map $\nabla : \mathfrak{X}(M) \times \Gamma(M, E) \to \Gamma(M, E)$ (where $\nabla(X, s)$ is written as $\nabla_X s$) satisfying the following four properties*

i) $\nabla_{fX}(s) = f\nabla_X s$ for all $f \in C^\infty$, $X \in \mathfrak{X}(M)$ and $s \in \Gamma(M, E)$

ii) $\nabla_{X_1+X_2} s = \nabla_{X_1} s + \nabla_{X_2} s$ for all $X_1, X_2 \in \mathfrak{X}(M)$ and $s \in \Gamma(M, E)$

iii) $\nabla_X(s_1 + s_2) = \nabla_X s_1 + \nabla_X s_2$ for all $X \in \mathfrak{X}(M)$ and $s_1, s_2 \in \Gamma(M, E)$

iv) $\nabla_X(fs) = (Xf)s + f\nabla_X(s)$ for all $f \in C^\infty$, $X \in \mathfrak{X}(M)$ and $s \in \Gamma(M, E)$

*For a fixed $X \in \mathfrak{X}(M)$ the map $\nabla_X : \Gamma(M, E) \to \Gamma(M, E)$ is called the **covariant derivative with respect to** X.*

As we will see below, for finite dimensional E and M this definition is enough to imply that ∇ induces maps $\nabla^U : \mathfrak{X}(U)\times\Gamma(U,E) \to \Gamma(U,E)$ that are naturally related in the sense we make precise below. Furthermore the value $(\nabla_X s)(p)$ depends only on the value X_p and only on the values of s along any smooth curve c representing X_p. The proof of these facts depends on the existence of cut-off functions. We have already developed the tools to obtain the proof easily in sections 4.9 and 8.5 and so we leave the verification of this to the reader. In any case we shall take a different route.

We would also like to be able to differentiate sections of a vector bundle along maps $f : N \to M$.

In the case of the tangent bundle of $\mathbb{R}^n$ one can identify vector fields with maps X:$\mathbb{R}^n \to \mathbb{R}^n$ and thus it makes sense to differentiate a vector field just as we would a function. For instance, if $\mathsf{X} = (f^1, ..., f^n)$ then we can define the directional derivative in the direction of v at $p \in T_p\mathbb{R}^n \cong \mathbb{R}^n$ by $D_v\mathsf{X} =(D_pf^1 \cdot v, ..., D_pf^n \cdot v)$ and we get a vector in $T_p\mathbb{R}^n_\nu$ as an answer. Taking the derivative of a vector field seems to require involve the limit of difference quotient of the type

$$\lim_{t\to 0} \frac{X(p + tv) - X(p)}{t}$$

and yet how can we interpret this in a way that makes sense for a vector field on a general manifold? One problem is that $p + tv$ makes no sense if the manifold isn't a vector space. This problem is easily solve by replacing $p + tv$ by $c(t)$ where $\dot{c}(0) = v$ and $c(0) = p$. We still have the more serious problem that $X(c(t)) \in T_{c(t)}M$ while $X(p) = X(c(0)) \in T_pM$. The difficulty is that $T_{c(t)}M$ is not likely to be the same vector space as T_pM and so what sense does $X(c(t)) - X(p)$ make? In the case of a vector space (like $\mathbb{R}^n$) every tangent space is canonically isomorphic to the vector space itself so there is sense to be made of a difference quotient involving vectors from different tangent spaces. In order to get an idea of how we might define a covariant derivative on a general manifold, let us look again at the case of a submanifold M of $\mathbb{R}^n$. Let $X \in \mathfrak{X}(M)$ and $v \in T_pM$. Form a curve with $\dot{c}(0) = v$ and $c(0) = p$ and consider the composition $X \circ c$. Since every vector tangent to M is also a vector in $\mathbb{R}^n$ we can consider $X \circ c$ to take values in $\mathbb{R}^n$ and then take the derivative

$$\left.\frac{d}{dt}\right|_0 X \circ c.$$

This is well defined but while $X \circ c(t) \in T_{c(t)}M \subset T_{c(t)}\mathbb{R}^n$ we only know that $\left.\frac{d}{dt}\right|_0 X \circ c \in T_p\mathbb{R}^n$. It would be more nature if the result of differentiation of a vector field tangent to M should have been in T_pM. The simple solution is to take the orthogonal projection of $\left.\frac{d}{dt}\right|_0 X\circ c$ onto $T_{c(0)}M$. Denote this orthogonal projection of a vector onto its tangent part by $X \mapsto X^\top$. Our definition of a covariant derivative operator on M induced by D is then

$$\nabla_v X := \left(\left.\frac{d}{dt}\right|_0 X \circ c \right)^\top \in T_pM.$$

This turns out to be a very good definition. In fact, one may easily verify that we have the following properties:

1. (Smoothness) If X and Y are smooth vector fields then the map
$$p \mapsto \nabla_{X_p} Y$$
is also a smooth vector field on M. This vector filed is denoted $\nabla_X Y$.

2. (Linearity over $\mathbb{R}$ in second "slot") For two vector fields X and Y and any $a, b \in \mathbb{R}$ we have
$$\nabla_v(aX_1 + bX_2) = a\nabla_v X_1 + b\nabla_v X_2.$$

3. (Linearity over $C^\infty(M)$ in first "slot") For any three vector fields X, Y and Z and any $f, g \in C^\infty(M)$ we have
$$\nabla_{fX+gY} Z = f\nabla_X Z + g\nabla_Y Z.$$

4. (Product rule) For $v \in T_pM$, $X \in \mathfrak{X}(M)$ and $f \in C^\infty(M)$ we have
$$\begin{aligned}\nabla_v fX &= f(p)\nabla_v X + (vf)X(p) \\ &= f(p)\nabla_v X + df(v)X(p).\end{aligned}$$

5. $\nabla_v(X \cdot Y) = \nabla_v X \cdot Y + X \cdot \nabla_v Y$ for all v, X, Y.

Notice that if $p \mapsto \nabla_{X_p} Y$ is denoted by $\nabla_X Y$ then $\nabla : X, Y \longmapsto \nabla_X Y$ is a Koszul connection 19.1. But is 19.1 the best abstraction of the properties we need? In finding the correct abstraction we could use the properties of our example on the tangent bundle of a submanifold of $\mathbb{R}^n$ with the aim of defining a so called covariant derivative it is a bit unclear whether we should define $\nabla_X Y$ for a pair of fields X, Y or define $\nabla_v X$ for a tangent vector v and a field X. Different authors take different approaches but it turns out that one can take either approach and, at least in finite dimensions, end up with equivalent notions. We shall make the following our basic definition of a covariant derivative.

Definition 19.2 *A* ***natural covariant derivative*** ∇ *on a smooth vector bundle* $E \to M$ *is an assignment to each open set* $U \subset M$ *of a map* $\nabla^U : \mathfrak{X}(U) \times \Gamma(U, E) \to \Gamma(U, E)$ *written* $\nabla^U : (X, \sigma) \to \nabla^U_X \sigma$ *such that the following hold:*

1. $\nabla^U_X \sigma$ *is* $C^\infty(U)$*-linear in* X,

2. $\nabla^U_X \sigma$ *is* $\mathbb{R}$*-linear in* σ,

3. $\nabla^U_X(f\sigma) = f\nabla^U_X \sigma + (Xf)\sigma$ *for all* $X \in \mathfrak{X}(U)$, $\sigma \in \Gamma(U, E)$ *and all* $f \in C^\infty(U)$.

4. *If $V \subset U$ then $r_V^U(\nabla_X^U \sigma) = \nabla_{r_V^U X}^V r_V^U \sigma$ (naturality with respect to restrictions $r_V^U : \sigma \mapsto \sigma|_V$).*

5. *$(\nabla_X^U Y)(p)$ only depends of the value of X at p (infinitesimal locality).*

Here $\nabla_X^U Y$ is called the covariant derivative of Y with respect to X. We will denote all of the maps ∇^U by the single symbol ∇ when there is no chance of confusion.

We have worked naturality with respect to restrictions and infinitesimal locality into the definition of a natural covariant derivative in order to have a handy definition for infinite dimensional manifolds which circumvents the problem of localization. We will show that on finite dimensional manifolds a Koszul connections induces a natural covariant derivative. We have the following intermediate result still stated for possibly infinite dimensional manifolds.

Lemma 19.1 *Suppose that M admits cut-off functions and $\nabla : \mathfrak{X}(M) \times \Gamma(E) \to \Gamma(E)$ is such that (1), (2), and (3) hold (for $U = M$). Then if on some open U either $X = 0$ or $\sigma = 0$ then*

$$(\nabla_X \sigma)(p) = 0 \text{ for all } p \in U.$$

Proof. We prove the case of $\sigma|_U = 0$ and leave the case of $X|_U = 0$ to the reader.

Let $q \in U$. Then there is some function f that is identically one on a neighborhood $V \subset U$ of q and that is zero outside of U thus $f\sigma \equiv 0$ on M and so since ∇ is linear we have $\nabla(f\sigma) \equiv 0$ on M. Thus since (3) holds for global fields we have

$$\begin{aligned} \nabla(f\sigma)(q) &= f(p)(\nabla_X \sigma)(q) + (X_q f)\sigma(q) \\ &= (\nabla_X \sigma)(q) = 0. \end{aligned}$$

Since $q \in U$ was arbitrary we have the result. ∎

In the case of finite dimensional manifolds we have

Proposition 19.1 *Let M be a finite dimensional smooth manifold. Suppose that there exist an operator $\nabla : \mathfrak{X}(M) \times \Gamma(E) \to \Gamma(E)$ such that (1), (2), and (3) hold (for $U = M$) for example a Koszul connection. If we set $\nabla_X^U \sigma := r_U^M(\nabla_{\widetilde{X}}^M \widetilde{\sigma})$ for any extensions $\widetilde{X}$ and $\widetilde{\sigma}$ of X and $\sigma \in \Gamma(U, E)$ to global fields then $U \mapsto \nabla^U$ is a natural covariant derivative.*

Proof. By the previous lemma $\nabla_X^U \sigma := r_U^M(\nabla_{\widetilde{X}}^M \widetilde{\sigma})$ is a well defined operator that is easily checked to satisfy (1), (2), (3), and (4) of definition 19.2. We now prove property (5). Let $\alpha \in T^*E$ and fix $\sigma \in \Gamma(U, E)$. define a map $\mathfrak{X}(U) \to C^\infty(U)$ by $X \mapsto \alpha(\nabla_X^U \sigma)$. By theorem 4.1 we see that $\alpha(\nabla_X^U \sigma)$ depend only on the value of X at $p \in U$. ∎

Since many authors only consider finite dimensional manifolds they define a covariant to be a map $\nabla : \mathfrak{X}(M) \times \Gamma(E) \to \Gamma(E)$ satisfying (1), (2), and (3).

It is common to write expressions like $\nabla_{\frac{\partial}{\partial x^i}}\sigma$ where σ is a global field and $\frac{\partial}{\partial x^i}$ is defined only on a coordinate domain U. This still makes sense as a field $p \mapsto \nabla_{\frac{\partial}{\partial x^i}(p)}\sigma$ on U by virtue of (5) or by interpreting $\nabla_{\frac{\partial}{\partial x^i}}\sigma$ as $\nabla_{\frac{\partial}{\partial x^i}}\ \sigma|_U$ and invoking (4) if necessary.

In the next section we indicate a local condition which guarantees that we will have a natural covariant derivative on $\pi : E \to M$. We now introduce the notion of a system of connection forms.

19.2 Connection Forms

Let $\pi : E \to M$ be a rank r vector bundle with a connection ∇. Recall that a choice of a local frame field over an open set $U \subset M$ is equivalent to a trivialization of the restriction $\pi_U : E|_U \to U$. Namely, if $\Phi = (\pi, \phi)$ is such a trivialization over U then defining $e_i(x) = \Phi^{-1}(x, \mathsf{e}_i)$ where (e_i) is the standard basis of $\mathbb{F}^n$ we have that $\sigma = (e_1, ..., e_k)$. We now examine the expression for the connection from the viewpoint of such a local frame. It is not hard to see that for every such frame field there must be a matrix of 1-forms $A_\sigma = (A^i_j)_{1\le i,j\le r}$ such that for $X \in \Gamma(U, E)$ we may write

$$\nabla_X e_i = A^b_i(X)e_b. \text{ (sum over } b)$$

For a section $s = s^i e_i$

$$\begin{aligned}
\nabla_X s &= \nabla_X(s^i e_i) \\
&= (Xs^a)e_a + s^a \nabla_X e_a \\
&= (Xs^i)e_i + s^i A^j_i(X)e_j \\
&= (Xs^i)e_i + s^r A^i_r(X)e_i \\
&= (Xs^i + A^i_r(X)s^r)e_i
\end{aligned}$$

So the $a-$component of $\nabla_X s$ is $(\nabla_X s)^i = Xs^i + A^i_r(X)s^r$. Of course the frames are defined only locally say on some open set U. We may surely choose U small enough that it is also the domain of a moving frame $\{E_\mu\}$ for M. Thus we have $\nabla_{E_\mu} e_j = \Gamma^k_{\mu j} e_k$ where $A^k_{\mu j} = A^j_i(E_\mu)$. We now have the formula

$$\nabla_X s = (X^\mu E_\mu s^i + X^\mu A^i_r(E_\mu)s^r)e_i$$

and in the usual case where $E_\mu = \partial_\mu$ is a holonomic frame associated to a chart then $\nabla_X s = (X^\mu \partial_\mu s^i + X^\mu A^i_{\mu j} s^r)e_i$.

Now suppose that we have two moving frames whose domains overlap, say $\sigma = (e_j)$ and $\sigma' = (e'_j)$. Let us examine how the matrix of forms $A_\sigma = \left(A^j_i\right)$ is related to the forms $A_{\sigma'} = \left(A'^j_i\right)$ on this open set. The change of frame is

$$\sigma'^b = g^b_a \sigma_b$$

which in matrix notation is

$$\sigma' = \sigma g.$$

for some smooth $g : U \cap U' \to \mathrm{GL}(n)$. Differentiating both sides

$$\begin{aligned}
\sigma' &= \sigma g \\
\nabla \sigma' &= \nabla(\sigma g) \\
\sigma' A' &= (\nabla \sigma) g + \sigma dg \\
\sigma' A' &= \sigma g g^{-1} A g + \sigma g g^{-1} dg \\
\sigma' A' &= f \sigma' g^{-1} A g + \sigma' g^{-1} dg \\
A' &= g^{-1} A g + g^{-1} dg
\end{aligned}$$

Conversely, we have the following theorem which we state without proof:

Theorem 19.1 *Let $\pi : E \to M$ be a smooth $\mathbb{F}-$vector bundle of rank k. Suppose that for every moving frame $\sigma = (e_1, ..., e_k)$ we are given a matrix of $1-$forms A_σ so that $A_{\sigma g} = g^{-1} A_\sigma g + g^{-1} dg$. Then there is a unique covariant derivative ∇ on $\pi : E \to M$ such that for a moving frame σ on U_σ*

$$\nabla_X s = (X s^j + A_r^j(X) s^r) e_j$$

for $s = \sum s^j e_j$ and where $A_\sigma = (A_r^j)$.

Definition 19.3 *A family of matrix valued 1-forms related as in the previous theorem will be called a system of* ***connection forms****.*

Sometimes one hears that the A_σ are locally elements of $\Omega(M, \mathrm{End}(E))$ but the transformation law just discovered says that we cannot consider the forms A_σ as coming from a section of $\mathrm{End}(E)$. However, we have the following:

Proposition 19.2 *If A_α, U_α and A_α, U_α are two systems of connection forms for the same cover then the difference forms $\triangle A_\alpha = A_\alpha - A'_\alpha$ are the local expressions of a global $\triangle A \in \Omega(M, \mathrm{End}(E))$.*

$$\begin{aligned}
\triangle A_\alpha &= A_\alpha - A'_\alpha \\
&= g^{-1} A_\alpha g + g^{-1} dg - (g^{-1} A'_\alpha g + g^{-1} dg) \\
&= g^{-1} A_\alpha g - g^{-1} A'_\alpha g \\
&= g^{-1} (\triangle A_\alpha) g
\end{aligned}$$

so that the forms $\triangle A_\alpha$ are local expressions of a global section of $\Omega(M, \mathrm{End}(E))$

Exercise 19.1 *Show that the set of all connections on E is naturally an affine space $C(E)$ whose vector space of differences is* $\mathrm{End}(E)$. *For any fixed connection ∇^0 we have an affine isomorphism* $\mathrm{End}(E) \to C(E)$ *given by $\triangle A \mapsto \nabla^0 + \triangle A$.*

Now in the special case mentioned above for a trivial bundle, the connection form in the defining frame is zero and so in that case $\triangle A_\alpha = A_\alpha$. So in this case A_α determines a section of $\mathrm{End}(E)$. Now any bundle is locally trivial so in this sense A_α is locally in $\mathrm{End}(E)$. But this is can be confusing since we have changed (by force so to speak) the transformation law for A_α among frames defined on the same open set to that of a difference $\triangle A_\alpha$ rather than A_α. The point is that even though $\triangle A_\alpha$ and A_α are equal in the distinguished frame they are not the same after transformation to a new frame.

In the next section we give another way to obtain a covariant derivative which is more geometric and global in character.

19.3 Vertical and Horizontal

One way to get a natural covariant derivative on a vector bundle is through the notion of connection in another sense. To describe this we will need the notion of a vertical bundle. We give the definition not just for vector bundles but also for a general fiber bundle as this will be useful to us later.

Lemma 19.2 *Let $E \xrightarrow{\pi} M$ be a fiber bundle; fix $p \in M$. Set $N := E_p$, and let $\imath : N \hookrightarrow E$ be the inclusion. For all $\xi \in N$,*

$$T_\xi \imath \left(N_\xi\right) = \ker[T_\xi \imath : T_\xi E \to M_p] = \left(T_\xi \pi\right)^{-1}(0_p) \subset T_\xi E$$

where $0_p \in M_p$ is the zero vector. If $\varphi : N \to F$ is a diffeomorphism and $\mathtt{x}$ a chart on an open set V in F, then for all $\xi \in \varphi^{-1}(V)$, $d\mathtt{x} \circ T_\xi \varphi$ maps $T_\xi N$ isomorphically onto $\mathbb{R}^m$, where $m := \dim F$.

Proof. $\pi \circ \imath \circ \gamma$ is constant for each C^∞ curve γ in N, so $T\pi \cdot (T\imath \cdot \dot{\gamma}(0)) = 0_p$; thus $T_\xi \imath \left(N_\xi\right) \subset \left(T_\xi \pi\right)^{-1}(0_p)$. On the other hand,

$$\dim(\left(T_\xi \pi\right)^{-1}(0_p)) = \dim E - \dim M = \dim F = \dim N_\xi$$

so$(T_\xi \imath)(N) = \left(T_\xi \pi\right)^{-1}(0_p)$. The rest follows since $d\mathtt{x} = pr_2 \circ T\mathtt{x}$. ■

Let $\mathcal{V}_p E := \left(T_\xi \pi\right)^{-1}(0_p)$ and $\mathcal{V}_p E$.

Definition 19.4 *Let $\pi : E \to M$ be a fiber bundle with typical fiber F and $\dim F = m$. The **vertical bundle on** $\pi : E \to M$ is the real vector bundle $\pi_{\mathcal{V}} : \mathcal{V}\ E \to E$ with total space defined by the disjoint union*

$$\mathcal{V}\ E := \bigsqcup_{p \in E} \mathcal{V}_p E \subset TE$$

The projection map is defined by the restriction $\pi_{\mathcal{V}} := T\pi|\,\mathcal{V}\ E$. A vector bundle atlas on $\mathcal{V}E$ consists of the vector bundle charts of the form

$$(\pi_{\mathcal{V}}, d\mathtt{x} \circ T\phi) : \pi_V^{-1}(\pi^{-1}U \cap \phi^{-1}V) \to (\pi^{-1}U) \cap \phi^{-1}V) \times \mathbb{R}^m$$

where $\Phi = (\pi, \phi)$ is a bundle chart on E over U and $\mathtt{x}$ is a chart on V in F.

Exercise 19.2 *Prove: Let $f : N \to M$ be a smooth map and $\pi : E \to M$ a fiber bundle with typical fiber F. Then $\mathcal{V}f^*E \to N$ is bundle isomorphic to $pr_2^*\mathcal{V}E \to N$.* ***Hint:*** *It is enough to prove that $\mathcal{V}f^*E$ is isomorphic to $\mathcal{V}E$ along the map $pr_2 : f^*E \to E$. Also check that $(Tf)^*E = \{(u,v) \in TN \times TE | Tf \cdot u = T\pi \cdot v\}$.*

Suppose now that $\pi : E \to M$ is a vector bundle. The vertical vector bundle $\mathcal{V}E$ is isomorphic to the vector bundle π^*E over E (we say that $\mathcal{V}E$ is isomorphic to E along π). To see this note that if $(\zeta, \xi) \in \pi^*E$, then $\pi(\zeta + t\xi)$ is constant in t. From this we see that the map from π^*E to TE given by $(\zeta, \xi) \mapsto d/dt|_0(\zeta + t\xi)$ maps into $\mathcal{V}E$. It is easy to see that this map (real) vector bundle isomorphism. Let $\mathcal{V}E$ be the vertical bundle on a vector bundle E over M. Denote the isomorphism from π^*E to $\mathcal{V}E$ by j:

$$j : \pi^*E \to \mathcal{V}\mathcal{E} \quad (\zeta, \xi) \mapsto \frac{d}{dt}_0(\zeta + t\xi) =: \xi_\zeta$$

Denote the vector bundle isomorphism from $\mathcal{V}E$ to E along π by pr_2 :

$$pr_2 : \mathcal{V}E \to E \quad \xi_\zeta \mapsto \xi$$

The map $pr_2 : \mathcal{V}E \to E$ pick off the "second component" ξ of ξ_ζ, whereas the projection map $\pi_\mathcal{V} : \mathcal{V}E \to E$ yields the "first component" ζ. The reader should think of ξ_ζ and (ζ, ξ) as the "same object".

Definition 19.5 *A connection on a vector bundle $\pi : E \to M$ is a smooth distribution $\mathcal{H}$ on the total space E such that*

(i) $\mathcal{H}$ is complementary to the vertical bundle:

$$TE = \mathcal{H} \oplus \mathcal{V}E$$

(ii) $\mathcal{H}$ is homogeneous: $T_\xi\mu_r(\mathcal{H}_\xi) = \mathcal{H}_{r\xi}$ for all $\xi \in \mathcal{H}$, $r \in \mathbb{R}$ and where μ_r is the multiplication map $\mu_r : \xi \mapsto r\xi$. The subbundle $\mathcal{H}$ is called the horizontal subbundle (or horizontal distribution).

Any $\xi \in TE$ has a decomposition $\xi = \mathcal{H}v + \mathcal{V}v$. Here, $\mathcal{H} : v \mapsto \mathcal{H}v$ and $\mathcal{V} : v \mapsto \mathcal{V}v$ are the obvious projections. An individual element $v \in T_\xi E$ is **horizontal** if $v \in \mathcal{H}_\xi$ and **vertical** if $v \in \mathcal{V}_\xi E$. A vector field $\widetilde{X} \in \mathfrak{X}(E)$ is said to be a **horizontal vector field** (resp. **vertical vector field**) if $\widetilde{X}(\xi) \in \mathcal{H}_\xi$ (resp. $\widetilde{X}(\xi) \in \mathcal{V}_\xi E$) for all $\xi \in E$.

A map $f : N \to E$ is said to be **horizontal** or parallel if $Tf \cdot v$ is horizontal for all $v \in TN$. The most important case comes when we start with a section of E along a map $f : N \to M$. A horizontal lift of f is a section $\widetilde{f}$ along f such that $\widetilde{f}$ is horizontal; in other words, the following diagram commutes and $T\widetilde{f}$ has image in $\mathcal{H}$:

$$\begin{array}{ccc} & & E \\ & \widetilde{f} \nearrow & \downarrow \\ N & \to & M \end{array}$$

Now if we have a connection on $\pi : E \to M$ we have a so called connector (connection map) which is the map $\kappa : TE \to E$ defined by $\kappa := pr_2(\mathcal{V}v)$. The connector is a vector bundle homomorphism along the map $\pi : E \to M$:

$$\begin{array}{ccc} TE & \xrightarrow{\kappa} & E \\ \downarrow & & \downarrow \\ E & \xrightarrow{\pi} & M \end{array}$$

Now an interesting fact is that TE is also a vector bundle over TM (via the map $T\pi$) and κ gives a vector bundle homomorphism along $\pi_{TM} : TM \to M$. More precisely, we have the following

Theorem 19.2 *If κ is the connector for a connection on a vector bundle $\pi : E \to M$ then κ gives a vector bundle homomorphism from the bundle $T\pi : TE \to TM$ to the bundle $\pi : E \to M$.*

$$\begin{array}{ccc} TE & \xrightarrow{\kappa} & E \\ \downarrow & & \downarrow \\ TM & \to & M \end{array}$$

Now using this notion of connection with associated connector κ we can get a natural covariant derivative.

Definition 19.6 *If $\mathcal{H}$ is a connection on $E \to M$ with connector κ then we define the **covariant derivative** of a section $\sigma \in \Gamma_f(E)$ along a map $f : N \to M$ with respect to $v \in T_pN$ by*

$$\nabla_v^f \sigma := \kappa \left(T_p\sigma \cdot v \right).$$

If V is a vector field on N then $\left(\nabla_V \sigma \right)(p) := \nabla_{V(p)}\sigma$.

Now let $\epsilon_1, ..., \epsilon_n$ be a frame field defined over $U \subset M$. Since f is continuous $O = f^{-1}(U)$ is open and on O, the $\epsilon_1 \circ f, ..., \epsilon_n \circ f$ are fields along f and so locally (on O) we may write $\sigma = \sigma^i \left(\epsilon_i \circ f \right)$ for some functions $\sigma^a : U \subset N \to \mathbb{R}$.

For any $p \in O$ and $v \in T_pN$ we have

$$\nabla_v^f \sigma := \left(d\sigma^a \cdot v \right) \epsilon_a(f\left(p\right)) + \left. A_r^a \right|_p \left(Tf \cdot v \right) \sigma^r(p) \epsilon_a(f\left(p\right))$$

Exercise 19.3 *Prove this last statement.*

Exercise 19.4 *Prove that if $f = id_M$ then for $v \in TM$, $\nabla_v : \mathfrak{X}(M) \to \mathfrak{X}(M)$ is a Koszul connection and that we recover our previous results.*

If all we had were the connection forms we could use the above formula as the definition of $\nabla_v^f \sigma$. Since f might not be even be an immersion this definition would only make sense because of the fact that it is independent of the frame.

In the same way that one extends a derivation to a tensor derivation one may show that a covariant derivative on a vector bundle induces naturally related covariant derivatives on all the multilinear bundles. In particular, $\Pi : E^* \to M$

denotes the dual bundle to $\pi : E \to M$ we may define connections on $\Pi : E^* \to M$ and on $\pi \otimes \Pi : E \otimes E^* \to M$. We do this in such a way that for $s \in \Gamma(M, E)$ and $s^* \in \Gamma(M, E^*)$ we have

$$\nabla_X^{E \otimes E^*}(s \otimes s^*) = \nabla_X s \otimes s^* + s \otimes \nabla_X^{E^*} s^*$$

and

$$(\nabla_X^{E^*} s^*)(s) = X(s^*(s)) - s^*(\nabla_X s).$$

Of course this last formula follows from our insistence that covariant differentiation commutes with contraction:

$$\begin{aligned} X(s^*(s)) &= \\ (\nabla_X C(s \otimes e^*)) &= C(\nabla_X^{E \otimes E^*}(s \otimes s^*)) \\ &= C\left(\nabla_X s \otimes s^* + s \otimes \nabla_X^{E^*} s^*\right) \\ &= s^*(\nabla_X s) + (\nabla_X^{E^*} s^*)(s) \end{aligned}$$

where C denotes the contraction $s \otimes \alpha \mapsto \alpha(s)$. All this works like the tensor derivation extension procedure discussed previously.

Now the bundle $E \otimes E^* \to M$ is naturally isomorphic to $\mathrm{End}(E)$ and by this isomorphism we get a connection on $\mathrm{End}(E)$.

$$(\nabla_X A)(s) = \nabla_X(A(s)) - A(\nabla_X s)$$

Indeed, since $c : s \otimes A \mapsto A(s)$ is a contraction we must have

$$\begin{aligned} \nabla_X(A(s)) &= c\,(\nabla_X s \otimes A + s \otimes \nabla_X A) \\ &= A(\nabla_X s) + (\nabla_X A)(s) \end{aligned}$$

19.4 Parallel Transport

Once again let $E \to M$ be a vector bundle with a connection given by a system of connection forms $\{\overset{\alpha}{A}\}_\alpha$. Suppose we are given a curve $c : I \to M$ together with a section along c; that is a map $\sigma : I \to E$ such that the following diagram commutes:

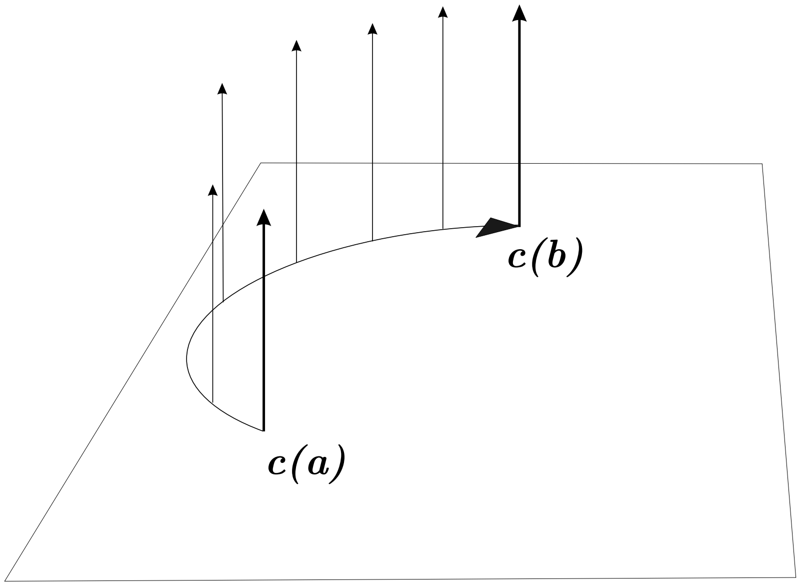
$$\begin{array}{ccc} & & E \\ & \overset{\sigma}{\nearrow} & \downarrow \\ I & \overset{c}{\to} & M \end{array}$$

Motivation: If c is an integral curve of a field X then we have

$$\begin{aligned} (\nabla_X s)(c(t)) &= (X(c(t)).s^a(c(t)) + (\,A_r^a|_{c(t)}\, X_{c(t)})s^r(c(t)))\epsilon_a(c(t)) \\ &= (\dot{c}(t).s^a(t))\epsilon_a(t) + (\,A_r^a|_{c(t)}\, \dot{c}(t))s^r(t)\epsilon_a(t) \end{aligned}$$

where we have abbreviated $s^a(t) := s^a(c(t))$ and $\epsilon_a(t) := \epsilon_a(c(t))$. As a special case of covariant differentiation along a map we have

$$\nabla_{\partial_t}\sigma := (\frac{d}{dt}\sigma^a(t))\epsilon_a(t) + (\,A^a_r|_{c(t)}\,\dot{c}(t))\sigma^r(t)\epsilon_a(t)$$



Simple schematic of parallel transport

Now on to the important notion of parallel transport.

Definition 19.7 *Let $c : [a, b] \to M$ be a smooth curve. A section σ along c is said to be* ***parallel along*** *c if*

$$\frac{\nabla\sigma}{dt}(t) := (\nabla_{\partial_t}\sigma)(t) = 0 \text{ for all } t \in [a, b].$$

Similarly, a section $\sigma \in \Gamma(M, E)$ is said to be ***parallel*** *if $\nabla_X\sigma = 0$ for all $X \in \mathfrak{X}(M)$.*

Exercise 19.5 *Show that if $t \mapsto X(t)$ is a curve in E_p then we can consider X a vector field along the constant map $p : t \mapsto p$ and then $\nabla_{\partial_t}X(t) = X'(t) \in E_p$.*

Exercise 19.6 *Show that $\sigma \in \Gamma(M, E)$ is a parallel section if and only if $X \circ c$ is parallel along c for every curve $c : I \to M$*

Exercise 19.7 *Show that for $f : I \to \mathbb{R}$ and $\sigma : I \to M$ is a section of E along c then $\nabla_{\partial_t}(f\sigma) = \frac{df}{dt}\sigma + f\nabla_{\partial_t}\sigma$.*

Exercise 19.8 *Show that if $\sigma : I \to U \subset M$ where U is the domain of a local frame field $\{\epsilon_1, ..., \epsilon_k\}$ then $\sigma(t) = \sum_{i=1}^{k}\sigma^i(t)\epsilon_i(c(t))$. Write out the local expression in this frame field for $\nabla_v\sigma$ where $v = v^j\epsilon_j$.*

Theorem 19.3 *Given a smooth curve $c : [a, b] \to M$ and numbers $t_0 \in [a, b]$ with $c(t_0) = p$ and vector $v \in E_p$ there is a unique parallel section σ_c along c such that $\sigma_c(t_0) = v$.*

Proof. In local coordinates this reduces to a first order initial value problem that may be shown to have a unique smooth solution. Thus if the image of the curve lies completely inside a coordinate chart then we have the result. The general result follows from patching these together. This is exactly what we do below when we generalize to piecewise smooth curves so we will leave this last part of the proof to the skeptical reader. ■

Under the conditions of this last theorem the value $\sigma_c(t)$ is a vector in the fiber $E_{c(t)}$ and is called the **parallel transport** or **parallel translation** of v along c from $c(t_0)$ to $c(t)$. Let us denote this by $P(c)^t_{t_0} v$. Next we suppose that $c : [a, b] \to M$ is a (continuous) piecewise smooth curve. Thus we may find a monotonic sequence $t_0, t_1, ...t_j = t$ such that $c_i := c|_{[t_{i-1}, t_i]}$ (or $c|_{[t_i, t_{i-1}]}$) is smooth[1]. In this case we define

$$P(c)^t_{t_0} := P(c)^t_{t_{j-1}} \circ \cdots \circ P(c)^{t_1}_{t_0}$$

Now given $v \in E_{c(t_0)}$ as before, the obvious sequence of initial value problems gives a unique piecewise smooth section σ_c along c such that $\sigma_c(t_0) = v$ and the solution must clearly be $P(c)^t_{t_0} v$ (why?).

Exercise 19.9 *$P(c)^t_{t_0} : E_{c(t_0)} \to E_{c(t)}$ is a linear isomorphism for all t with inverse $P(c)^{t_0}_t$ and the map $t \mapsto \sigma_c(t) = P(c)^t_{t_0} v$ is a section along c that is smooth wherever c is smooth.*

We may approach the covariant derivative from the direction of parallel transport. Indeed some authors given an axiomatic definition of parallel transport, prove its existence and then use it to define covariant derivative. For us it will suffice to have the following theorem:

Theorem 19.4 *For any smooth section σ of E defined along a smooth curve $c : I \to M$. Then we have*

$$(\nabla_{\partial_t} \sigma)(t) = \lim_{\varepsilon \to 0} \frac{P(c)^t_{t+\varepsilon} \sigma(t + \varepsilon) - \sigma(t)}{\varepsilon}$$

Proof. Let $e_1, ..., e_k$ be a basis of $E_{c(t_0)}$ for some fixed $t_0 \in I$. Let $\epsilon_i(t) := P(c)^t_{t_0} e_i$. Then $\nabla_{\partial_t} \epsilon_i(t) \equiv 0$ and $\sigma(t) = \sum \sigma^i(t) e_i(t)$. Then

$$\begin{aligned}
&\lim_{\varepsilon \to 0} \frac{P(c)^t_{t+\varepsilon} \sigma(t + \varepsilon) - \sigma(t)}{\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{\sigma^i(t + \varepsilon) P(c)^t_{t+\varepsilon} \epsilon_i(t + \varepsilon) - \sigma(t)}{\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{\sigma^i(t + \varepsilon) \epsilon_i(t) - \sigma^i(t) e_i(t)}{\varepsilon} \\
&= \sum \frac{d\sigma^i}{dt}(t) \epsilon_i(t).
\end{aligned}$$

[1] It may be that $t < t_0$.

On the other hand

$$\begin{aligned}(\nabla_{\partial_t}\sigma)(t) &= \nabla_{\partial_t}(\sigma^i(t)\epsilon_i(t)) \\ &= \sum \frac{d\sigma^i}{dt}(t)\epsilon_i(t) + \sum \sigma^i(t)\nabla_{\partial_t}\epsilon_i(t) \\ &= \sum \frac{d\sigma^i}{dt}(t)\epsilon_i(t).\end{aligned}$$

■

Exercise 19.10 *Let ∇ be a connection on $E \to M$ and let $\alpha : [a,b] \to M$ and $\beta : [a,b] \to E_{\alpha(t_0)}$. If $X(t) := P(\alpha)_{t_0}^t(\beta(t))$ then*

$$\left(\nabla_{\partial_t} X\right)(t) = P(\alpha)_{t_0}^t(\beta'(t))$$

Note: $\beta'(t) \in E_{\alpha(t_0)}$ for all t.

An important fact about covariant derivatives is that they don't need to commute. If $\sigma : M \to E$ is a section and $X \in \mathfrak{X}(M)$ then $\nabla_X\sigma$ is a section also and so we may take it's covariant derivative $\nabla_Y\nabla_X\sigma$ with respect to some $Y \in \mathfrak{X}(M)$. In general, $\nabla_Y\nabla_X\sigma \neq \nabla_X\nabla_Y\sigma$ and this fact has an underlying geometric interpretation which we will explore later. A measure of this lack of commutativity is the **curvature operator** that is defined for a pair $X, Y \in \mathfrak{X}(M)$ to be the map $F(X,Y) : \Gamma(E) \to \Gamma(E)$ defined by

$$\begin{aligned}F(X,Y)\sigma &:= \nabla_X\nabla_Y\sigma - \nabla_Y\nabla_X\sigma - \nabla_{[X,Y]}\sigma. \\ &\text{i.e.} \\ F(X,Y)\sigma &:= [\nabla_X, \nabla_Y]\sigma - \nabla_{[X,Y]}\sigma\end{aligned}$$

Theorem 19.5 *For fixed σ the map $(X,Y) \mapsto F(X,Y)\sigma$ is $C^\infty(M)$ bilinear and antisymmetric.*
$F(X,Y) : \Gamma(E) \to \Gamma(E)$ is a $C^\infty(M)$ module homomorphism; that is it is linear over the smooth functions:

$$F(X,Y)(f\sigma) = fF(X,Y)(\sigma)$$

Proof. We leave the proof of the first part as an easy exercise. For the second part we just calculate:

$$\begin{aligned}F(X,Y)(f\sigma) &= \nabla_X\nabla_Y f\sigma - \nabla_Y\nabla_X f\sigma - \nabla_{[X,Y]}f\sigma \\ &= \nabla_X(f\nabla_Y\sigma + (Yf)\sigma) - \nabla_Y(f\nabla_X\sigma + (Xf)\sigma) \\ &- f\nabla_{[X,Y]}\sigma - ([X,Y]f)\sigma \\ &= f\nabla_X\nabla_Y\sigma + (Xf)\nabla_Y\sigma + (Yf)\nabla_X\sigma + X(Yf) \\ &- f\nabla_Y\nabla_X\sigma - (Yf)\nabla_X\sigma - (Xf)\nabla_Y\sigma - Y(Xf) \\ &- f\nabla_{[X,Y]}\sigma - ([X,Y]f)\sigma \\ &= f[\nabla_X, \nabla_Y] - f\nabla_{[X,Y]}\sigma = fF(X,Y)\sigma\end{aligned}$$

■

Thus we also have F as a map $F : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \Gamma(M, \mathrm{End}(E))$. But then since R is tensorial in the first two slot and antisymmetric we also may think in the following terms

$$F \in \Gamma(M, \mathrm{End}(E) \otimes \wedge^2 M)$$
or
$$F \in \Gamma(M, E \otimes E^* \otimes \wedge^2 M).$$

In the current circumstance it is harmless to identify $E \otimes E^* \otimes \wedge^2 M$ with $\wedge^2 M \otimes E \otimes E^*$ the second one seems natural too although when translating into matrix notation the first is more consistent.

We will have many occasions to consider differentiation along maps and so we should also look at the curvature operator for sections along maps:

Let $f : N \to M$ be a smooth map and σ a section of $E \to M$ along f. For $U, V \in \mathfrak{X}(N)$ we define a map $F^f(U, V) : \Gamma_f(E) \to \Gamma_f(E)$ by

$$F^f(U, V)X := \nabla_U \nabla_V X - \nabla_V \nabla_U X - \nabla_{[U,V]} X$$

for all $X \in \Gamma_f(E)$.

Notice that if $X \in \Gamma(E)$ then $X \circ f \in \Gamma_f(E)$ and if $U \in \mathfrak{X}(N)$ then $Tf \circ U \in \Gamma_f(TM) := \mathfrak{X}_f(M)$ is a vector field along f. Now as a matter of notation we let $F(Tf \circ U, Tf \circ V)X$ denote the map $p \mapsto F(Tf \cdot U_p, Tf \cdot V_p)X_p$ which makes sense because the curvature is tensorial. Thus $F(Tf \circ U, Tf \circ V)\ X \in \Gamma_f(E)$. Then we have the following useful fact:

Proposition 19.3 *Let $X \in \Gamma_f(E)$ and $U, V \in \mathfrak{X}(N)$. Then*

$$F^f(U, V)X = F(Tf \circ U, Tf \circ V)X.$$

Proof. Exercise! ■

19.5 The case of the tangent bundle: Linear connections

A connection on the tangent bundle TM of a smooth manifold M is called a linear connection (or affine connection; especially in older literature) on M. A natural linear connection can be given in terms of connection forms but in this case one often sees the concept of Christoffel symbols. This is just another way of presenting the connection forms and so the following is a bit redundant.

Definition 19.8 *A* ***system of Christoffel symbols*** *on a smooth manifold M (modelled on M) is an assignment of a differentiable map*

$$\Gamma_\alpha : \psi_\alpha(U_\alpha) \to L(\mathsf{M}, \mathsf{M}; \mathsf{M})$$

to every admissible chart U_α, ψ_α such that if U_α, ψ_α and U_β, ψ_β are two such charts with $U_\alpha \cap U_\beta \neq \emptyset$, then for all $p \in U_\alpha \cap U_\beta$

$$\begin{aligned}&D(\psi_\beta \circ \psi_\alpha^{-1}) \cdot \Gamma_\alpha(x) \\ &= D^2(\psi_\beta \circ \psi_\alpha^{-1}) + \Gamma_\beta(y) \circ (D(\psi_\beta \circ \psi_\alpha^{-1}) \times D(\psi_\beta \circ \psi_\alpha^{-1}))\end{aligned}$$

where $y = \psi_\beta(p)$ and $x = \psi_\alpha(p)$. For finite dimensional manifolds with $\psi_\alpha = (x^1, ..., x^n)$ and $\psi_\beta = (y^1, ..., y^n)$ this last condition reads

$$\frac{\partial y^r}{\partial x^k}(x)\Gamma^k_{ij}(x) = \frac{\partial^2 y^r}{\partial x^i \partial x^j}(x) + \bar{\Gamma}^r_{pq}(y(x))\frac{\partial y^p}{\partial x^i}(x)\frac{\partial y^q}{\partial x^j}(x)$$

where $\Gamma^k_{ij}(x)$ are the components of $\Gamma_\alpha(x)$ and $\bar{\Gamma}^r_{pq}$ the components of $\Gamma_\beta(y(x)) = \Gamma_\beta(\psi_\beta \circ \psi_\alpha^{-1}(x))$ with respect to the standard basis of $L(\mathbb{R}^n, \mathbb{R}^n; \mathbb{R}^n)$.

Proposition 19.4 *Given a system of Christoffel symbols on a smooth manifold M there is a unique natural covariant derivative (a linear connection) ∇ on M such that the local expression $(\nabla_X Y)_U$ of $\nabla_X Y$ with respect to a chart $U, \mathtt{x}$ is given by $DY_U(x) \cdot X_U(x) + \Gamma_U(x)(X_U(x), Y_U(x))$ for $x \in \mathtt{x}(U)$. Conversely, a natural covariant derivative determines a system of Christoffel symbols.*

Proof. The proof of this is not significantly different from the proof of 19.1.Let a system of Christoffel symbols be given. Now for any open set $U \subset M$ we may let $\{U_a, \mathtt{x}_a\}_a$ be any family of charts such that $\bigcup_a U_a = U$. Given vector fields $X, Y \in \mathfrak{X}(U)$ we define

$$s_{X,Y}(U_a) := \nabla^{U_a}_{r^U_{U_a} X} r^U_{U_a} Y$$

to have principal representation

$$(\nabla_X Y)_{U_\alpha} = DY_{U_\alpha} \cdot X_{U_\alpha} \cdot + \Gamma_{U_\alpha}(X_{U_\alpha}, Y_{U_\alpha}).$$

One should check that the transformation law in the definition of a system of Christoffel symbols implies that $DY_{U_\alpha} \cdot X_{U_\alpha} \cdot + \Gamma_{U_\alpha}(X_{U_\alpha}, Y_{U_\alpha})$ transforms as the principal local representative of a vector. It is straightforward to check that the change of chart formula for Christoffel symbols implies that

$$r^{U_a}_{U_a \cap U_b} s_{X,Y}(U_a) = s_{X,Y}(U_a \cap U_b) = r^{U_b}_{U_a \cap U_b} s_{X,Y}(U_a)$$

and so by there is a unique section

$$\nabla_X Y \in \mathfrak{X}(U)$$

such that

$$r^U_{U_a} \nabla_X Y = s_{X,Y}(U_a)$$

The verification that this defines a natural covariant derivative is now a straightforward (but tedious) verification of (1)-(5) in the definition of a natural covariant derivative.

For the converse, suppose that ∇ is a natural covariant derivative on M. Define the Christoffel symbol for a chart U_a, ψ_α to be in the following way. For fields X and Y one may define $\Theta(X_{U_\alpha}, Y_{U_\alpha}) := (\nabla_X Y)_{U_\alpha} - DY_{U_\alpha} \cdot X_{U_\alpha}$ and then use the properties (1)-(5) to show that $\Theta(X_{U_\alpha}, Y_{U_\alpha})(x)$ depends only on the values of X_{U_α} and Y_{U_α} at the point x. Thus there is a function $\Gamma : U_\alpha \to L(\mathsf{M}, \mathsf{M}; \mathsf{M})$ such that $\Theta(X_{U_\alpha}, Y_{U_\alpha})(x) = \Gamma(x)(X_{U_\alpha}(x), Y_{U_\alpha}(x))$. We wish to show that this defines a system of Christoffel symbols. But this is just an application of the chain rule. ∎

In finite dimensions and using traditional notation

$$\nabla_X Y = \left(\frac{\partial Y^k}{\partial x^j} X^j + \Gamma^k_{ij} X^i Y^j \right) \frac{\partial}{\partial x^k}.$$

where $X = X^j \frac{\partial}{\partial x^j}$ and $Y = Y^j \frac{\partial}{\partial x^j}$. In particular,

$$\nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} = \Gamma^k_{ij} \frac{\partial}{\partial x^k}.$$

Let ∇ be a natural covariant derivative on $E \to M$. It is a consequence of proposition 8.4 that for each $X \in \mathfrak{X}(U)$ there is a unique tensor derivation ∇_X on $\mathfrak{T}^r_s(U)$ such that ∇_X commutes with contraction and coincides with the given covariant derivative on $\mathfrak{X}(U)$ (also denoted ∇_X) and with $\mathcal{L}_X f$ on $C^\infty(U)$.

To describe the covariant derivative on tensors more explicitly consider $\Upsilon \in \mathfrak{T}^0_1$ with a 1-form Since we have the contraction $Y \otimes \omega \mapsto C(Y \otimes \omega) = \omega(Y)$ we should have

$$\begin{aligned}
\nabla_X \Upsilon(Y) &= \nabla_X C(Y \otimes \omega) \\
&= C(\nabla_X (Y \otimes \omega)) \\
&= C(\nabla_X Y \otimes \omega + Y \otimes \nabla_X \omega) \\
&= \omega(\nabla_X Y) + (\nabla_X \omega)(Y)
\end{aligned}$$

and so we should define $(\nabla_X \omega)(Y) := \nabla_X(\omega(Y)) - \omega(\nabla_X Y)$. More generally, if $\Upsilon \in \mathfrak{T}^r_s$ then

$$(\nabla_X \Upsilon)(Y_1, ..., Y_s) = \nabla_X(\Upsilon(\omega_1, ..., \omega_r, Y_1, ..., Y_s)) - \sum_{j=1}^{r} \Upsilon(\omega_1, .\nabla_X \omega_j, .., \omega_r, Y_1, ...) \tag{19.1}$$

$$- \sum_{i=1}^{s} \Upsilon(\omega_1, ..., \omega_r, Y_1, ..., \nabla_X Y_i, ...) \tag{19.2}$$

Definition 19.9 *The covariant differential of a tensor field $\Upsilon \in \mathfrak{T}^k_l$ is denoted by $\nabla\Upsilon$ is defined to be the element of $\mathfrak{T}^k_{l+1}$ such that if $\Upsilon \in \mathfrak{T}^k_l$ then*

$$\nabla\Upsilon(\omega^1, ..., \omega^1, X, Y_1, ..., Y_s) := \nabla_X \Upsilon(\omega^1, ..., \omega^1, Y_1, ..., Y_s)$$

For any fixed frame field $E_1, ..., E_n$ we denote the components of $\nabla\Upsilon$ by $\nabla_i \Upsilon^{i_1 ... i_r}_{j_1 ... j_s}$.

Remark 19.1 *We have place the new variable at the beginning as suggested by our notation $\nabla_i \Upsilon^{i_1}_{j_1 \dots j_s}$ for the for the components of $\nabla \Upsilon$ but in opposition to the equally common $\Upsilon^{i_1}_{j_1 \dots j_s; i}$. This has the advantage of meshing well with exterior differentiation , making the statement of theorem 19.6 as simple as possible.*

The reader is asked in problem 1 to show that $T(X,Y) := \nabla_X Y - \nabla_Y X$ defines a tensor and so $T(X_p, Y_p)$ is well defined for $X_p, Y_p \in T_pM$.

Definition 19.10 *Let ∇ be a connection on M and let $T(X,Y) := \nabla_X Y - \nabla_Y X$. Then T is called the **torsion tensor** of ∇. If T is identically zero then we say that the connection ∇ is **torsion free**.*

19.5.1 Comparing the differential operators

On a smooth manifold we have the Lie derivative $\mathcal{L}_X : \mathfrak{T}^r_s(M) \to \mathfrak{T}^r_s(M)$ and the exterior derivative $d : \Omega^k(M) \to \Omega^{k+1}(M)$ and in case we have a torsion free covariant derivative ∇ then that makes three differential operators which we would like to compare. To this end we restrict attention to purely covariant tensor fields $\mathfrak{T}^0_s(M)$.

The extended map $\nabla_\xi : \mathfrak{T}^0_s(M) \to \mathfrak{T}^0_s(M)$ respects the subspace consisting of alternating tensors and so we have a map

$$\nabla_\xi : L^k_{alt}(M) \to L^k_{alt}(M)$$

which combine to give a degree preserving map

$$\nabla_\xi : L_{alt}(M) \to L_{alt}(M)$$

or in other notation

$$\nabla_\xi : \Omega(M) \to \Omega(M).$$

It is also easily seen that not only do we have $\nabla_\xi(\alpha \otimes \beta) = \nabla_\xi \alpha \otimes \beta + \alpha \otimes \nabla_\xi \beta$ but also

$$\nabla_\xi(\alpha \wedge \beta) = \nabla_\xi \alpha \wedge \beta + \alpha \wedge \nabla_\xi \beta$$

Now as soon as one realizes that $\nabla \omega \in \Omega^k(M)_{C^\infty} \otimes \Omega^1(M)$ instead of $\Omega^{k+1}(M)$ we search for a way to fix things. By antisymmetrizing we get a map $\Omega^k(M) \to \Omega^{k+1}(M)$ which turns out to be none other than our old friend the exterior derivative as will be shown below.

On a smooth manifold we have the Lie derivative $\mathcal{L}_X : \mathfrak{T}^r_s(M) \to \mathfrak{T}^r_s(M)$ and the exterior derivative $d : \Omega^k(M) \to \Omega^{k+1}(M)$ and in case we have a torsion free covariant derivative ∇ then that makes three differential operators which we would like to compare. To this end we restrict attention to purely covariant tensor fields $\mathfrak{T}^0_s(M)$.

The extended map $\nabla_\xi : \mathfrak{T}^0_s(M) \to \mathfrak{T}^0_s(M)$ respects the subspace consisting of alternating tensors and so we have a map

$$\nabla_\xi : L^k_{alt}(M) \to L^k_{alt}(M)$$

which combine to give a degree preserving map

$$\nabla_\xi : L_{alt}(M) \to L_{alt}(M)$$

or in other notation

$$\nabla_\xi : \Omega(M) \to \Omega(M).$$

It is also easily seen that not only do we have $\nabla_\xi(\alpha \otimes \beta) = \nabla_\xi \alpha \otimes \beta + \alpha \otimes \nabla_\xi \beta$ but also

$$\nabla_\xi(\alpha \wedge \beta) = \nabla_\xi \alpha \wedge \beta + \alpha \wedge \nabla_\xi \beta$$

Now as soon as one realizes that $\nabla\omega \in \Omega^k(M)_{C^\infty} \otimes \Omega^1(M)$ instead of $\Omega^{k+1}(M)$ we search for a way to fix things. By antisymmetrizing we get a map $\Omega^k(M) \to \Omega^{k+1}(M)$ which turns out to be none other than our old friend the exterior derivative as will be shown below.

Now recall 19.1. There is a similar formula for the Lie derivative:

$$(\mathcal{L}_X S)(Y_1, ..., Y_s) = X(S(Y_1, ..., Y_s)) - \sum_{i=1}^{s} S(Y_1, ..., Y_{i-1}, \mathcal{L}_X Y_i, Y_{i+1}, ..., Y_s). \tag{19.3}$$

On the other hand, if ∇ is torsion free so that $\mathcal{L}_X Y_i = [X, Y_i] = \nabla_X Y_i - \nabla_{Y_i} X$ then we obtain the

Proposition 19.5 *For a torsion free connection we have*

$$(\mathcal{L}_X S)(Y_1, ..., Y_s) = (\nabla_X S)(Y_1, ..., Y_s) + \sum_{i=1}^{s} S(Y_1, ..., Y_{i-1}, \nabla_{Y_i} X, Y_{i+1}, ..., Y_s). \tag{19.4}$$

Theorem 19.6 *If ∇ is a torsion free covariant derivative on M then*

$$d = k! Alt \circ \nabla$$

or in other words, if $\omega \in \Omega^k(M)$ then

$$d\omega(X_0, X_1, ..., X_k) = \sum_{i=0}^{k} (-1)^i (\nabla_{X_i} \omega)(X_0, ..., \widehat{X_i}, ..., X_k)$$

Proof. Note that $Alt \circ \nabla\omega(X_0, X_1, ..., X_k) = \frac{1}{k!} \sum_{i=0}^{k} (-1)^i (\nabla_{X_i} \omega)(X_0, ..., \widehat{X_i}, ..., X_k)$.

The proof is now just a computation

$$\begin{aligned}
d\omega(X_0, X_1, ..., X_k) &= \sum_{i=0}^{k}(-1)^i X_i(\omega(X_0, ..., \widehat{X_i}, ..., X_k)) \\
&+ \sum_{1\le r<s\le k} (-1)^{r+s}\omega([X_r, X_s], X_0, ..., \widehat{X_r}, ..., \widehat{X_s}, ..., X_k) \\
&= \sum_{i=0}^{k}(-1)^i X_i(\omega(X_0, ..., \widehat{X_i}, ..., X_k)) \\
&+ \sum_{1\le r<s\le k} (-1)^{r+s}\omega(\nabla_{X_r}X_s - \nabla_{X_r}X_s, X_0, ..., \widehat{X_r}, ..., \widehat{X_s}, ..., X_k) \\
&\sum_{i=0}^{k}(-1)^i X_i(\omega(X_0, ..., \widehat{X_i}, ..., X_k)) \\
&+ \sum_{1\le r<s\le k} (-1)^{r+1}\omega(X_0, ..., \widehat{X_r}, ..., \nabla_{X_r}X_s, ..., X_k) \\
&- \sum_{1\le r<s\le k} (-1)^{s}\omega(X_0, ..., \nabla_{X_s}X_r, ..., \widehat{X_s}, ..., X_k) \\
&= \sum_{i=0}^{k}(-1)^i \nabla_{X_i}\omega(X_0, ..., \widehat{X_i}, ..., X_k) \text{ (by using 19.1)}
\end{aligned}$$

■

19.5.2 Higher covariant derivatives

Now let us suppose that we have a connection ∇^{E_i} on every vector bundle $E_i \to M$ in some family $\{E_i\}_{i\in I}$. We then also have the connections $\nabla^{E_i^*}$ induced on the duals $E_i^* \to M$. By demanding a product formula be satisfied as usual we can form a related family of connections on all bundles formed from tensor products of the bundles in the family $\{E_i, E_i^*\}_{i\in I}$. In this situation it might be convenient to denote any and all of these connections by a single symbol as long as the context makes confusion unlikely. In particular we have the following common situation: By the definition of a connection we have that $X \mapsto \nabla_X^E \sigma$ is $C^\infty(M)$ linear and so $\nabla^E\sigma$ is an section of the bundle $T^*M \otimes E$. We can use the Levi-Civita connection ∇ on M together with ∇^E to define a connection on $E \otimes T^*M$. To get a clear picture of this connection we first notice that a section ξ of the bundle $E \otimes T^*M$ can be written locally in terms of a local frame field $\{\theta^i\}$ on T^*M and a local frame field $\{e_i\}$ on E. Namely, we may write $\xi = \xi^i_j e_i \otimes \theta^j$. Then the connection $\nabla^{E\otimes T^*M}$ on $E \otimes T^*M$ is defined so that a product rule holds:

$$\begin{aligned}
\nabla_X^{E\otimes T^*M}\xi &= \nabla_X^E\left(\xi_j^i e_i\right)\otimes\theta^j + \xi_j^i e_i\otimes\nabla_X\theta^j \\
&= \left(X\xi_j^i e_i + \xi_j^i\nabla_X^E e_i\right)\otimes\theta^j + \xi_j^i e_i\otimes\nabla_X\theta^j \\
&= \left(X\xi_l^r e_r + e_r A_i^r(X)\xi_l^i\right)\otimes\theta^l - \omega_l^j(X)\theta^l\otimes\xi_j^r e_r \\
&= \left(X\xi_l^r + A_i^r(X)\xi_l^i - \omega_l^j(X)\xi_j^r\right)e_r\otimes\theta^l
\end{aligned}$$

Now let $\xi = \nabla^E\sigma$ for a given $\sigma\in\Gamma(E)$. The map $X\mapsto\nabla_X^{E\otimes T^*M}(\nabla^E\sigma)$ is $C^\infty(M)$ linear and $\nabla^{E\otimes T^*M}(\nabla^E\sigma)$ is an element of $\Gamma\left(E\otimes T^*M\otimes T^*M\right)$ which can again be given the obvious connection. The process continues and denoting all the connections by the same symbol we may consider the $k-$th covariant derivative $\nabla^k\sigma\in\Gamma(E\otimes T^*M^{\otimes k})$ for each $\sigma\in\Gamma(E)$.

It is sometimes convenient to change viewpoints just slightly and define the **covariant derivative operators** $\nabla_{X_1,X_2,\ldots,X_k}:\Gamma(E)\to\Gamma(E)$. The definition is given inductively as

$$\begin{aligned}
\nabla_{X_1,X_2}\sigma &:= \nabla_{X_1}\nabla_{X_2}\sigma - \nabla_{\nabla_{X_1}X_2}\sigma \\
\nabla_{X_1,\ldots,X_k}\sigma &:= \nabla_{X_1}\left(\nabla_{X_2,X_3,\ldots,X_k}\sigma\right) \\
&\quad - \nabla_{\nabla_{X_1}X_2,X_3,\ldots,X_k} - \ldots - \nabla_{X_2,X_3,\ldots,\nabla_{X_1}X_k}
\end{aligned}$$

Then we have following convenient formula which is true by definition:

$$\nabla^{(k)}\sigma(X_1,\ldots,X_k) = \nabla_{X_1,\ldots,X_k}\sigma$$

Warning. $\nabla_{\partial_\iota}\nabla_{\partial_j}\tau$ is a section of E but not the same section as $\nabla_{\partial_\iota\partial_j}\tau$ since in general

$$\nabla_{\partial_\iota}\nabla_{\partial_j}\tau \neq \nabla_{\partial_i}\nabla_{\partial_j}\sigma - \nabla_{\nabla_{\partial_i}\partial_j}\sigma$$

Now we have that

$$\begin{aligned}
\nabla_{X_1,X_2}\sigma - \nabla_{X_2,X_1}\sigma &= \nabla_{X_1}\nabla_{X_2}\sigma - \nabla_{\nabla_{X_1}X_2}\sigma - \left(\nabla_{X_2}\nabla_{X_1}\sigma - \nabla_{\nabla_{X_2}X_1}\sigma\right) \\
&= \nabla_{X_1}\nabla_{X_2}\sigma - \nabla_{X_2}\nabla_{X_1}\sigma - \nabla_{(\nabla_{X_1}X_2-\nabla_{X_2}X_1)}\sigma \\
&= F(X_1,X_2)\sigma - \nabla_{T(X_1,X_2)}\sigma
\end{aligned}$$

so if ∇ (the connection on he base M) is torsion free then we recover the curvature

$$\nabla_{X_1,X_2}\sigma - \nabla_{X_2,X_1}\sigma = F(X_1,X_2)\sigma$$

One thing that is quite important to realize is that F depends only on the connection on E while the operators ∇_{X_1,X_2} involve a torsion free connection on the tangent bundle TM.

Clearly there must be some kind of cancellation in the above formula for curvature. This leads naturally to the notion of "exterior covariant derivative".

19.5.3 Exterior covariant derivative

The exterior covariant derivative essentially antisymmetrizes the higher covariant derivatives just defined in such a way that the dependence on the auxiliary torsion free linear connection on the base cancel out. Of course this means that there must be a definition that does not involve this connection on the base at all. We give the definitions below but first we point out a little more algebraic structure. We can give the space of $E-$valued forms $\Omega(M,E)$ the structure of a $(\Omega(M),\Omega(M))-$ "bi-module" over $C^{\infty}(M)$. This means that we define a product $\wedge : \Omega(M) \times \Omega(M,E) \rightarrow \Omega(M,E)$ and another product $\wedge : \Omega(M,E) \times \Omega(M) \rightarrow \Omega(M,E)$ which are compatible in the sense that

$$(\alpha \wedge \omega) \wedge \beta = \alpha \wedge (\omega \wedge \beta)$$

for $\omega \in \Omega(M,E)$ and $\alpha \in \Omega(M)$. These products are defined by extending linearly the rules

$$\begin{aligned}
\alpha \wedge (\sigma \otimes \omega) &:= \sigma \otimes \alpha \wedge \omega \text{ for } \sigma \in \Gamma(E), \alpha, \omega \in \Omega(M) \\
(\sigma \otimes \omega) \wedge \alpha &:= \sigma \otimes \omega \wedge \alpha \text{ for } \sigma \in \Gamma(E), \alpha, \omega \in \Omega(M) \\
\alpha \wedge \sigma = \sigma \wedge \alpha &= \sigma \otimes \alpha \text{ for } \alpha \in \Omega(M) \text{ and } \sigma \in \Gamma(E)
\end{aligned}$$

for. It follows that

$$\alpha \wedge \omega = (-1)^{kl} \omega \wedge \alpha \text{ for } \alpha \in \Omega^k(M) \text{ and } \omega \in \Omega^l(M,E)$$

In the situation where $\sigma \in \Gamma(E) := \Omega^0(M,E)$ and $\omega \in \Omega(M)$ all three of the following conventional equalities:

$$\sigma\omega = \sigma \wedge \omega = \sigma \otimes \omega \text{ (special case)}$$

The proof of the following theorem is analogous the proof of the existence of the exterior derivative

Theorem 19.7 *Given a connection ∇^E on a vector bundle $\pi : E \rightarrow M$, there exists a unique operator $d^E : \Omega(M,E) \rightarrow \Omega(M,E)$ such that*

(i) $d^E(\Omega^k(M,E)) \subset \Omega^{k+1}(M,E)$
(ii) For $\alpha \in \Omega^k(M)$ and $\omega \in \Omega^l(M,E)$ we have

$$\begin{aligned}
d^E(\alpha \wedge \omega) &= d\alpha \wedge \omega + (-1)^k \alpha \wedge d^E\omega. \\
d^E(\omega \wedge \alpha) &= d^E\omega \wedge \alpha + (-1)^l \omega \wedge d\alpha
\end{aligned}$$

(iii) $d^E\alpha = d\alpha$ *for* $\alpha \in \Omega(M)$

(iv) $d^E\sigma = \nabla^E\sigma$ *for* $\sigma \in \Gamma(E)$

In particular, if $\alpha \in \Omega^k(M)$ and $\sigma \in \Omega^0(M,E) = \Gamma(E)$ we have

$$d^E(\sigma \otimes \alpha) = d^E\sigma \wedge \alpha + \sigma \otimes d\alpha.$$

It can be shown that we have the following formula:

$$\begin{aligned} d\omega(X_0,...,X_k) = &\sum_{0\leq i\leq k} (-1)^i \nabla^E_{X_i}(\omega(X_0,...,\widehat{X_i},...,X_k)) \\ &+ \sum_{0\leq i<j\leq k} (-1)^{i+j}\omega([X_i,X_j],X_0,...,\widehat{X_i},...,\widehat{X_j},...,X_k). \end{aligned}$$

Definition 19.11 *The operator whose existence is given by the pervious theorem is called the* ***exterior covariant derivative****.*

Remark 19.2 *In the literature it seems that it isn't always appreciated that there is a difference between $\left(d^E\right)^k$ and $\left(\nabla^E\right)^k$. The higher derivative given by $\left(d^E\right)^k$ are not appropriate for defining $k-th$ order Sobolev spaces since as we shall see $\left(d^E\right)^2$ is zero for any flat connection and $\left(d^E\right)^3 = 0$ is always true (Bianchi identity).*

If ∇^{TM} is a torsion free covariant derivative on M and ∇^E is a connection on the vector bundle $E \to M$ then as before we get a covariant derivative ∇ on all the bundles $E \otimes \wedge^k T^*M$ and as for the ordinary exterior derivative we have the formula

$$d^E = k!Alt \circ \nabla$$

or in other words, if $\omega \in \Omega^k(M,E)$ then

$$d\omega(X_0,X_1,...,X_k) = \sum_{i=0}^{k}(-1)^i(\nabla_{X_i}\omega)(X_0,...,\widehat{X_i},...,X_k)$$

Lets look at the local expressions with respect to a moving frame. In doing so we will rediscover and generalize the local expression obtained above. Let $\{e_i\}$ be a frame field for E on $U \subset M$. Then locally, for $\omega \in \Omega^k(M,E)$ we may write

$$\omega = e_i \otimes \omega^i$$

where $\omega_i \in \Omega^k(M,E)$ and $\sigma_i \in \Omega^0(M,E) = \Gamma(E)$. We have

$$\begin{aligned} d^E\omega &= d^E\left(e_i \otimes \omega^i\right) = e_i \otimes d\omega^i + d^E e_i \wedge \omega^i \\ &= e_j \otimes d\omega^j + A^j_i e_j \wedge \omega^i \\ &= e_j \otimes d\omega^j + e_j \otimes A^j_i \wedge \omega^i \\ &= e_j \otimes \left(d\omega^j + A^j_i \wedge \omega^i\right). \end{aligned}$$

Thus the "$k+1-$form coefficients" of $d^E\omega$ with respect to the frame $\{\sigma_j\}$ are given by $d\omega^j + A_i^j \wedge \omega^i$. Notice that if $k=0$ and we write $s = e_i \otimes s^i$ then we get

$$\begin{aligned} d^E s &= d^E\left(e_i \otimes s^i\right) \\ &= e_j \otimes \left(ds^j + A_i^j s^i\right) \end{aligned}$$

and if we apply this to X we get

$$\begin{aligned} d^E s = \nabla_X^E s &= e_j\left(Xs^j + A_i^j(X)s^i\right) \\ \left(\nabla_X^E s\right)^j &= Xs^j + A_i^j(X)s^i \end{aligned}$$

which is our earlier formula. As before, it is quite convenient to employ matrix notation. If $e_U = (e_1, ..., e_r)$ is the frame over U written as a row vector and $\omega_U = (\omega^1, ..., \omega^r)^t$ is the column vector of $k-$form coefficients then $d^E e_U = e_U A_U$ and the above calculation looks quite nice:

$$\begin{aligned} d^E\left(e_U\omega_U\right) &= e_U d\omega_U + d^E e_U \wedge \omega_U \\ &= e_U\left(d\omega_U + A_U \wedge \omega_U\right) \end{aligned}$$

Now there is another module structure we should bring out. We identify $\mathrm{End}(E) \otimes \wedge^k TM$ with $E \otimes E^* \otimes \wedge^k T^*M$. The space $\Omega(M, \mathrm{End}(E))$ is an algebra over $C^\infty(M)$ where the multiplication is according to $(L_1 \otimes \omega_1) \wedge (L_2 \otimes \omega_2) = (L_1 \circ L_2) \otimes \omega_1 \wedge \omega_2$. Now $\Omega(M, E)$ is a module over this algebra because we can multiply as follows

$$(L \otimes \alpha) \wedge (\sigma \otimes \beta) = L\sigma \otimes (\alpha \wedge \beta)$$

To understand what is going on a little better lets consider how the action of $\Omega(M, \mathrm{End}(E))$ on $\Omega(M, E)$ comes about from a slightly different point of view. We can identify $\Omega(M, \mathrm{End}(E))$ with $\Omega(M, E \otimes E^*)$ and then the action of $\Omega(M, E \otimes E^*)$ on $\Omega(M, E)$ is given by tensoring and then contracting: For $\alpha, \beta \in \Omega(M)$, $s, \sigma \in \Gamma(E)$ and $s^* \in \Gamma(E)$ we have

$$\begin{aligned} (s \otimes s^* \otimes \alpha) \wedge (\sigma \otimes \beta) &:= C(s \otimes s^* \otimes \sigma \otimes (\alpha \wedge \beta)) \\ &= s^*(\sigma)s \otimes (\alpha \wedge \beta) \end{aligned}$$

Now from ∇^E get related connections on E^*, $E \otimes E^*$and $E \otimes E^* \otimes E$. The connection on $E \otimes E^*$ is also a connection on $\mathrm{End}(E)$. Of course we have

$$\begin{aligned} &\nabla_X^{\mathrm{End}(E)\otimes E}(L \otimes \sigma) \\ &= (\nabla_X^{\mathrm{End}(E)} L) \otimes \sigma + L \otimes \nabla_X^E \sigma \end{aligned}$$

and after contraction

$$\begin{aligned} &\nabla_X^E(L(\sigma)) \\ &= (\nabla_X^{\mathrm{End}(E)} L)(\sigma) + L(\nabla_X^E \sigma) \end{aligned}$$

The connections on $\mathrm{End}(E)$ and $\mathrm{End}(E) \otimes E$ gives corresponding exterior covariant derivative operators

$$d^{\mathrm{End}(E)} : \Omega^k(M, \mathrm{End}(E)) \to \Omega^{k+1}(M, \mathrm{End}(E))$$

and

$$d^{\mathrm{End}(E)\otimes E} : \Omega^k(M, \mathrm{End}(E) \otimes E) \to \Omega^{k+1}(M, \mathrm{End}(E) \otimes E)$$

Let C denote the contraction $A \otimes \sigma \mapsto A(\sigma)$. For $\alpha \in \Omega^k(M), \beta \in \Omega(M)$, $s \in \Gamma(E)$ and $s^* \in \Gamma(E)$ we have $\nabla^E_X(L(\sigma)) = \nabla^E_X C(L \otimes \sigma) = C(\nabla^{\mathrm{End}(E)}_X L \otimes \sigma + L \otimes \nabla^E_X \sigma) = (\nabla^{\mathrm{End}(E)}_X L)(\sigma) + L(\nabla^E_X \sigma)$. But by definition $X \mapsto L(\nabla^E_X \sigma)$ is just $L \wedge \nabla^E \sigma$. Using this we have

$$\begin{aligned}
d^E((A \otimes \alpha) \wedge (\sigma \otimes \beta)) &:= d^E(A(\sigma) \otimes (\alpha \wedge \beta)) \\
&= \nabla^E(A(\sigma)) \wedge (\alpha \wedge \beta) \\
&+ A(\sigma) \otimes d\,(\alpha \wedge \beta) \\
&= (-1)^k \left(\alpha \wedge \nabla^E(A(\sigma)) \wedge \beta\right) + A(\sigma) \otimes d\,(\alpha \wedge \beta) \\
&= (-1)^k \left(\alpha \wedge \{\nabla^{\mathrm{End}(E)} A \wedge \sigma + A \wedge \nabla^E \sigma\} \wedge \beta\right) \\
&+ A(\sigma) \otimes d\,(\alpha \wedge \beta) + A(\sigma) \otimes d\alpha \wedge \beta + (-1)^k A(\sigma) \otimes \alpha \wedge d\beta \\
&= (\nabla^{\mathrm{End}(E)} A) \wedge \alpha \wedge \sigma \wedge \beta + (-1)^k A \wedge \alpha \wedge (\nabla^E \sigma) \wedge \beta \\
&+ A(\sigma) \otimes d\alpha \wedge \beta + (-1)^k A(\sigma) \otimes \alpha \wedge d\beta \\
&= (\nabla^{\mathrm{End}(E)} A \wedge \alpha + A \otimes d\alpha) \wedge (\sigma \otimes \beta) \\
&+ (-1)^k\,(A \otimes \alpha) \wedge (\nabla^E \sigma \wedge \beta + \sigma \otimes d\beta) \\
&= (\nabla^{\mathrm{End}(E)} A \wedge \alpha + A \otimes d\alpha) \wedge (\sigma \otimes \beta) \\
&+ (-1)^k\,(A \otimes \alpha) \wedge (\nabla^E \sigma \wedge \beta + \sigma \otimes d\beta) \\
&= d^{\mathrm{End}(E)}(A \otimes \alpha) \wedge (\sigma \otimes \beta) + (-1)^k\,(A \otimes \alpha) \wedge d^E\,(\sigma \otimes \beta)
\end{aligned}$$

By linearity we conclude that for $\Phi \in \Omega^k(M, \mathrm{End}(E))$ and $\omega \in \Omega(M, E)$ we have

$$d^E(\Phi \wedge \omega) = d^{\mathrm{End}(E)} \Phi \wedge \omega + (-1)^k \Phi \wedge d^E \omega$$

Proposition 19.6 *The map $\Omega^k(M, E) \to \Omega^{k+2}(M, E)$ given by $d^E \circ d^E$ given the action of the curvature 2-form of ∇^E.*

Proof. Let us check the case where $k = 0$ first:

$$\begin{aligned}
\left(d^E \circ d^E \sigma\right)(X_1, X_2) &= \left(2!Alt \circ \nabla^E \circ Alt \circ \nabla^E \sigma\right)(X_1, X_2) \\
&= \left(2!Alt \circ \nabla^E \nabla^E \sigma\right)(X_1, X_2) \\
&= \nabla_{X_1, X_2} \sigma - D_{X_2, X_1} \sigma = F(X_1, X_2)\sigma \\
&= (F \wedge \sigma)(X_1, X_2)
\end{aligned}$$

where the last equality is a trivial consequence of the definition of the action of $.\Omega^2(M, \mathrm{End}(E))$ on $\Omega^0(M, E)$.

Now more generally we just check $d^E \circ d^E$ on elements of the form $\sigma \otimes \theta$:

$$\begin{aligned} d^E \circ d^E \omega &= d^E \circ d^E (\sigma \otimes \theta) \\ &= d^E \left(d^E \sigma \wedge \theta + \sigma \otimes d\theta \right) \\ &= \left(d^E d^E \sigma \right) \wedge \theta - d^E \sigma \wedge d\theta + d^E \sigma \wedge d\theta + 0 \\ &= (F \wedge \sigma) \wedge \theta = F \wedge (\sigma \wedge \theta) \end{aligned}$$

■

Lets take a look at what things look like in a local frame field. As before restrict to an open set U on which $e_U = (e_1, ..., e_r)$ and then we write a typical element $s \in \Omega^k(M, E)$ as $s = e_U \omega_U$ where $\omega_U = (\omega^1, ..., \omega^r)^t$ is a column vector of smooth $k-$forms

$$\begin{aligned} d^E d^E s = d^E d^E \left(e_U \omega_U \right) &= d^E \left(e_U d\omega_U + d^E e_U \wedge \omega_U \right) \\ &= d^E \left(e_U \left(d\omega_U + A_U \wedge \omega_U \right) \right) \\ &= d^E e_U \wedge \left(d\omega_U + A_U \wedge \omega_U \right) \\ &+ e_U \wedge d^E \left(d\omega_U + A_U \wedge \omega_U \right) \\ &= e_U A \wedge \left(d\omega_U + A_U \wedge \omega_U \right) \\ &+ e_U \wedge d^E \left(d\omega_U + A_U \wedge \omega_U \right) \\ &= e_U A_U \wedge \left(d\omega_U + A_U \wedge \omega_U \right) \\ &+ e_U \wedge dA_U \wedge \omega_U - e_U \wedge A_U \wedge d\omega_U \\ &= e_U dA_U \wedge \omega_U + e_U A_U \wedge A_U \wedge \omega_U \\ &= e_U \left(dA_U + A_U \wedge A_U \right) \wedge \omega_U \end{aligned}$$

The matrix $dA_U + A_U \wedge A_U$ represents a section of $End(E)_U \otimes \wedge^2 TU$. In fact, we will now check that these local sections paste together to give a global section of $End(E) \otimes \wedge^2 TM$, i.e., and element of $\Omega^2(M, End(E))$ which is clearly the curvature form: $F : (X, Y) \mapsto F(X, Y) \in \Gamma(End(E))$. Let $F_U = dA_U + A_U \wedge A_U$ and let $F_V = dA_V + A_V \wedge A_V$ be corresponding form for a different moving frame $e_U := e_V g$ where $g : U \cap V \to \mathrm{GL}(\mathbb{F}^r), r = rank(E)$. What we need to verify this is the transformation law

$$F_V = g^{-1} F_U g$$

Recall that $A_V = g^{-1}A_U g + g^{-1}dg$. Using $d\left(g^{-1}\right) = -g^{-1}dgg^{-1}$ we have

$$\begin{aligned}
F_V &= dA_V + A_V \wedge A_V \\
&= d\left(g^{-1}A_U g + g^{-1}dg\right) \\
&+ \left(g^{-1}A_U g + g^{-1}dg\right) \wedge \left(g^{-1}A_U g + g^{-1}dg\right) \\
&= d\left(g^{-1}\right) A_U g + g^{-1}dA_U g - g^{-1}A_U dg \\
& g^{-1}A_U \wedge A_U g + g^{-1}dgg^{-1}A_U g + \\
& g^{-1}A_U dg + g^{-1}dg \wedge g^{-1}dg \\
&= g^{-1}dA_U g + g^{-1}A_U \wedge A_U g = g^{-1}F_U g.
\end{aligned}$$

where we have used $g^{-1}dg \wedge g^{-1}dg = d(g^{-1}g) = 0$.

19.6 Problem Set

1. Let M have a linear connection ∇ and let $T(X,Y) := \nabla_X Y - \nabla_Y X$. Show that T is tensorial ($C^\infty(M)-$bilinear). The resulting tensor is called the torsion tensor for the connection.

   Second order differential equations and sprays

2. A second order differential equation on a smooth manifold M is a vector field on TM, that is a section X of the bundle TTM (second tangent bundle) such that every integral curve α of X is the velocity curve of its projection on M. In other words, $\alpha = \dot\gamma$ where $\gamma := \pi_{TM} \circ \alpha$. A solution curve $\gamma : I \to M$ for a second order differential equation X is, by definition, a curve with $\ddot\alpha(t) = X(\dot\alpha(t))$ for all $\tau \in I$.

   In case $M = \mathbb{R}^n$ show that this concept corresponds to the usual systems of equations of the form

   $$\begin{aligned}
   y' &= v \\
   v' &= f(y, v)
   \end{aligned}$$

   which is the reduction to a first order system of the second order equation $y'' = f(y, y')$. What is the vector field on $T\mathbb{R}^n = \mathbb{R}^n \times \mathbb{R}^n$ which corresponds to this system?

   Notation: For a second order differential equation X the maximal integral curve through $v \in TM$ will be denoted by α_v and its projection will be denoted by $\gamma_v := \pi_{TM} \circ \alpha$.

3. A spray on M is a second order differential equation, that is, a section X of TTM as in the previous problem, such that for $v \in TM$ and $s \in \mathbb{R}$, a number $t \in \mathbb{R}$ belongs to the domain of γ_{sv} if and only if st belongs to the domain of γ_v and in this case

   $$\gamma_{sv}(t) = \gamma_v(st)$$

Show that there is an infinite number of sprays on any differentiable manifold M.

Hint: (1) Show that a vector field $X \in \mathfrak{X}(TM)$ is a spray if and only if $T\pi \circ X = \mathrm{id}_{TM}$ where $\pi = \pi_{TM}$:

$$\begin{array}{ccc} & & TTM \\ & X \nearrow & \downarrow \\ TM & \overset{\mathrm{id}_{TM}}{\to} & TM \end{array}$$

(2) Show that $X \in \mathfrak{X}(TM)$ is a spray if and only if for any $s \in \mathbb{R}$ and any $v \in TM$

$$X_{sv} = T\mu_s(sX_v)$$

where $\mu_s : v \mapsto sv$ is the multiplication map.

(3) Show the existence of a spray on an open ball in a Euclidean space.

(4) Show that X_1 and X_2 both satisfies one of the two characterizations of a spray above then so does any convex combination of X_1 and X_2.

4. Show that if one has a linear connection on M then there is a spray whose solutions are the geodesics of the connection.

5. Show that given a spray on a manifold there is a (not unique) linear connection ∇on the manifold such that $\gamma : I \to M$ is a solution curve of the spray if and only if $\left(\nabla_{\dot\gamma(t)}\dot\gamma\right)(t) = 0$ for all $t \in I$. Note: γ is called a geodesic for the linear connection or the spray. Does the stipulation that the connection be torsion free force uniqueness?

6. Let X be a spray on M. Equivalently, we may start with a linear connection on M which induces a spay. Show that the set $O_X := \{v \in TM : \gamma_v(1) \text{ is defined}\}$ is an open neighborhood of the zero section of TM.

Chapter 20

Riemannian and semi-Riemannian Geometry

The most beautiful thing we can experience is the mysterious. It is the source of all true art and science.
-Albert Einstein

Geometry has a long history indeed and hold a central position in mathematics but also in philosophy both ancient and modern and also in physics. In fact, we might safely find the origin of geometry in the physical universe itself and the fact ever mobile creature must be able to some kind of implicit understanding of geometry. Knowledge of geometry remained largely implicitly even in the human understanding until the intellectual pursuits of the ancient Greeks introduced geometry as a specific and self conscious discipline. One of the earliest figures associated with geometry is the Greek philosopher Thales, born circa 600 BC in Miletus, Asia Minor. He is said to be the founder of philosophy, one of the Seven Wise Men of Greece. Even more famous, at least among mathematicians, is Pythagoras who was instructed in the teachings of Thales. Pythagoras was the founder of religious sect which put mathematics, especially number and geometry, in a central and sacred position. According to the Pythagoreans, the world itself was as bottom nothing but number and proportion. This ontology may seem ridiculous at first glance but modern physics tells a story that is almost exclusively mathematical. Of course, these days we have a deeper appreciation for the role of experimentation and measurement. Philosophers have also become more sensitive to the ways in which our embodiment in the world and our use of metaphor and other tropes influences and constrains our mathematical understanding.

The theorem of Pythagoras relating the sides of right triangles is familiar to every student of mathematics and is of interest in differential geometry (more specifically Riemannian geometry) since the theorem fails to hold in the presence of curvature. Curvature is defined any time one has a connection but the curvature appropriate to semi-Riemannian geometry is of a very special sort since it arises completely out of the metric structure on the manifold. Thus curvature is truly connected to what we might loosely call the shape of the manifold.

The next historical figure we mention is Euclid (300 BC), also a Greek, who changed the rules of the game by introducing the notions of *axiom* and *proof*. Euclid's axioms are also called postulates. The lasting influence on mathematics of Euclid's approach goes without saying but the axiomatic approach to geometric knowledge has also always been a paradigmatic example for philosophy. Theories of a priori knowledge and apodeictic knowledge seem to have Euclid's axiomatics in the background. Euclid's major work is a 13 volume treatise called Elements. A large portion of what is found in this treatise is the axiomatic geometry of planar figures which roughly corresponds to what has been taught in secondary school for many years. What stands out for our purposes is Euclid's fifth axiom. The point is that exactly the same geometries for which the Pythagorean theorem fails are also geometries for which this fifth postulate fails to hold. In modern terms the postulates are equivalent to the following:

1. For every point P and every point $Q \neq P$ there exists a unique line that passes through P and Q.

2. For every pair of segments AB and CD there exists a unique point E such that B is between A and E and segment CD is congruent to segment BE.

3. For every point O and every point $A \neq O$ there exists a circle with center O and radius OA.

4. All right angles are congruent to each other.

5. For every line ℓ and for every point P that is not on ℓ there exists a unique line m through P that is parallel to ℓ.

Of course, "parallel" is a term that must be defined. One possible definition (the one we use) is that two lines are parallel if they do not intersect.

It was originally thought on the basis of intuition that the fifth postulate was a necessary truth and perhaps a logical necessity following from the other 4 postulates. The geometry of a sphere has geodesics as lines (great circles in this case). By identifying antipodal points we get projective space where the lines are the projective great circles. This geometry satisfies the first four axioms above but the fifth does not hold when we make the most useful definition for what it means to be parallel. In fact, on a sphere we find that given a geodesic, and a point not on the geodesic there are no parallel lines since all lines (geodesics) must intersect. After projectivization we get a geometry for which the first four postulates hold but the fifth does not. Thus the fifth cannot be a logical consequence of the first four.

It was Rene Descarte that made geometry susceptible to study via the real number system through the introduction of the modern notion of coordinate system and it is to this origin that the notion of a differentiable manifold can be traced. Of course, Descartes did not define the notion of a differentiable manifold. That task was left to Gauss, Riemann and Whitney. Descartes did make the first systematic connection between the real numbers and the geometry of the line, the plane, and 3-dimensional space. Once the introduction of Descartes' coordinates made many theorems of geometry become significantly easier to prove. In fact, it is said that Descartes was annoyed axiomatic geometry because to much tedious work was required obtain results. Modern differential geometry carries on this tradition by framing everything in the setting of differentiable manifolds where one can always avail oneself of various coordinates.

At first it was not clear that there should be a distinction between the purely differential topological notion of a manifold (such as the plane) which does not include the notion of distance, and the notion of the Riemannian manifold (the rigid plane being an example). A (semi-) Riemannian manifold comes with notions of distance, volume, curvature and so on, all of which are born out of the fundamental tensor called the metric tensor. Today we clearly distinguish between the notions of topological, differential topological and the more rigid notion of (semi-) Riemannian manifold even though it is difficult to achieve a mental representation of even a plane or surface without imposing a metric in imagination. I usually think of a purely differential topological version of, say, a plane or sphere as a wobbling shape shifting surface with the correct topology. Thus the picture is actually of a family of manifolds with various metrics. This is only a mental trick but it is important in (semi-) Riemannian geometry to keep the distinction between a (semi-) Riemannian manifold (defined below) and the underlying differentiable manifold. For example, one may ask questions like 'given a specific differentiable manifold M, what type of metric with what possible curvatures can be imposed on M?'.

The distinction between a Riemannian manifold and the more inclusive notion of a semi-Riemannian manifold is born out of the distinction between a merely nondegenerate symmetric, bilinear form and that of a positive definite, symmetric bilinear form. One special class of semi-Riemannian manifolds called the Lorentz manifold make their appearance in the highly differential geometric theory of gravitation know as Einstein's general theory of relativity. After an exposition of the basics of general semi-Riemannian geometry we will study separately the special cases of Riemannian geometry and Lorentz geometry.

20.1 Tensors and Scalar Products

20.1.1 Scalar Products

Definition 20.1 *A scalar product on a (real) finite dimensional vector space* V *is a nondegenerate symmetric bilinear form* $\mathrm{g} : \mathrm{V} \times \mathrm{V} \to \mathbb{R}$. *The scalar product is called*

1. *positive (resp. negative) definite if* $\mathrm{g}(v, v) \geq 0$ *(resp.* $\mathrm{g}(v, v) \leq 0$*) for all* $v \in \mathrm{V}$ *and* $\mathrm{g}(v, v) = 0 \Longrightarrow v = 0$.

2. *positive (resp. negative) semidefinite if* $\mathrm{g}(v, v) \geq 0$ *(resp.*$\mathrm{g}(v, v) \leq 0$*) for all* $v \in \mathrm{V}$.

Nondegenerate means that the map $\mathrm{g} : \mathrm{V} \to \mathrm{V}^*$ given by $v \mapsto \mathrm{g}(v, .)$ is a linear isomorphism or equivalently, if $\mathrm{g}(v, w) = 0$ for all $w \in V$ implies that

$v = 0$.

Definition 20.2 *A* ***scalar product space*** *is a pair* V, g *where* V *is a vector space and* g *is a scalar product. In case the scalar product is positive definite we shall also refer to it as an* ***inner product*** *and the pair* V, g *as an inner product space.*

Definition 20.3 *The* ***index*** *of a symmetric bilinear* g *form on* V *is the largest subspace* $W \subset V$ *such that the restriction* $\mathrm{g}|_W$ *is negative definite.*

Given a basis $\mathcal{F} = (f_1, ..., f_n)$ for V we may form the matrix $[\mathrm{g}]_{\mathcal{F}}$ which has as ij-th entry $g_{ij} := \mathrm{g}(f_i, f_j)$. This is the matrix that represents g with respect to the basis $\mathcal{F}$. So if $v = \mathcal{F}[v]^{\mathcal{F}} = \sum_{i=1}^n v^i f_i$, $w = \mathcal{F}[w]^{\mathcal{F}} = \sum_{i=1}^n w^i f_i$ then

$$\mathrm{g}(v, w) = [v]^{\mathcal{F}} [\mathrm{g}]_{\mathcal{F}} [w]^{\mathcal{F}} = g_{ij} v^i w^j.$$

It is a standard fact from linear algebra that if g is a scalar product then there exists a basis $e_1, ..., e_n$ for V such that the matrix representative of g with respect to this basis is a diagonal matrix $\mathrm{diag}(-1, ..., 1)$ with ones or minus ones along the diagonal and we may arrange for the minus ones come first. Such a basis is called an orthonormal basis for V, g. The number of minus ones appearing is the index $\mathrm{ind}(\mathrm{g})$ and so is independent of the orthonormal basis chosen. We sometimes denote the index by the Greek letter ν. It is easy to see that the index $\mathrm{ind}(\mathrm{g})$ is zero if and only if g positive semidefinite. Thus if $e_1, ..., e_n$ is an orthonormal basis for V, g then $\mathrm{g}(e_i, e_j) = \varepsilon_i \delta_{ij}$ where $\varepsilon_i = \mathrm{g}(e_i, e_i) = \pm 1$are the entries of the diagonal matrix the first $\mathrm{ind}(\mathrm{g})$ of which are equal to -1 and the remaining are equal to 1. Let us refer to the list of ± 1 given by $(\varepsilon_1, ..., \varepsilon_n)$ as the **signature**.

Remark 20.1 *The convention of putting the minus signs first is not universal and in fact we reserve the right to change the convention to a positive first convention but ample warning will be given. The negative signs first convention is popular in relativity theory but the reverse is usual in quantum field theory. It makes no physical difference in the final analysis as long as one is consistent but it can be confusing when comparing references from the literature.*

Another difference between the theory of positive definite scalar products and indefinite scalar products is the appearance of the ε_i from the signature in formulas that would be familiar in the positive definite case. For example we have the following:

Proposition 20.1 *Let* $e_1, ..., e_n$ *be an orthonormal basis for* V, g. *For any* $v \in V, \mathrm{g}$ *we have a unique expansion given by* $v = \sum_i \varepsilon_i \langle v, e_i \rangle e_i$.

Proof. The usual proof works. One just has to notice the appearance of the ε_i. ■

Definition 20.4 *If $v \in V$ then let $|v|$ denote the nonnegative number $|\mathrm{g}(v,v)|^{1/2}$ and call this the (absolute or positive)* ***length*** *of v. Some authors call $\mathrm{g}(v,v)$ or $\mathrm{g}(v,v)^{1/2}$ the length which would make it possible for the length to be complex valued. We will avoid this.*

Definition 20.5 *Let V, g be a scalar product space. We say that v and w are mutually orthogonal if and only if $\mathrm{g}(v,w) = 0$. Furthermore, given two subspaces W_1 and W_2 of V we say that W_1 is orthogonal to W_2 and write $\mathrm{W}_1 \perp \mathrm{W}_2$ if and only if every element of W_1 is orthogonal to every element of W_2.*

Since in general g is not necessarily positive definite or negative definite it may be that there are elements that are orthogonal to themselves.

Definition 20.6 *Given a subspace W of a scaler product space V we define the orthogonal complement as $\mathrm{W}^\perp = \{v \in V : \mathrm{g}(v,w) = 0 \text{ for all } w \in \mathrm{W}\}$.*

We always have $\dim(\mathrm{W}) + \dim(\mathrm{W}^\perp) = \dim(\mathrm{V})$ but unless g is definite we may *not* have $\mathrm{W} \cap \mathrm{W}^\perp = \emptyset$. Of course by nondegeneracy we will always have $V^\perp = 0$.

Definition 20.7 *A subspace W of a scaler product space V, g is called* ***nondegenerate*** *if $\mathrm{g}|_{\mathrm{W}}$ is nondegenerate.*

Lemma 20.1 *A subspace $\mathrm{W} \subset V, \mathrm{g}$ is nondegenerate if and only if $\mathrm{V} = \mathrm{W} \oplus \mathrm{W}^\perp$ (inner direct sum).*

Proof. Easy exercise in linear algebra. ■

Just as for inner product spaces we call a linear isomorphism $I : \mathrm{V}_1, \mathrm{g}_1 \to \mathrm{V}_2, \mathrm{g}_2$ from one scalar product space to another to an **isometry** if $\mathrm{g}_1(v,w) = \mathrm{g}_2(Iv, Iw)$. It is not hard to show that if such an isometry exists then g_1 and g_2 have the same index and signature.

20.1.2 Natural Extensions and the Star Operator

If we have a scalar product g on a finite dimensional vector space V then there is a natural way to induce a scalar product on the various tensor spaces $T^r_s(\mathrm{V})$ and on the Grassmann algebra. The best way to explain is by way of some examples.

First consider V^*. Since g is nondegenerate there is a linear isomorphism map $\mathrm{g}_\flat : \mathrm{V} \to \mathrm{V}^*$ defined by

$$\mathrm{g}_\flat(v)(w) = \mathrm{g}(v,w).$$

Denote the inverse by $\mathrm{g}^\sharp : \mathrm{V}^* \to \mathrm{V}$. We force this to be an isometry by defining the scalar product on V^* to be

$$\mathrm{g}^*(\alpha, \beta) = \mathrm{g}(\mathrm{g}^\sharp(\alpha), \mathrm{g}^\sharp(\beta)).$$

Under this prescription, the dual basis $e^1, ..., e^n$ corresponding to an orthonormal basis $e_1, ..., e_n$ for V will also be orthonormal. The signature (and hence the index) of g^* and g are the same.

Next consider $T^1_1(\mathrm{V}) = \mathrm{V} \otimes \mathrm{V}^*$. We define the scalar product of two simple tensors $v_1 \otimes \alpha_1$ and $v_2 \otimes \alpha_2 \in \mathrm{V} \otimes \mathrm{V}^*$ by

$$\mathrm{g}^1_1(v_1 \otimes \alpha_1, v_2 \otimes \alpha_2) = \mathrm{g}(v_1, v_2)\mathrm{g}^*(\alpha_1, \alpha_2).$$

One can then see that for orthonormal dual bases $e^1, ..., e^n$ and $e_1, ..., e_n$ we have that

$$\{e_i \otimes e^j\}_{1 \le i,j \le n}$$

is an orthonormal basis for $T^1_1(\mathrm{V}), \mathrm{g}^1_1$. In general one defines g^r_s so that the natural basis for $T^r_s(\mathrm{V})$ formed from the orthonormal basis $\{e^1, ..., e^n\}$ (and its dual $\{e_1, ..., e_n\}$), will also be orthonormal.

Notation 20.1 *In order to reduce notational clutter let us reserve the option to denote all these scalar products coming from* g *by the same letter* g *or, even more conveniently, by* $\langle .,. \rangle$.

Exercise 20.1 *Show that under the natural identification of* $\mathrm{V} \otimes \mathrm{V}^*$ *with* $L(\mathrm{V}, \mathrm{V})$ *the scalar product of linear transformations* A *and* B *is given by* $\langle A, B \rangle = \mathrm{trace}(A^t B)$.

Next we see how to extend the maps $\mathrm{g}_\flat$ and $\mathrm{g}^\sharp$ to maps on tensors. We give two ways of defining the extensions. In either case, what we want to define is maps $(\mathrm{g}_\flat)^i_\downarrow : T^r{}_s(\mathrm{V}) \to T^{r-1}{}_{s+1}(\mathrm{V})$ and $(\mathrm{g}^\sharp)^\uparrow_j : T^r{}_s(\mathrm{V}) \to T^{r+1}{}_{s-1}(\mathrm{V})$ where $0 \le i \le r$ and $0 \le j \le s$. It is enough to give the definition for simple tensors:

$$\begin{aligned}&(\mathrm{g}_\flat)^i_\downarrow(w_1 \otimes \cdots \otimes w_r \otimes \omega^1 \otimes \cdots \otimes \omega^s)\\ &:= w_1 \otimes \cdots \otimes \widehat{w_i} \otimes \cdots w_r \otimes \mathrm{g}_\flat(w_i) \otimes \omega^1 \otimes \cdots \otimes \omega^s\end{aligned}$$

and

$$\begin{aligned}&(\mathrm{g}^\sharp)^\uparrow_j(w_1 \otimes \cdots \otimes w_r \otimes \omega^1 \otimes \cdots \otimes \omega^s)\\ &:= w_1 \otimes \cdots \otimes w_r \otimes \mathrm{g}^\sharp(\omega^j) \otimes \omega^1 \otimes \cdots \otimes \widehat{\omega^j} \otimes \cdots \otimes \omega^s.\end{aligned}$$

This definition is extended to all of $T^r_s(\mathrm{V})$ by linearity. For our second, equivalent definition let $\Upsilon \in T^r{}_s(\mathrm{V})$. Then

$$\begin{aligned}&((\mathrm{g}_\flat)^i_\downarrow \Upsilon)(\alpha^1, ..., \alpha^{r-1}; v_1, ..., v_{s+1})\\ &:= \Upsilon(\alpha^1, ..., \alpha^{r-1}, \mathrm{g}_\flat(v_1); v_2, ..., v_{s+1})\end{aligned}$$

and similarly

$$\begin{aligned}&((\mathrm{g}^\sharp)^\uparrow_j \Upsilon)(\alpha^1, ..., \alpha^{r+1}; v_1, ..., v_{s-1})\\ &:= \Upsilon(\alpha^1, ..., \alpha^r, \mathrm{g}^\sharp(\alpha^{r+1}); v_2, ..., v_{s+1})\end{aligned}$$

Lets us see what this looks like by viewing the components. Let $f_1, ..., f_n$ be an arbitrary basis of V and let $f^1, ..., f^n$ be the dual basis for V^*. Let $g_{ij} := g(f_i, f_j)$ and $g^{ij} = g^*(f^i, f^j)$. The reader should check that $\sum_k g^{ik} g_{kj} = \delta^i_j$. Now let $\tau \in T^r_s(V)$ and write

$$\tau = \tau^{i_1,..,i_r}{}_{j_1,...,j_s} f_{i_1} \otimes \cdots \otimes f_{i_r} \otimes f^{j_1} \otimes \cdots \otimes f^{j_s}.$$

Define

$$\tau^{i_1,...,i_r j}_{j_1,...,j_{b-1},\widehat{k},j_b...,j_{s-1}} := \tau^{i_1,...,i_r}{}_{j_1,...,j_{j-1}}{}^{j}{}_{j_{b+1}...,j_{s-1}} := \sum_m b^{km} \tau^{i_1,..,i_r}_{j_1,...,j_{b-1},m,j_b...,j_{s-1}}$$

then

$$(g^\sharp)^{\uparrow}_a \tau = \tau^{i_1,...,i_r j_a}_{j_1,...,j_{a-1},\widehat{j_a},j_{a+1}...,j_{s-1}} f_{i_1} \otimes \cdots \otimes f_{i_r} \otimes f_{j_a} \otimes f^{j_1} \otimes \cdots \otimes f^{j_{s-1}}.$$

Thus the $(g^\sharp)^{\uparrow}_a$ visually seems to raise an index out of a-th place and puts it up in the last place above. Similarly, the component version of lowering $(g_\flat)^a_\downarrow$ takes

$$\tau^{i_1,..,i_r}{}_{j_1,...,j_s}$$

and produces

$$\tau^{i_1,..,\widehat{i_a},...i_r}_{i_a j_1,...,j_s} = \tau^{i_1,..,}{}_{i_a}{}^{,...i_r}{}_{j_1,...,j_s}.$$

How and why would one do this so called index raising and lowering? What motivates the choice of the slots? In practice one applies this type changing only in specific well motivated situations and the choice of slot placement is at least partially conventional. We will comment further when we actually apply these operators. The notation is suggestive and the $g^\sharp$ and $g_\flat$ and their extensions are referred to as **musical isomorphisms** . One thing that is useful to know is that if we raise all the lower indices and lower all the upper ones on a tensor then we can "completely contract" against another tensor of the original type with the result being the scalar product. For example, let $\tau = \sum \tau_{ij} f^i \otimes f^j$ and $\chi = \sum \chi_{ij} f^i \otimes f^j$. Then letting the components of $(g^\sharp)^{\uparrow}_1 \circ (g^\sharp)^{\uparrow}_1(\chi)$ by χ^{ij} we have

$$\chi^{ij} = g^{ik} g^{jl} \chi_{kl}$$

and

$$\langle \chi, \tau \rangle = \sum \chi_{ij} \tau^{ij}.$$

In general, unless otherwise indicated, we will preform repeated index raising by raising from the first slot $(g^\sharp)^{\uparrow}_1 \circ \cdots \circ (g^\sharp)^{\uparrow}_1$ and similarly for repeated lowering $(g_\flat)^1_\downarrow \circ \cdots \circ (g_\flat)^1_\downarrow$. For example,

$$A_{ijkl} \mapsto A^i{}_{jkl} = g^{ia} A_{ajkl} \mapsto A^{ij}{}_{kl} = g^{ia} g^{jb} A_{abkl}$$

Exercise 20.2 *Verify the above claim directly from the definition of* $\langle \chi, \tau \rangle$.

Even though elements of $L^k_{alt}(\mathrm{V}) \cong \bigwedge^k(\mathrm{V}^*)$ can be thought of as tensors of type $0, k$ that just happen to be anti-symmetric, it is better in most cases to give a scalar product to this space in such a way that the basis

$$\{e^{i_1} \wedge \cdots \wedge e^{i_k}\}_{i_1 < \ldots < i_k} = \{e^{\vec{I}}\}$$

is orthonormal if $e^1, ..., e^n$ is orthonormal. Now given any k-form $\alpha = a_{\vec{I}} e^{\vec{I}}$ where $e^{\vec{I}} = e^{i_1} \wedge \cdots \wedge e^{i_k}$ with $i_1 < ... < i_k$ as explained earlier, we can also write $\alpha = \frac{1}{k!} a_I e^I$ and then as a tensor

$$\begin{aligned} \alpha &= \frac{1}{k!} a_I e^I \\ &= \frac{1}{k!} a_{i_1 \ldots i_k} e^{i_1} \otimes \cdots \otimes e^{i_k}. \end{aligned}$$

Thus if $\alpha = a_{\vec{I}} e^{\vec{I}}$ α and $\beta = b_{\vec{I}} e^{\vec{I}}$ are k-forms considered as covariant tensor fields we have

$$\begin{aligned} \langle \alpha, \beta \rangle &= \frac{1}{(k!)^2} a_{i_1 \ldots i_k} b^{i_1 \ldots i_k} \\ &= a_I b^I \end{aligned}$$

But when $\alpha = a_{\vec{I}} e^{\vec{I}}$ α and $\beta = b_{\vec{I}} e^{\vec{I}}$ are viewed as k-forms then we must have

$$\begin{aligned} \langle \alpha, \beta \rangle &= a_{\vec{I}} b^{\vec{I}} \\ &= \frac{1}{k!} a_I b^I \end{aligned}$$

so the two definitions are different by a factor of $k!$. The reason for the discrepancy might be traced to our normalization when we defined the exterior product. If we had defined the exterior product as $Alt(\alpha \otimes \beta)$ we would not have this problem. Of course the cost would be that some other formulas would become cluttered. The definition for forms can be written succinctly in terms of $1-$forms as

$$\begin{aligned} &\langle \alpha^1 \wedge \alpha^2 \wedge \cdots \wedge \alpha^k, \beta^1 \wedge \beta^2 \wedge \cdots \wedge \beta^k \rangle \\ &= \det(\langle \alpha^i, \beta^j \rangle) \end{aligned}$$

where the α^i and β^i are 1-forms.

Definition 20.8 *We define the scalar product on* $\bigwedge^k \mathrm{V}^* \cong L^k_{alt}(\mathrm{V})$ *by first using the above formula for wedge products of 1-forms and then we extending (bi)linearly to all of* $\bigwedge^k \mathrm{V}^*$. *We can also extend to the whole Grassmann algebra* $\bigwedge \mathrm{V}^* = \bigoplus \bigwedge^k \mathrm{V}^*$ *by declaring forms of different degree to be orthogonal. We also have the obvious similar definition for* $\bigwedge^k \mathrm{V}$ *and* $\bigwedge \mathrm{V}$.

We would now like to exhibit the definition of the very useful star operator. This will be a map from $\bigwedge^k \mathrm{V}^*$ to $\bigwedge^{n-k} \mathrm{V}^*$ for each k, $1 \le k \le n$ where $n = \dim(M)$. First of all if we have an orthonormal basis $e^1, ..., e^n$ for V^* then $e^1 \wedge \cdots \wedge e^n \in \bigwedge^n \mathrm{V}^*$. But $\bigwedge^n \mathrm{V}^*$ is one dimensional and if $\ell : \mathrm{V}^* \to \mathrm{V}^*$ is any isometry of V^* then $\ell e^1 \wedge \cdots \wedge \ell e^n = \pm e^1 \wedge \cdots \wedge e^n$. In particular, for any permutation σ of the letters $\{1, 2, ..., n\}$ we have $e^1 \wedge \cdots \wedge e^n = sgn(\sigma) e^{\sigma 1} \wedge \cdots \wedge e^{\sigma n}$.

For a given $\mathcal{E}^* = \{e^1, ..., e^n\}$ for V^* (with dual basis for orthonormal basis $\mathcal{E} = \{e_1, ..., e_n\}$) us denote $e^1 \wedge \cdots \wedge e^n$ by $\varepsilon(\mathcal{E}^*)$. Then we have

$$\langle \varepsilon(\mathcal{E}^*), \varepsilon(\mathcal{E}^*) \rangle = \varepsilon_1 \varepsilon_2 \cdots \varepsilon_n = \pm 1$$

and the only elements ω of $\bigwedge^n \mathrm{V}^*$ with $\langle \omega, \omega \rangle = \pm 1$ are $\varepsilon(\mathcal{E}^*)$ and $-\varepsilon(\mathcal{E}^*)$. Given a fixed orthonormal basis $\mathcal{E}^* = \{e^1, ..., e^n\}$, all other orthonormal bases $\mathcal{B}^*$ bases for V^* fall into two classes. Namely, those for which $\varepsilon(\mathcal{B}^*) = \varepsilon(\mathcal{E}^*)$ and those for which $\varepsilon(\mathcal{B}^*) = -\varepsilon(\mathcal{E}^*)$. Each of these two top forms $\pm\varepsilon(\mathcal{E}^*)$ is called a metric volume element for $\mathrm{V}^*, \mathrm{g}^* = \langle , \rangle$. A choice of orthonormal basis determines one of these two volume elements and provides an orientation for V^*. On the other hand, we have seen that any nonzero top form ω determines an orientation. If we have an orientation given by a top form ω then $\mathcal{E} = \{e_1, ..., e_n\}$ determines the same orientation if and only if $\omega(e_1, ..., e_n) > 0$.

Definition 20.9 *Let an orientation be chosen on* V^* *and let* $\mathcal{E}^* = \{e^1, ..., e^n\}$ *be an oriented orthonormal frame so that* $vol := \varepsilon(\mathcal{E}^*)$ *is the corresponding volume element. Then if* $\mathcal{F} = \{f_1, ..., f_n\}$ *is a basis for* V *with dual basis* $\mathcal{F}^* = \{f^1, ..., f^n\}$ *then*

$$vol = \sqrt{|\det(\mathrm{g}_{ij})|} f^1 \wedge \cdots \wedge f^n$$

where $\mathrm{g}_{ij} = \langle f_i, f_j \rangle$.

Proof. Let $e^i = a^i_j f^j$ then

$$\begin{aligned} \varepsilon_i \delta^{ij} &= \pm\delta^{ij} = \langle e^i, e^j \rangle = \langle a^i_k f^k, a^j_m f^m \rangle \\ &= a^i_k a^j_m \langle f^k, f^m \rangle = a^i_k a^j_m \mathrm{g}^{km} \end{aligned}$$

so that $\pm 1 = \det(a^i_k)^2 \det(\mathrm{g}^{km}) = (\det(a^i_k))^2 (\det(\mathrm{g}_{ij}))^{-1}$ and so

$$\sqrt{|\det(\mathrm{g}_{ij})|} = \det(a^i_k).$$

On the other hand,

$$\begin{aligned} \mathrm{vol} &:= \varepsilon(\mathcal{E}^*) = e^1 \wedge \cdots \wedge e^n \\ &= a^1_{k_1} f^{k_1} \wedge \cdots \wedge a^n_{k_1} f^{k_1} = \det(a^i_k) f^1 \wedge \cdots \wedge f^n \end{aligned}$$

and the result follows. ■

Fix an orientation and let $\mathcal{E}^* = \{e^1, ..., e^n\}$ be an orthonormal basis in that orientation class. Then we have chosen one of the two volume forms, say $\mathrm{vol} = \varepsilon(\mathcal{E}^*)$. Now we define $* : \bigwedge^k \mathrm{V}^* \to \bigwedge^{n-k} \mathrm{V}^*$ by first giving the definition on basis elements and then extending by linearity.

Definition 20.10 *Let* V^* *be oriented and let* $\{e^1, ..., e^n\}$ *be a positively oriented orthonormal basis. Let* σ *be a permutation of* $(1, 2, ..., n)$*. On the basis elements* $e^{\sigma(1)} \wedge \cdots \wedge e^{\sigma(k)}$ *for* $\bigwedge^k \mathrm{V}^*$ *define*

$$*(e^{\sigma(1)} \wedge \cdots \wedge e^{\sigma(k)}) = \varepsilon_{\sigma_1}\varepsilon_{\sigma_2}\cdots\varepsilon_{\sigma_k} sgn(\sigma) e^{\sigma(k+1)} \wedge \cdots \wedge e^{\sigma(n)}.$$

In other words,

$$*(e^{i_1} \wedge \cdots \wedge e^{i_k}) = \pm(\varepsilon_{i_1}\varepsilon_{i_2}\cdots\varepsilon_{i_k}) e^{j_1} \wedge \cdots \wedge e^{j_{n-k}}$$

where we take the $+$ *sign if and only if* $e^{i_1} \wedge \cdots \wedge e^{i_k} \wedge e^{j_1} \wedge \cdots \wedge e^{j_{n-k}} = e^1 \wedge \cdots \wedge e^n$.

Remark 20.2 *In case the scalar product is positive definite* $\varepsilon_1 = \varepsilon_2 \cdots = \varepsilon_n = 1$ *and so the formulas are a bit less cluttered.*

We may develop a formula for the star operator in terms of an arbitrary basis.

Lemma 20.2 *For* $\alpha, \beta \in \bigwedge^k \mathrm{V}^*$ *we have*

$$\langle \alpha, \beta \rangle vol = \alpha \wedge *\beta$$

Proof. It is enough to check this on typical basis elements $e^{i_1} \wedge \cdots \wedge e^{i_k}$ and $e^{m_1} \wedge \cdots \wedge e^{m_k}$. We have

$$\begin{aligned} &(e^{m_1} \wedge \cdots \wedge e^{m_k}) \wedge *(e^{i_1} \wedge \cdots \wedge e^{i_k}) \\ &= e^{m_1} \wedge \cdots \wedge e^{m_k} \wedge (\pm e^{j_1} \wedge \cdots \wedge e^{j_{n-k}}) \end{aligned} \tag{20.1}$$

This latter expression is zero unless $\{m_1, ..., m_k\} \cup \{j_1, ..., j_{n-k}\} = \{1, 2, ..., n\}$ or in other words, unless $\{i_1, ..., i_k\} = \{m_1, ..., m_k\}$. But this is also true for

$$\langle e^{m_1} \wedge \cdots \wedge e^{m_k}, e^{i_1} \wedge \cdots \wedge e^{i_k} \rangle \,\mathrm{vol}. \tag{20.2}$$

On the other hand if $\{i_1, ..., i_k\} = \{m_1, ..., m_k\}$ then both 20.1 and 20.2 give $\pm\,\mathrm{vol}$. So the lemma is proved up to a sign. We leave it to the reader to show that the definitions are such that the signs match. ■

Proposition 20.2 *The following identities hold for the star operator:*

1) $*1 = vol$
2) $*vol = (-1)^{\mathrm{ind(g)}}$
3) $**\alpha = (-1)^{\mathrm{ind(g)}}(-1)^{k(n-k)}\alpha$ *for* $\alpha \in \bigwedge^k \mathrm{V}^*$.

Proof. (1) and (2) follow directly from the definitions. For (3) we must first compute $*(e^{j_{k+1}} \wedge \cdots \wedge e^{j_n})$. We must have $*(e^{j_{k+1}} \wedge \cdots \wedge e^{j_n}) = ce^{j_1} \wedge \cdots \wedge e^{j_k}$ for some constant c. On the other hand,

$$\begin{aligned} c\varepsilon_1\varepsilon_2\cdots\varepsilon_n \,\mathrm{vol} &= \langle e^{j_{k+1}} \wedge \cdots \wedge e^{j_n}, ce^{j_1} \wedge \cdots \wedge e^{j_k} \rangle \\ &= \langle e^{j_{k+1}} \wedge \cdots \wedge e^{j_n}, *(e^{j_{k+1}} \wedge \cdots \wedge e^{j_n}) \rangle \\ &= \varepsilon_{j_{k+1}}\cdots\varepsilon_{j_n} sgn(j_{k+1}\cdots j_n, j_1\cdots j_k) e^{j_{k+1}} \wedge \cdots \wedge e^{j_n} \wedge e^{j_1} \wedge \cdots \wedge e^{j_k} \\ &\varepsilon_{j_{k+1}}\cdots\varepsilon_{j_n} sgn(j_{k+1}\cdots j_n, j_1\cdots j_k) \,\mathrm{vol} \\ &= (-1)^{k(n-k)}\varepsilon_{j_{k+1}}\cdots\varepsilon_{j_n} \,\mathrm{vol} \end{aligned}$$

so that $c = \varepsilon_{j_{k+1}} \cdots \varepsilon_{j_n} (-1)^{k(n-k)}$. Using this we have, for any permutation $J = (j_1, ..., j_n)$,

$$\begin{aligned} * * (e^{j_1} \wedge \cdots \wedge e^{j_k}) &= * \varepsilon_{j_1} \varepsilon_{j_2} \cdots \varepsilon_{j_k} sgn(J) e^{j_{k+1}} \wedge \cdots \wedge e^{j_n} \\ &= \varepsilon_{j_1} \varepsilon_{j_2} \cdots \varepsilon_{j_k} \varepsilon_{j_{k+1}} \cdots \varepsilon_{j_n} sgn(J) e^{j_1} \wedge \cdots \wedge e^{j_k} \\ &= (-1)^{\mathrm{ind(g)}} (-1)^{k(n-k)} e^{j_1} \wedge \cdots \wedge e^{j_k} \end{aligned}$$

which implies the result. ∎

20.2 Riemannian and semi-Riemannian Metrics

Consider a regular submanifold M of a Euclidean space, say $\mathbb{R}^n$. Since we identify T_pM as a subspace of $T_p\mathbb{R}^n \cong \mathbb{R}^n$ and the notion of length of tangent vectors makes sense on $\mathbb{R}^n$ it also makes sense for vectors in T_pM. In fact, if $X_p, Y_p \in T_pM$ and c_1, c_2 are some curves with $\dot{c}_1(0) = X_p$, $\dot{c}_2(0) = Y_p$ then c_1 and c_2 are also a curves in $\mathbb{R}^n$. Thus we have an inner product defined $\mathrm{g}_p(X_p, Y_p) = \langle X_p, Y_p \rangle$. For a manifold that is not given as submanifold of some $\mathbb{R}^n$ we must have an inner product assigned to each tangent space as part of an extra structure. The assignment of a nondegenerate symmetric bilinear form $\mathrm{g}_p \in T_pM$ for every p in a smooth way defines a tensor field $\mathrm{g} \in \mathfrak{X}^0_2(M)$ on M called metric tensor.

Definition 20.11 *If* $\mathrm{g} \in \mathfrak{X}^0_2(M)$ *is nondegenerate, symmetric and positive definite at every tangent space we call* g *a* ***Riemannian metric*** *(tensor). If* g *is a Riemannian metric then we call the pair* M, g *a* ***Riemannian manifold*** .

The Riemannian manifold as we have defined it is the notion that best generalizes to manifolds the metric notions from surfaces such as arc length of a curve, area (or volume), curvature and so on. But because of the structure of spacetime as expressed by general relativity we need to allow the metric to be indefinite. In this case, some nonzero vectors v might have zero or negative self scalar product $\langle v, v \rangle$.

Recall that the index $\nu = \mathrm{ind(g)}$ of a bilinear form is the number of negative ones appearing in the signature $(-1, ...1)$.

Definition 20.12 *If* $\mathrm{g} \in \mathfrak{X}^0_2(M)$ *is symmetric nondegenerate and has constant index on* M *then we call* g *a semi-Riemannian metric and* M, g *a* ***semi-Riemannian manifold*** *or* ***pseudo-Riemannian manifold.*** *The index is called the* ***index*** *of* M, g *and denoted* $\mathrm{ind}(M)$. *The signature is also constant and so the manifold has a signature also. If the index of a semi-Riemannian manifold (with* $\dim(M) \geq 2$) *is* $(-1, +1, +1+1, ...)$ *(or according to some conventions* $(1, -1, -1-1, ...)$*) then the manifold is called a* ***Lorentz manifold*** .

The simplest family of semi-Riemannian manifolds are the spaces $\mathbb{R}^{n-v,v}$ which are the Euclidean spaces $\mathbb{R}^n$ endowed with the scalar products given by

$$\langle x, y\rangle_\nu = -\sum_{i=1}^{\nu} x^i y^i + \sum_{i=\nu+1}^{n} x^i y^i.$$

Since ordinary Euclidean geometry does not use indefinite scalar products we shall call the spaces $\mathbb{R}^{\nu,n-\nu}$ **semi-Euclidean** spaces when the index ν is not zero. If we write just $\mathbb{R}^n$ then either we are not concerned with a scalar product at all or the scalar product is assumed to be the usual inner product ($\nu = 0$).Thus a Riemannian metric is just the special case of index 0. Analogous to the groups $S\mathrm{O}(n), \mathrm{O}(n)$ and $\mathrm{Euc}(n)$ are we have, associated to $\mathbb{R}^{\nu,n-\nu}$, the groups $S\mathrm{O}(\nu, n-\nu), \mathrm{O}(\nu, n-\nu)$ (generalizing the Lorentz group) and $\mathrm{Euc}(\nu, n-\nu)$ (generalizing the Poincaré group):

$$\begin{aligned}
\mathrm{O}(\nu, n-\nu) &= \{Q \in End(\mathbb{R}^{\nu,n-\nu}) : \langle Qx, Qy\rangle_\nu = \langle x, y\rangle_\nu \text{ for all } x, y \in \mathbb{R}^{\nu,n-\nu}\} \\
S\mathrm{O}(\nu, n-\nu) &= \{Q \in \mathrm{O}(\nu, n-\nu) : \det Q = \pm 1 \text{ (preserves a choice of orientation)} \\
\mathrm{Euc}(\nu, n-\nu) &= \{L : L(x) = Qx + x_0 \text{ for some } Q \in \mathrm{O}(\nu, n-\nu) \text{ and } x_0 \in \mathbb{R}^{\nu,n-\nu}
\end{aligned}$$

The group $\mathrm{Euc}(\nu, n-\nu)$ is the group of semi-Euclidean motions and we have the expected homogeneous space isomorphism $\mathbb{R}^{\nu,n-\nu} = \mathrm{Euc}(\nu, n-\nu)/\mathrm{O}(\nu, n-\nu)$. To be consistent we should denote this homogeneous space by $\mathbf{E}^{\nu,n-\nu}$.

Remark 20.3 (Notation) *We will usually write $\langle X_p, Y_p\rangle$ or $\mathrm{g}(X_p, Y_p)$ in place of $\mathrm{g}(p)(X_p, X_p)$. Also, just as for any tensor field we define the function $\langle X, Y\rangle$ that, for a pair of vector fields X and Y, is given by $\langle X, Y\rangle(p) = \langle X_p, Y_p\rangle$.*

In local coordinates $(x^1, ..., x^n)$ on $U \subset M$ we have that $\mathrm{g}|_U = \sum \mathrm{g}_{ij} dx^i \otimes dx^j$ where $\mathrm{g}_{ij} = \langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\rangle$. Thus if $X = \sum X^i \frac{\partial}{\partial x^i}$ and $Y = \sum Y^i \frac{\partial}{\partial x^i}$ on U then

$$\langle X, Y\rangle = \sum \mathrm{g}_{ij} X^i Y^i \tag{20.3}$$

Remark 20.4 *The expression $\langle X, Y\rangle = \sum \mathrm{g}_{ij} X^i Y^i$ means that for all $p \in U$ we have $\langle X(p), Y(p)\rangle = \sum \mathrm{g}_{ij}(p) X^i(p) Y^i(p)$ where as we know that functions X^i and Y^i are given by $X^i = dx^i(X)$ and $Y^i = dx^i(Y)$.*

The definitions we gave for volume element, star operator and the musical isomorphisms are all defined on each tangent space T_pM of a semi-Riemannian manifold and the concepts globalize to the whole of TM accordingly. We have

1. Let M be oriented. The volume element induced by the metric tensor g is defined to be the $n-$form vol such that vol_p is the metric volume form on T_pM matching the orientation. If $U, \mathtt{x}$ is a chart on M then we have

$$vol_U = \sqrt{|\det(\mathrm{g}_{ij})|} f^1 \wedge \cdots \wedge f^n$$

If f is a smooth function we may integrate f over M by using the volume form:

$$\int_M f\, vol$$

The volume of a domain $D \subset M$ is defined as

$$vol(D) := \{\int_M f vol : \operatorname{supp}(f) \subset D \text{ and } 0 \leq f \leq 1\}$$

2. The musical isomorphisms are defined in the obvious way from the isomorphisms on each tangent space.

$$((\mathrm{g}^\sharp)^{\uparrow}_a \tau)(p) := (\mathrm{g}^\sharp(p))^{\uparrow}_a \tau(p)$$
$$\left((\mathrm{g}_\flat)^a_{\downarrow} \tau\right)(p) = (\mathrm{g}_\flat(p))^a_{\downarrow} \tau(p)$$

If we choose a local frame field F_i and its dual F^i on $U \subset M$ then in terms of the components of a tensor τ and the metric tensor with respect to this frame we have

$$(\mathrm{g}^\sharp)^{\uparrow}_a \tau_U = \tau^{i_1,\ldots,i_r j_a}_{j_1,\ldots,j_{a-1},\widehat{j_a},j_{a+1}\cdots,j_{s-1}} F_{i_1} \otimes \cdots \otimes F_{i_r} \otimes F_{j_a} \otimes F^{j_1} \otimes \cdots \otimes F^{j_{s-1}}.$$

where

$$\tau^{i_1,\ldots,i_r j}_{j_1,\ldots,j_{b-1},\widehat{k},j_b\cdots,j_{s-1}}(p)$$
$$= \tau^{\,i_1,\ldots,i_r}{}_{j_1,\ldots,j_{j-1}}{}^{j}{}_{j_{b+1}\cdots,j_{s-1}}(p) := \sum_m \mathrm{g}^{jm}(p) \tau^{i_1,\ldots,i_r}_{j_1,\ldots,j_{b-1},m,j_b\cdots,j_{s-1}}(p).$$

A similar formula holds for $(\mathrm{g}_\flat)^a_{\downarrow}\tau$ and we often suppress the dependence on the point p and the reference to the domain U.

3. Once again assume that M is oriented with metric volume form vol. The star operator gives two types of maps which are both denoted by $*$. Namely, the bundle maps $* : \bigwedge^k T^*M \to \bigwedge^{n-k} T^*M$ which has the obvious definition in terms of the star operators on each fiber and the maps on forms $* : \Omega^k(M) \to \Omega^{n-k}(M)$ which is also induce in the obvious way. The definitions are set up so that $*$ is natural with respect to restriction and so that for any oriented orthonormal frame field $\{E_1, ..., E_n\}$ with dual $\{\theta^1, ..., \theta^n\}$ we have $*(\theta^{i_1} \wedge \cdots \wedge \theta^{i_k}) = \pm(\varepsilon_{i_1}\varepsilon_{i_2} \cdots \varepsilon_{i_k})\theta^{j_1} \wedge \cdots \wedge \theta^{j_{n-k}}$ where as before we use the $+$ sign if and only if $\theta^{i_1} \wedge \cdots \wedge \theta^{i_k} \wedge \theta^{j_1} \wedge \cdots \wedge \theta^{j_{n-k}} = \theta^1 \wedge \cdots \wedge \theta^n = vol$

As expected we have $*1 = vol$, $*vol = (-1)^{\mathrm{ind}(\mathrm{g})}$ and $**\alpha = (-1)^{\mathrm{ind}(\mathrm{g})}(-1)^{k(n-k)}\alpha$ for $\alpha \in \Omega^k(M)$.

The inner products induced by $g(p)$ on $\bigwedge^k T_p^*M$ for each p combine to give a pairing $\langle .,. \rangle : \Omega^k(M) \times \Omega^k(M) \to C^\infty(M)$ with $\langle \alpha, \beta \rangle(p) := \langle \alpha(p), \beta(p) \rangle_p = \mathrm{g}_p(\alpha(p), \beta(p))$. We see that $\langle \alpha, \beta \rangle vol = \alpha \wedge *\beta$.

Now we can put an inner product on the space $\Omega(M) = \sum_k \Omega^k(M)$ by letting $(\alpha|\beta) = 0$ for $\alpha \in \Omega^{k_1}(M)$ and $\beta \in \Omega^{k_2}(M)$ with $k_1 \neq k_2$ and

$$(\alpha|\beta) = \int_M \alpha \wedge *\beta = \int_M \langle \alpha, \beta \rangle vol \ \text{ if } \alpha, \beta \in \Omega^k(M)$$

Definition 20.13 *Let M, g and N, h be two semi-Riemannian manifolds. A diffeomorphism $\Phi : M \to N$ is called an* ***isometry*** *if $\Phi^* \mathrm{h} = \mathrm{g}$. Thus for an isometry $\Phi : M \to N$ we have $\mathrm{g}(v, w) = \mathrm{h}(T\Phi \cdot v, T\Phi \cdot w)$ for all $v, w \in TM$. If $\Phi : M \to N$ is a local diffeomorphism such that $\Phi^* \mathrm{h} = \mathrm{g}$ then Φ is called a* ***local isometry.*** *If there is an isometry $\Phi : M \to N$ then we say that M, g and N, h are* ***isometric****.*

Definition 20.14 *Let $\widetilde{M}$ and M be semi-Riemannian manifolds. If $\wp : \widetilde{M} \to M$ is a covering map such that $\wp$ is a local isometry we call $\wp : \widetilde{M} \to M$ a* ***semi-Riemannian*** *covering.*

Definition 20.15 *The set of all isometries of a semi-Riemannian manifold M to itself is a group called the isometry group. It is denoted by $Isom(M)$.*

The isometry group of a generic manifold is most likely trivial but many examples of manifolds with relatively large isometry groups are easy to find using Lie group theory. This will be taken up at later point in the exposition.

Example 20.1 *We have seen that a regular submanifold of a Euclidean space $\mathbb{R}^n$ is a Riemannian manifold with the metric inherited from $\mathbb{R}^n$. In particular, the sphere $S^{n-1} \subset \mathbb{R}^n$ is a Riemannian manifold. Every isometry of S^{n-1} is the restriction to S^{n-1} of an isometry of $\mathbb{R}^n$ that fixed the origin (and consequently fixes S^{n-1}).*

If we have semi-Riemannian manifolds M, g and N, h then we can consider the product manifold $M \times N$ and the projections $pr_1 : M \times N \to M$ and $pr_2 : M \times N \to N$. The tensor $\mathrm{g} \times \mathrm{h} = pr_1^* \mathrm{g} + pr_2^* \mathrm{h}$ provides a semi-Riemannian metric on the manifold $M_1 \times M_2$ and $M_1 \times M_2, pr_1^* \mathrm{g} + pr_2^* \mathrm{h}$ is called the semi-Riemannian product of M, g and N, h. Let $U_1 \times U_2, (\mathtt{x}, \mathtt{y}) = (x^1, ..., x^{n_1}, y^1, ..., y^{n_2})$ be a natural product chart where we write x^i and y^i instead of the more pedantic $x^i \circ pr_1$ and $y^i \circ pr_2$. We then have coordinate frame fields $\frac{\partial}{\partial x^1}, ..., \frac{\partial}{\partial x^{n_1}}, \frac{\partial}{\partial y^1}, ..., \frac{\partial}{\partial y^{n_2}}$ and the components of $\mathrm{g} \times \mathrm{h} = pr_1^* \mathrm{g} + pr_2^* \mathrm{h}$ in this frame discovered by choosing a point $(p_1, p_2) \in U_1 \times U_2$ and then calculating. However, we will be less likely to be misled if we temporarily return to the notation introduced earlier where

$\widetilde{x}^i = x^i \circ pr_1, \widetilde{y}^i = y^i \circ pr_1$ and $\frac{\partial}{\partial \widetilde{x}^i}$ etc. We then have

$$
\begin{aligned}
&\mathrm{g} \times \mathrm{h}(\left.\frac{\partial}{\partial \widetilde{x}^i}\right|_{(p,q)}, \left.\frac{\partial}{\partial \widetilde{y}^j}\right|_{(p,q)}) \\
&= pr_1^* \mathrm{g}(\left.\frac{\partial}{\partial \widetilde{x}^i}\right|_{(p,q)}, \left.\frac{\partial}{\partial \widetilde{y}^j}\right|_{(p,q)}) + pr_2^* \mathrm{h}(\left.\frac{\partial}{\partial \widetilde{x}^i}\right|_{(p,q)}, \left.\frac{\partial}{\partial \widetilde{y}^j}\right|_{(p,q)}) \\
&= \mathrm{g}(Tpr_1 \left.\frac{\partial}{\partial \widetilde{x}^i}\right|_{(p,q)}, Tpr_1 \left.\frac{\partial}{\partial \widetilde{y}^j}\right|_{(p,q)}) + \mathrm{h}(Tpr_2 \left.\frac{\partial}{\partial \widetilde{x}^i}\right|_{(p,q)}, Tpr_2 \left.\frac{\partial}{\partial \widetilde{y}^j}\right|_{(p,q)}) \\
&= \mathrm{g}(\left.\frac{\partial}{\partial x^i}\right|_p, 0_p) + \mathrm{h}(0_q, \left.\frac{\partial}{\partial y^j}\right|_q) \\
&= 0 + 0 = 0
\end{aligned}
$$

and (abbreviating a bit)

$$
\begin{aligned}
\mathrm{g} \times \mathrm{h}(\left.\frac{\partial}{\partial \widetilde{x}^i}\right|_{(p,q)}, \left.\frac{\partial}{\partial \widetilde{x}^j}\right|_{(p,q)}) &= \mathrm{g}(\left.\frac{\partial}{\partial x^i}\right|_p, \left.\frac{\partial}{\partial x^i}\right|_p) + \mathrm{h}(0_q, 0_q) \\
&= \mathrm{g}_{ij}(p)
\end{aligned}
$$

Similarly $\mathrm{g} \times \mathrm{h}(\frac{\partial}{\partial y^i}, \frac{\partial}{\partial y^j})(p,q) = \mathrm{h}_{ij}(q)$. So with respect to the coordinates $(x^1, ..., x^{n_1}, y^1, ..., y^{n_2})$ the matrix of $\mathrm{g} \times \mathrm{h}$ is of the form

$$
\begin{pmatrix} (\mathrm{g}_{ij} \circ pr_1) & 0 \\ 0 & (\mathrm{h}_{ij} \circ pr_2) \end{pmatrix}.
$$

Forming semi-Riemannian product manifolds is but one way to get more examples of semi-Riemannian manifolds. Another way to get a variety of examples of semi-Riemannian manifolds is via a group action by isometries (assuming one is lucky enough to have a big supply of isometries). Let us here consider the case of a discrete group that acts smoothly, properly, freely and by isometries. We have already seen that if we have an action $\rho : G \times M \to M$ satisfying the first three conditions then the quotient space M/G has a unique structure as a smooth manifold such that the projection $\kappa : M \to M/G$ is a covering. Now since G acts by isometries $\rho_g^* \langle ., . \rangle = \langle ., . \rangle$ for all $g \in G$. The tangent map $T\kappa : TM \to T(M/G)$ is onto and so for any $\bar{v}_{\kappa(p)} \in T_{\kappa(p)}(M/G)$ there is a vector $v_p \in T_pM$ with $T_p\kappa \cdot v_p = \bar{v}_{\kappa(p)}$. In fact there is more than one such vector in T_pM (except in the trivial case $G = \{e\}$) but if $T_p\kappa \cdot v_p = T_q\kappa \cdot w_q$ then there is a $g \in G$ such that $\rho_g p = q$ and $T_p\rho_g v_p = w_q$. Conversely, if $\rho_g p = q$ then $T_p\kappa \cdot (T_p\rho_g v_p) = T_q\kappa \cdot w_q$. Now for $\bar{v}_1, \bar{v}_2 \in T_pM$ define $h(\bar{v}_1, \bar{v}_2) = \langle v_1, v_2 \rangle$ where v_1 and v_2 are chosen so that $T\kappa \cdot v_i = \bar{v}_i$. From our observations above this is well defined. Indeed, if $T_p\kappa \cdot v_i = T_q\kappa \cdot w_i = \bar{v}_i$ then there is an isometry ρ_g with $\rho_g p = q$ and $T_p\rho_g v_i = w_i$ and so

$$
\langle v_1, v_2 \rangle = \langle T_p\rho_g v_1, T_p\rho_g v_2 \rangle = \langle w_1, w_2 \rangle.
$$

It is easy to show that $x \mapsto h_x$ defined a metric on M/G with the same signature as that of $\langle ., . \rangle$ and $\kappa^* h = \langle ., . \rangle$. In fact we will use the same notation for either the metric on M/G or on M which is not such an act of violence since κ is now a local isometry.

If we have a local isometry $\phi : N \to M$ then any lift $\phi : N \to \widetilde{M}$ is also a local isometry. Deck transformations are lifts of the identify map $M \to M$ and so are diffeomorphisms which are local isometries. Thus deck transformations are in fact, isometries. We conclude that the group of deck transformations of a Riemannian cover is a subgroup of the group of isometries $Isom(\widetilde{M})$.

The simplest example of this construction is $\mathbb{R}^n/\Gamma$ for some lattice Γ and where we use the canonical Riemannian metric on $\mathbb{R}^n$. In case the lattice is isomorphic to $\mathbb{Z}^n$ then $\mathbb{R}^n/\Gamma$ is called a flat torus of dimension n. Now each of these tori are locally isometric but may not be globally so. To be more precise, suppose that $f_1, f_2, ..., f_n$ is a basis for $\mathbb{R}^n$ which is not necessarily orthonormal. Let Γ_f be the lattice consisting of integer linear combinations of $f_1, f_2, ..., f_n$. The question now is what if we have two such lattices Γ_f and $\Gamma_{\bar{f}}$ when is $\mathbb{R}^n/\Gamma_f$ isometric to $\mathbb{R}^n/\Gamma_{\bar{f}}$? Now it may seem that since these are clearly diffeomorphic and since they are locally isometric then they must be (globally) isometric. But this is not the case. We will be able to give a good reason for this shortly but for now we let the reader puzzle over this.

Every smooth manifold that admits partitions of unity also admits at least one (in fact infinitely may) Riemannian metrics. This includes all finite dimensional paracompact manifolds. The reason for this is that the set of all Riemannian metric tensors is, in an appropriate sense, convex. To wit:

Proposition 20.3 *Every smooth manifold that admits a smooth partition of unity admits a Riemannian metric.*

Proof. As in the proof of 20.11 above we can transfer the Euclidean metric onto the domain U_α of any given chart via the chart map ψ_α. The trick is to piece these together in a smooth way. For that we take a smooth partition of unity U_α, ρ_α subordinate to a cover by charts U_α, ψ_α. Let g_α be any metric on U_α and define

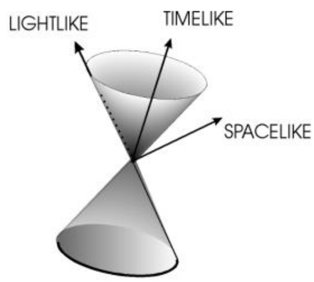
$$\mathrm{g}(p) = \sum \rho_\alpha(p) g_\alpha(p).$$

The sum is finite at each $p \in M$ since the partition of unity is locally finite and the functions $\rho_\alpha \mathrm{g}_\alpha$ are extended to be zero outside of the corresponding U_α. The fact that $\rho_\alpha \geq 0$ and $\rho_\alpha > 0$ at p for at least one α easily gives the result that g positive definite is a Riemannian metric on M. ■

The length of a tangent vector $X_p \in T_pM$ in a *Riemannian* manifold is given by $\sqrt{\mathrm{g}(X_p, X_p)} = \sqrt{\langle X_p, X_p \rangle}$. In the case of an indefinite metric ($\nu > 0$) we will need a classification:

Definition 20.16 *A tangent vector $\nu \in T_pM$ to a semi-Riemannian manifold M is called*

1. *spacelike if* $\langle \nu, \nu \rangle > 0$
2. *lightlike or null if*?? $\langle \nu, \nu \rangle = 0$



3. *timelike if* $\langle \nu, \nu \rangle < 0$.

4. *nonnull if* ν *is either timelike of spacelike.*

   The terms spacelike,lightlike, timelike indicate the ***causal character*** *of* v.

Definition 20.17 *The set of all timelike vectors* T_pM *in is called the* ***light cone*** *at* p.

Definition 20.18 *Let* $I \subset \mathbb{R}$ *be some interval. A curve* $c : I \to M, \mathrm{g}$ *is called spacelike, lightlike,timelike, or nonnull according as* $\dot{c}(t) \in T_{c(t)}M$ *is spacelike, lightlike, timelike, or nonnull respectively for all* $t \in I$.

For Lorentz spaces, that is for semi-Riemannian manifolds with index equal to 1 and dimension greater than or equal to 2, we may also classify subspaces into three categories:

Definition 20.19 *Let* M, g *be a Lorentz manifold. A subspace* $\mathrm{W} \subset T_pM$ *of the tangents space is called*

1. *spacelike if* $\mathrm{g}|_{\mathrm{W}}$ *is positive definite,*

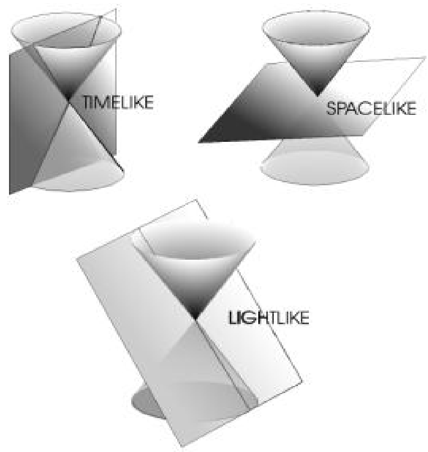
2. *time like if* $\mathrm{g}|_{\mathrm{W}}$ *nondegenerate with index* 1*,*

3. *lightlike if* $\mathrm{g}|_{\mathrm{W}}$ *is degenerate.*

Theorem 20.1 *A smooth manifold admits a an indefinite metric of index k if it the tangent bundle has some rank k subbundle.*

For the proof of this very plausible theorem see (add here)??

Recall that a continuous curve $c : [a, b] \to M$ into a smooth manifold is called **piecewise smooth** if there exists a partition $a = t_0 < t_1 < \cdots < t_k = b$ such that c restricted to $[t_i, t_{i+1}]$ is smooth for $0 \leq i \leq k - 1$. Also, a curve $c : [a, b] \to M$ is called **regular** if it has a nonzero tangent for all $t \in [a, b]$.



Definition 20.20 *Let M, g be Riemannian. If $c : [a, b] \to M$ is a (piecewise smooth) curve then the length of the curve from $c(a)$ to $c(b)$ is defined by*

$$L(c) := L_{c(a)\to c(b)}(c) := \int_a^t \langle \dot{c}(t), \dot{c}(t)\rangle^{1/2} dt. \tag{20.4}$$

Definition 20.21 *Let M, g be semi-Riemannian. If $c : [a, b] \to M$ is a (piecewise smooth) timelike or spacelike curve then*

$$L_{c(a),c(b)}(c) = \int_a^b |\langle \dot{c}(t), \dot{c}(t)\rangle|^{1/2}\, dt$$

is called the ***length*** *of the curve. For a general (piecewise smooth) curve $c : [a, b] \to M$, where M is a Lorentz manifold, the quantity*

$$\tau_{c(a),c(b)}(c) = \int_a^b \langle \dot{c}(t), \dot{c}(t)\rangle^{1/2} dt$$

will be called the ***proper time*** *of c.*

In general, if we wish to have a positive real number for a length then in the semi-Riemannian case we need to include absolute value signs in the definition so the proper time is just the timelike special case of a generalized arc length defined for any smooth curve by $\int_a^b |\langle \dot{c}(t), \dot{c}(t)\rangle|^{1/2}\, dt$ but unless the curve is either timelike or spacelike this arc length can have some properties that are decidedly not like our ordinary notion of length. In particular, curve may connect two different points and the generalized arc length might still be zero! It becomes clear that we are not going to be able to define a metric distance function as we soon will for the Riemannian case.

Definition 20.22 *A* ***positive reparametrization*** *of a piecewise smooth curve* $c : I \to M$ *is a curve defined by composition* $c \circ f^{-1} : J \to M$ *where* $f : I \to J$ *is a piecewise smooth bijection that has* $f' > 0$ *on each subinterval* $[t_{i-1}, t_i] \subset I$ *where* c *is smooth.*

Remark 20.5 (important fact) *The integrals above are well defined since* $c'(t)$ *is defined except for a finite number of points in* $[a, b]$. *Also, it is important to notice that by standard change of variable arguments a positive reparametrization* $\widetilde{c}(u) = c(f^{-1}(u))$ *where* $u = f(t)$ *does not change the (generalized) length of the curve*

$$\int_a^b |\langle \dot{c}(t), \dot{c}(t)\rangle|^{1/2} dt = \int_{f^{-1}(a)}^{f^{-1}(b)} |\langle \widetilde{c}'(u), \widetilde{c}'(u)\rangle|^{1/2} du.$$

Thus the (generalized) length of a piecewise smooth curve is a geometric property of the curve; i.e. a semi-Riemannian invariant.

20.3 Relativity and Electromagnetism

In this section we take short trip into physics. Lorentz manifolds of dimension 4 are the basis of Einstein's General Theory of Relativity. The geometry which is the infinitesimal model of all Lorentz manifolds and provides that background geometry for special relativity is Minkowski space. Minkowski space M^4 is $\mathbb{R}^{1,3}$ acted upon by the Poincaré group. More precisely, Minkowski space is the affine space modeled on $\mathbb{R}^{1,3}$ but we shall just take $\mathbb{R}^{1,3}$ itself and "forget" the preferred origin and coordinates as explained earlier. All coordinates related to the standard coordinate system by an element of the Poincaré group are to be put on equal footing. The reader will recall that these special coordinate systems are referred to as (Lorentz) inertial coordinates. Each tangent space of M^4 is a scalar product space canonically isomorphic $\mathbb{R}^{1,3}$. Of course, M^4 is a very special semi-Riemannian manifold. M^4 is flat (see below definition 20.31 below), simply connected and homogeneous. The Poincaré group is exactly the isometry group of M^4. We shall now gain some insight into Minkowski space by examining some of the physics that led to its discovery. Each component of the electric field associated with an instance of electromagnetic radiation in free space satisfies the wave equation $\Box f = 0$ where $\Box = -\partial_0^2 + \partial_1^2 + \partial_2^2 + \partial_3^2$ is the wave operator). The form of this equation already suggests a connection with the Lorentz metric on M^4. In fact, the groups $\mathrm{O}(1, 3)$ and $P = \mathrm{Euc}(1, 3)$ preserve the operator. On the other hand, things aren't quite so simple since it is the components of a vector that appears in the wave equation and those have their own transformation law. Instead of pursuing this line of reasoning let us take a look at the basic equations of electromagnetism; Maxwell's equations[1]:

[1]Actually, this is the form of Maxwell's equations after a certain convenient choice of units and we are ignoring the somewhat subtle distinction between the two types of electric fields E and D and the two types of magnetic fields B and H and also their relation in terms of dialectic constants.

$$\nabla \cdot \mathbf{B} = 0$$
$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0$$
$$\nabla \cdot \mathbf{E} = \varrho$$
$$\nabla \times \mathbf{B} - \frac{\partial \mathbf{E}}{\partial t} = \mathbf{j}$$

Here $\mathbf{E}$ and $\mathbf{B}$, the electric and magnetic fields are functions of space and time. We write $\mathbf{E} = \mathbf{E}(t, \mathbf{x})$, $\mathbf{B} = \mathbf{B}(t, \mathbf{x})$. The notation suggests that we have conceptually separated space and time as if we were stuck in the conceptual framework of the Galilean spacetime. Our purpose is to slowly discover how much better the theory becomes when we combine space and time in Minkowski spacetime.

Before continuing we will try to say a few words about the meaning of these equations since the readership may include a few who have not looked at these equations in a while. The electric field $\mathbf{E}$ is produced by the presence of charged particles. Under normal conditions a generic material is composed of a large number of atoms. To simplify matters we will think of the atom as being composed of just three types of particle; electrons, protons and neutrons. Protons carry a positive charge and electrons carry a negative charge and neutrons carry no charge. Normally, each atom will have a zero net charge since it will have an equal number of electrons and protons. It is a testament to the relative strength of electrical force that if a relatively small percent of the atoms in a material a stripped from their atoms and conducted away then there will be a net positive charge. In the vicinity of the material there will be an electric field the evidence of which is that if small "test particle" with a positive charge moves into the vicinity of the positively charged material it will be deflected, that is, accelerated away from what would otherwise presumably be a constant speed straight path. Let us assume for simplicity that the charged body which has the larger, positive charge is stationary at $\mathbf{r}_0$ with respect to a rigid rectangular coordinate system which is also stationary with respect to the laboratory. We must assume that our test particle carries is sufficiently small charge so that the electric field that it creates contribute negligibly to the field we are trying to detect (think of a single electron). Let the test particle be located at $\mathbf{r}$. Careful experiments show that when both charges are positive, the force experience by the test particle is directly away from the charged body located at $\mathbf{r}_0$ and has magnitude proportional to

$$qe/r^2$$

where $r = |\mathbf{r} - \mathbf{r}_0|$ is the distance between the charged body and the test particle, and where q and e are positive numbers which represent the amount of charge carried by the stationary body and the test particle respectively. If the units are chosen in an appropriate way we can say that the force F is given by

$$F = qe/r^2 = qe\frac{\mathbf{r} - \mathbf{r}_0}{|\mathbf{r} - \mathbf{r}_0|^3}.$$

By definition, the electric field at the location $\mathbf{r}$ of the test particle is

$$\mathbf{E} = q\frac{\mathbf{r} - \mathbf{r}_0}{|\mathbf{r} - \mathbf{r}_0|^3} \tag{20.5}$$

Now if the test particle has charge opposite to that of the source body then one of q or e will be negative an the force is directed toward the source. The direction of the electric field is completely determined by the sign of q and whether q is positive or negative the correct formula for $\mathbf{E}$ is still equation 20.5. The test particle could have been placed anywhere in space and so the electric field is implicitly defined at each point in space and so gives a vector field on $\mathbb{R}^3$. If the charge is modeled as a smoothly distributed charge density q which is nonzero in some region $U \subset \mathbb{R}^3$, then the total charge is given by integration $Q = \int_U q(t, \mathbf{r})dV_{\mathbf{r}}$ and the field at $\mathbf{r}$ is now given by $\mathbf{E}(t, \mathbf{r}) = \int_U q(t, \mathbf{y})\frac{\mathbf{r}-\mathbf{y}}{|\mathbf{r}-\mathbf{y}|^3}dV_{\mathbf{y}}$. Now since the source particle is stationary at $\mathbf{r}_0$ the electric field will be independent of time t. It turns out that a magnetic field is only produced by a changing electric field which would be produced, for example, if the source body was in motion. When charge is put into motion we have an electric current. For example, suppose a region of space is occupied by a mist of charged particles moving in such a way as to be essentially a fluid and so modeled by a smooth vector field $\mathbf{j}$ on $\mathbb{R}^3$ which in general also depend on time. Of course if the fluid of charge have velocity $\mathbf{V}(t, \mathbf{r})$ at $\mathbf{r}$ and the charge density is given by $\rho(t, \mathbf{y})$ then $\mathbf{j} = \mathbf{V}(t, \mathbf{r})\rho(t, \mathbf{y})$. But once again we can imagine a situation where the electric field does not depend of t. This would be the case even when the field is created by a moving fluid of charge as long at the charge density is constant in time. Even then, if the test particle is at rest in the laboratory frame (coordinate system) then there will be no evidence of a magnetic field felt by the test particle. In fact, suppose that there is a magnetic field $\mathbf{B}$ with an accompanying electric field $\mathbf{B}$ produced by a circulating charge density. Then the force felt by the test particle of charge e at $\mathbf{r}$ is $\mathbf{F} = e\mathbf{E} + \frac{e}{c}\mathbf{v} \times \mathbf{B}$ where $\mathbf{v}$ is the velocity of the test particle. The test particle has to be moving to feel the magnetic part of the field! This is strange already since even within the Galilean framework we would have expected the laws of physics to be the same in any inertial frame. But here we see that the magnetic field would disappear in any frame which is moving with the charge since in that frame the charge would have no velocity. At this point it is worth pointing out that from the point of view of spacetime, we are not staying true to the spirit of differential geometry since a vector field should have a geometric reality that is independent of its expression in an coordinate system. But here it seems that the fields just *are* those local representatives and so are dependent on the choice of spacetime coordinates; they seem to have no reality independent of the choice of inertial coordinate system. This is especially obvious for the magnetic field since a change to a new inertial system in spacetime (Galilean at this point) can effect a disappearance of the magnetic field. Let us ignore this problem for a while by just thinking of time as a parameter and sticking to one coordinate system.

Our next task is to write Maxwell's equations in terms of differential forms. We already have a way to convert (time dependent) vector fields $\mathbf{E}$ and $\mathbf{B}$ on

$\mathbb{R}^3$ into (time dependent) differential forms on $\mathbb{R}^3$. Namely, we use the flatting operation with respect to the standard metric on $\mathbb{R}^3$. For the electric field we have

$$\mathbf{E} = E^x\partial_x + E^y\partial_y + E^z\partial_z \mapsto \mathcal{E} = E_x dx + E_y dy + E_z dz$$

For the magnetic field we do something a bit different. Namely, we flat and then apply the star operator. In rectangular coordinates we have

$$\mathbf{B} = B^x\partial_x + B^y\partial_y + B^z\partial_z \mapsto \mathcal{B} = B_x dy \wedge dz + B_y dy \wedge dx + B_z dx \wedge dz$$

Now if we stick to rectangular coordinates (as we have been) the matrix of the standard metric is just $I = (\delta_{ij})$ and so we see that the above operations do not numerically change the components of the fields. Thus in any rectangular coordinates we have

$$\begin{aligned} E^x &= E_x \\ E^y &= E_y \\ E_z &= E_z \end{aligned}$$

and similarly for the B's. Now it is not hard to check that in the static case where $\mathcal{E}$ and $\mathcal{B}$ are time independent the first pair of (static) Maxwell's equations are equivalent to

$$d\mathcal{E} = 0 \text{ and } d\mathcal{B} = 0$$

This is pretty nice but if we put time dependent back into the picture we need to do a couple more things to get a nice viewpoint. So assume now that $\mathbf{E}$ and $\mathbf{B}$ and hence the forms $\mathcal{E}$ and $\mathcal{B}$ are time dependent and lets view these as differential forms on spacetime M^4 and in fact, ; let us combine $\mathcal{E}$ and $\mathcal{B}$ into a single 2−form on M^4 by letting

$$\mathcal{F} = \mathcal{B} + \mathcal{E} \wedge dt.$$

Since $\mathcal{F}$ is a 2−form it can be written in the form $\mathcal{F} = \frac{1}{2}F_{\mu\nu}dx^\mu \wedge dx^\nu$ where $F_{\mu\nu} = -F_{\nu\mu}$ and where the Greek indices are summed over $\{0, 1, 2, 3\}$. It is traditional in physics to let the Greek indices run over this set and to let Latin indices run over just the space indices $\{1, 2, 3, 4\}$. We will follow this convention for a while. Now if we compare $\mathcal{F} = \mathcal{B} + \mathcal{E} \wedge dt$ with $\frac{1}{2}F_{\mu\nu}dx^\mu \wedge dx^\nu$ we see that the $F_{\mu\nu}$'s form a antisymmetric matrix which is none other than

$$\begin{bmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & B_z & -B_y \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{bmatrix}$$

Our goal now is to show that the first pair of Maxwell's equations are equivalent to the single differential form equation

$$d\mathcal{F} = 0$$

Let N be ant manifold and let $M = (a,b)\times N$ for some interval (a,b). Let the coordinate on (a,b) be $t = x^0$ (time). Let $U, (x^1, ..., x^n)$ be a coordinate system on M. With the usual abuse of notation $(x^0, x^1, ..., x^n)$ is a coordinate system on $(a,b)\times N$. One can easily show that the local expression $d\omega = \partial_\mu f_{\mu_1 \dots \mu_k} \wedge dx^\mu \wedge dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_k}$ for the exterior derivative of a form $\omega = f_{\mu_1 \dots \mu_k} dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_k}$ can be written as

$$\begin{aligned} d\omega = {} & \sum_{i=1}^{3} \partial_i \omega_{\mu_1 \dots \mu_k} \wedge dx^i \wedge dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_k} \\ & + \partial_0 \omega_{\mu_1 \dots \mu_k} \wedge dx^0 \wedge dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_k} \end{aligned} \tag{20.6}$$

where the μ_i sum over $\{0, 1, 2, ..., n\}$. Thus we may consider spatial d_S part of the exterior derivative operator d on $(a,b)\times S = M$. To wit, we think of a given form ω on $(a,b)\times S$ as a time dependent for on N so that $d_S\omega$ is exactly the first term in the expression 20.6 above. Then we may write $d\omega = d_S\omega + dt \wedge \partial_t \omega$ as a compact version of the expression 20.6. The part $d_S\omega$ contains no dt's. Now we have by definition $\mathcal{F} = \mathcal{B} + \mathcal{E} \wedge dt$ on $\mathbb{R} \times \mathbb{R}^3 = M^4$ and so

$$\begin{aligned} d\mathcal{F} &= d\mathcal{B} + d(\mathcal{E} \wedge dt) \\ &= d_S\mathcal{B} + dt \wedge \partial_t \mathcal{B} + (d_S\mathcal{E} + dt \wedge \partial_t \mathcal{E}) \wedge dt \\ &= d_S\mathcal{B} + (\partial_t \mathcal{B} + d_S\mathcal{E}) \wedge dt. \end{aligned}$$

Now the $d_S\mathcal{B}$ is the spatial part and contains no dt's. It follows that $d\mathcal{F}$ is zero if and only if both $d_S\mathcal{B}$ and $\partial_t\mathcal{B} + d_S\mathcal{E}$ are zero. But unraveling the definitions shows that the pair of equations $d_S\mathcal{B} = 0$ and $\partial_t\mathcal{B} + d_S\mathcal{E} = 0$ (which we just showed to be equivalent to $d\mathcal{F} = 0$) are Maxwell's first two equations disguised in a new notation so in summary we have

$$d\mathcal{F} = 0 \quad \Longleftrightarrow \quad \begin{array}{c} d_S\mathcal{B} = 0 \\ \partial_t\mathcal{B} + d_S\mathcal{E} = 0 \end{array} \quad \Longleftrightarrow \quad \begin{array}{c} \nabla \cdot \mathbf{B} = 0 \\ \nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 \end{array}$$

Now we move on to rewrite the last pair of Maxwell's equations where the advantage of combining time and space together is Manifest to an even greater degree. One thing to notice about the about what we have done so far is the following. Suppose that the electric and magnetic fields were really all along most properly thought of a differential forms. Then we see that the equation $d\mathcal{F} = 0$ has nothing to do with the metric on Minkowski space at all. In fact, if $\phi : M^4 \to M^4$ is any diffeomorphism at all we have $d\mathcal{F} = 0$ if and only if $d(\phi^*\mathcal{F}) = 0$ and so the truth of the equation $d\mathcal{F} = 0$ is really a differential topological fact; a certain form is closed. The metric structure of Minkowski space is irrelevant. The same will not be true for the second pair. Even if we start out with the form $\mathcal{F}$ on spacetime it will turn out that the metric will necessarily be implicitly in the differential forms version of the second pair of Maxwell's equations. In fact, what we will show is that if we use the star operator for the Minkowski metric then the second pair can be rewritten as the single equation $*d*\mathcal{F} = *\mathcal{J}$ where $\mathcal{J}$ is formed from $\mathbf{j} = (j^1, j^2, j^3)$ and

ρ as follows: First we form the $4-$vector field $J = \rho\partial_t + j^1\partial_x + j^2\partial_y + j^3\partial_z$ (called the $4-$current) and then using the flatting operation we produce $\mathcal{J} = -\rho dt + j^1dx + j^2dy + j^3dz = J_0dt + J_1dx + J_2dy + J_3dz$ which is the covariant form of the $4-$current. We will only outline the passage from $*d*\mathcal{F} = *\mathcal{J}$ to the pair $\nabla \cdot \mathbf{E} = \varrho$ and $\nabla \times \mathbf{B} - \frac{\partial \mathbf{E}}{\partial t} = \mathbf{j}$. Let $*_S$be the operator one gets by viewing differential forms on M^4 as time dependent forms on $\mathbb{R}^3$ and then acting by the star operator with respect to the standard metric on $\mathbb{R}^3$. The first step is to verify that the pair $\nabla \cdot \mathbf{E} = \varrho$ and $\nabla \times \mathbf{B} - \frac{\partial \mathbf{E}}{\partial t} = \mathbf{j}$ is equivalent to the pair $*_Sd_S *_S \mathcal{E} = \varrho$ and $-\partial_t\mathcal{E} + *_Sd_S *_S \mathcal{B} = j$ where $j := j^1dx + j^2dy + j^3dz$ and $\mathcal{B}$ and $\mathcal{E}$ are as before. Next we verify that

$$*\mathcal{F} = *_S\mathcal{E} - *_S\mathcal{B} \wedge dt.$$

So the goal has become to get from $*d*\mathcal{F} = *\mathcal{J}$ to the pair $*_Sd_S *_S \mathcal{E} = \varrho$ and $-\partial_t\mathcal{E} + *_Sd_S *_S \mathcal{B} = j$. The following exercise finishes things off.

Exercise 20.3 *Show that* $*d*\mathcal{F} = -\partial_t\mathcal{E} - *_Sd_S *_S \mathcal{E} \wedge dt + *_Sd_S *_S \mathcal{B}$ *and then use this and what we did above to show that* $*d*\mathcal{F} = *\mathcal{J}$ *is equivalent to the pair* $*_Sd_S *_S \mathcal{E} = \varrho$ *and* $-\partial_t\mathcal{E} + *_Sd_S *_S \mathcal{B} = j$.

We have arrived at the pair of equations

$$d\mathcal{F} = 0$$
$$*d*\mathcal{F} = *\mathcal{J}$$

Now if we just think of this as a pair of equations to be satisfied by a $2-$form $\mathcal{F}$ where the $1-$form $\mathcal{J}$ is given then this last version of Maxwell's equations make sense on any semi-Riemannian manifold. In fact, on a Lorentz manifold that can be written as $(a,b)\times S = M$ with the twisted product metric $-ds^2 + \overset{3}{g}$ for some Riemannian metric on S then we can write our $2-$form $\mathcal{F}$ in the form $\mathcal{F} = \mathcal{B} + \mathcal{E} \wedge dt$ which allows us to identify the electric and magnetic fields in their covariant form.

The central idea in relativity is that the laws of physics should take on a simple and formally identical form when expressed in any Lorentz inertial frame. This is sometimes called covariance. We are now going to show how to interpret two of the four Maxwell's equations in their original form as conservation laws and then show that just these two conservation laws will produce the whole set of Maxwell's equations when combined with a simple assumption involving covariance.

Now it is time to be clear about how we are modeling spacetime. The fact that the Lorentz metric was part of the above procedure that unified and simplified Maxwell's equations suggests that the Minkowski spacetime is indeed the physically correct model.

We now discuss conservation in the context of electromagnetism and then show that we can arrive at the full set of 4 Maxwell's equations in their original form from combining the two simplest of Maxwell's equations with an reasonable assumption concerning Lorentz covariance.

20.4 Levi-Civita Connection

Let M, g be a semi-Riemannian manifold and ∇ a metric connection for M. Recall that the operator $T_\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ defined by $T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y]$ is a tensor called the torsion tensor of ∇. Combining the requirement that a connection be both a metric connection and torsion free pins down the metric completely.

Theorem 20.2 *For a given semi-Riemannian manifold M, g, there is a unique metric connection ∇ such that its torsion is zero; $T_\nabla \equiv 0$. This unique connection is called the **Levi-Civita derivative** for M, g.*

Proof. We will derive a formula that must be satisfied by ∇ and that can be used to actually define ∇. Let X, Y, Z, W be arbitrary vector fields on $U \subset M$. If ∇ exists as stated then on U we must have

$$\begin{aligned} X\langle Y,Z\rangle &= \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle \\ Y\langle Z,X\rangle &= \langle \nabla_Y Z, X\rangle + \langle Z, \nabla_Y X\rangle \\ Z\langle X,Y\rangle &= \langle \nabla_Z X, Y\rangle + \langle X, \nabla_Z Y\rangle. \end{aligned}$$

where we have written ∇^U simply as ∇. Now add the first two equations to the third one to get

$$\begin{aligned} &X\langle Y,Z\rangle + Y\langle Z,X\rangle - Z\langle X,Y\rangle \\ &= \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle + \langle \nabla_Y Z, X\rangle + \langle Z, \nabla_Y X\rangle \\ &- \langle \nabla_Z X, Y\rangle - \langle X, \nabla_Z Y\rangle. \end{aligned}$$

Now if we assume the torsion zero hypothesis then this reduces to

$$\begin{aligned} &X\langle Y,Z\rangle + Y\langle Z,X\rangle - Z\langle X,Y\rangle \\ &= \langle Y, [X,Z]\rangle + \langle X, [Y,Z]\rangle \\ &- \langle Z, [X,Y]\rangle + 2\langle \nabla_X Y, Z\rangle. \end{aligned}$$

Solving we see that $\nabla_X Y$ must satisfy

$$\begin{aligned} 2\langle \nabla_X Y, Z\rangle = X\langle Y,Z\rangle &+ Y\langle Z,X\rangle - Z\langle X,Y\rangle \\ &\langle Z, [X,Y]\rangle - \langle Y, [X,Z]\rangle - \langle X, [Y,Z]\rangle. \end{aligned}$$

Now since knowing $\langle \nabla_X Y, Z\rangle$ for all Z is tantamount to knowing $\nabla_X Y$ we conclude that if ∇ exists then it is unique. On the other hand, the patient reader can check that if we actually define $\langle \nabla_X Y, Z\rangle$ and hence $\nabla_X Y$ by this equation then all of the defining properties of a connection are satisfied and furthermore T_∇ will be zero. ■

It is not difficult to check that we may define a system of Christoffel symbols for the Levi-Civita derivative by the formula

$$\Gamma^\alpha(X_U, Y_U) := (\nabla_X Y)_U - DY_U \cdot X_U$$

where X_U, Y_U and $(\nabla_X Y)_U$ are the principal representatives of X, Y and $\nabla_X Y$ respectively for a given chart $U, \mathtt{x}$. Let M be a semi-Riemannian manifold of dimension n and let $U, \mathtt{x} = (x^1, ..., x^n)$ be a chart. Then we have the formula

$$\Gamma^k_{ij} = \frac{1}{2} g^{kl} \left(\frac{\partial g_{jl}}{\partial x^i} + \frac{\partial g_{li}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^l} \right).$$

where $g_{jk} g^{ki} = \delta^i_j$.

Let $F : N \to M$ be a smooth map. A **vector field along** F is a map $Z : N \to TM$ such that the following diagram commutes:

$$\begin{array}{ccc} & & TM \\ Z & \nearrow & \downarrow^{\tau_M} \\ N & \xrightarrow{F} & M \end{array}$$

We denote the set of all smooth vector fields along a map F by $\mathfrak{X}_F$. Let $F : N \to M$ be a smooth map and let the model spaces of M and N be M and N respectively.

We shall say that a pair of charts Let $U, \mathtt{x}$ be a chart on M and let $V, \mathtt{u}$ be a chart on N such that $\mathtt{u}(V) \subset U$. We shall say that such a pair of charts is adapted to the map F.

Assume that there exists a system of Christoffel symbols on M. We may define a covariant derivative $\nabla_X Z$ of a vector field Z along F with respect to a field $X \in \mathfrak{X}(N)$ by giving its principal representation with respect to any pair of charts adapted to F. Then $\nabla_X Z$ will itself be a vector fields along F. The map F has a local representation $F_{V,U} : V \to \psi(U)$ defined by $F_{V,U} := \mathtt{x} \circ F \circ \mathtt{u}^{-1}$. Similarly the principal representation $\mathsf{Z} : \mathtt{u}(V) \to \mathsf{M}$ of Z is given by $T\mathtt{x} \circ Z \circ \mathtt{u}^{-1}$ followed by projection onto the second factor of $\mathtt{x}(U) \times \mathsf{M}$. Now given any vector field $X \in \mathfrak{X}(N)$ with principal representation $\mathsf{X} : \psi(U) \to \mathsf{N}$ we define the covariant derivative $\nabla_X Z$ of Z with respect to X as that vector field along F whose principal representation with respect to any arbitrary pair of charts adapted to F is

$$D\mathsf{Z}(u) \cdot \mathsf{X}(u) + \Gamma(F_{V,U}(u))(DF_{V,U}(u) \cdot \mathsf{X}(u), \mathsf{Z}(u)).$$

In traditional notation if $Z = Z^i \frac{\partial}{\partial x^i} \circ F$ and $X = X^r \frac{\partial}{\partial u^r}$

$$(\nabla_X Z)^i = \frac{\partial Z^i}{\partial u^j} X^j + \Gamma^i_{jk} \frac{\partial F^j_{V,U}}{\partial u^r} X^r Z^k$$

The resulting map $\nabla : \mathfrak{X}(N) \times \mathfrak{X}_F \to \mathfrak{X}(N)$ has the following properties:

1. $\nabla : \mathfrak{X}(N) \times \mathfrak{X}_F \to \mathfrak{X}(N)$ is $C^\infty(N)$ linear in the first argument.

2. For the second argument we have

$$\nabla_X (fZ) = f \nabla_X Z + X(f) Z$$

for all $f \in C^\infty(N)$.

3. If Z happens to be of the form $Y \circ F$ for some $Y \in \mathfrak{X}(M)$ then we have

$$\nabla_X (Y \circ F)(p) = (\nabla_{TF \cdot X(p)} Y)(F(p)).$$

4. $(\nabla_X Z)(p)$ depends only on the value of X at $p \in N$ and we write $(\nabla_X Z)(p) = \nabla_{X_p} Z$.

For a curve $c :: \mathbb{R} \to M$ and $Z :: \mathbb{R} \to TM$ we define

$$\frac{\nabla Z}{dt} := \nabla_{d/dt} Z \in \mathfrak{X}_F$$

If Z happens to be of the form $Y \circ c$ then we have the following alternative notations with varying degrees of precision:

$$\nabla_{d/dt}(Y \circ c) = \nabla_{\dot{c}(t)} Y = \nabla_{d/dt} Y = \frac{\nabla Y}{dt}$$

Exercise 20.4 *Show that if $\alpha : I \to M$ is a curve and X, Y vector fields along α then $\frac{d}{dt}\langle X, Y\rangle = \langle \frac{\nabla}{dt} X, Y\rangle + \langle X, \frac{\nabla}{dt} Y\rangle$.*

Exercise 20.5 *Show that if $h : (a,b) \times (c,d) \to M$ is smooth then $\partial h/\partial t$ and $\partial h/\partial s$ are vector fields along h we have $\nabla_{\partial h/\partial t} \partial h/\partial s = \nabla_{\partial h/\partial s} \partial h/\partial t$. (Hint: Use local coordinates and the fact that ∇ is torsion free).*

20.5 Geodesics

In this section I will denote an interval (usually containing 0). The interval may be closed, open or half open. I may be an infinite or "half infinite" such as $(0, \infty)$. Recall that if I contains an endpoint then a curve $\gamma : I \to M$ is said to be smooth if there is a slightly larger open interval containing I to which the curve can be extended to a smooth curve.

Let M, g be a semi-Riemannian manifold. Suppose that $\gamma : I \to M$ is a smooth curve that is self parallel in the sense that

$$\nabla_{\dot{\gamma}} \dot{\gamma} = 0$$

along γ. We call γ a **geodesic**. More precisely, $\dot{\gamma}$ is vector field along γ and γ is a geodesic if $\dot{\gamma}$ is parallel: $\frac{\nabla \dot{\gamma}}{dt}(t) = 0$ for all $t \in I$. If M is finite dimensional, $\dim M = n$,and $\gamma(I) \subset U$ for a chart $U, \mathtt{x} = (x^1, ... x^n)$ then the condition for γ to be a geodesic is

$$\frac{d^2 x^i \circ \gamma}{dt^2}(t) + \Gamma^i_{jk}(\gamma(t)) \frac{dx^j \circ \gamma}{dt}(t) \frac{dx^k \circ \gamma}{dt}(t) = 0 \text{ for all } t \in I \text{ and } 1 \leq i \leq n. \tag{20.7}$$

and this system of equations is (thankfully) abbreviated to $\frac{d^2 x^i}{dt^2} + \Gamma^i_{jk} \frac{dx^j}{dt} \frac{dx^k}{dt} = 0$. These are the local geodesic equations. Now if γ is a curve whose image is not necessarily contained in the domain then by continuity we can say that for every

$t_0 \in \not I$ there is an $\epsilon > 0$ such that $\gamma|_{(t_0-\epsilon,t+\epsilon)}$ is contained in the domain of a chart. Then it is not hard to see that γ is a geodesic if each such restriction satisfies the corresponding local geodesic equations as in 20.7 (for each chart which meets the image of γ). We can convert the local geodesic equations 20.7 , which is a system of n-first order equations, into a system of $2n$ first order equations by the usual reduction of order trick. We let v denote a new dependent variable and then we get

$$\frac{dx^i}{dt} = v^i \ , \ 1 \leq i \leq n$$

$$\frac{dv^i}{dt} + \Gamma^i_{jk} v^j v^k = 0 \ , \ 1 \leq i \leq n.$$

We can think of x^i and v^i as coordinates on TM. Once we do this we recognize that the first order system above is the local expression of the equations for the integral curves of vector field on TM.

To be precise, one should distinguish various cases as follows: If $\gamma : I \to M$ is a geodesic we refer to it as a (parameterized) **closed geodesic segment** if $I = [a, b]$ for some finite numbers $a, b \in \mathbb{R}$. In case $I = [a, \infty)$ (resp. $(\infty, a]$) we call γ a positive (resp. negative) **geodesic ray**. If $I = (-\infty, \infty)$ then we call γ a **complete geodesic**.

So far, and for the most part from here on, this all makes sense for infinite dimensional manifolds modeled on a separable Hilbert space. We will write all our local expressions using the standard notation for finite dimensional manifolds. However, we will point out that the local equations for a geodesic can be written in notation appropriate to the infinite dimensional case as follows:

$$\frac{d^2 \mathrm{x} \circ \gamma}{dt^2}(t) + \Gamma(\gamma(t), \frac{d\mathrm{x} \circ \gamma}{dt}(t) \frac{d\mathrm{x} \circ \gamma}{dt}(t)) = 0$$

Exercise 20.6 *Show that there is a vector field $G \in \mathfrak{X}(TM)$ such that α is an integral curve of G if and only if $\gamma := \pi_{TM} \circ \alpha$ is a geodesic. Show that the local expression for G is*

$$v^i \frac{\partial}{\partial x^i} + \Gamma^i_{jk} v^j v^k \frac{\partial}{\partial v^i}$$

(The Einstein summation convention is in force here).

The vector field G from this exercise is an example of a spray (see problems 1). The flow if G in the manifold TM is called the **geodesic flow**.

Lemma 20.3 *If $v \in T_pM$ then there is a unique geodesic γ_v such that $\dot{\gamma}_v(0) = v$.*

Proof. This follows from standard existence and uniqueness results from differential equations. One may also deduce this result from the facts about flows since as the exercise above shows, geodesics are projections of integral curves of the vector field G. The reader who did not do the problems on sprays in 1 would do well to look at those problems before going on. ■

Lemma 20.4 *Let γ_1 and γ_2 be geodesics $I \to M$. If $\dot{\gamma}_1(t_0) = \dot{\gamma}_2(t_0)$ for some $t_0 \in I$, then $\gamma_1 = \gamma_2$.*

Proof. If not there must be $t' \in I$ such that $\gamma_1(t') \neq \gamma_2(t')$. Let us assume that $t' > a$ since the proof of the other cases is similar. The set $\{t \in I : t > t_0$ and $\gamma_1(t) \neq \gamma_2(t)\}$has a greatest lower bound $b \geq t_0$. Claim $\dot{\gamma}_1(b) = \dot{\gamma}_2(b)$: Indeed if $b = t_0$ then thus is assumed in the hypothesis and if $b > t_0$ then $\dot{\gamma}_1(t) = \dot{\gamma}_2(t)$ on the interval (t_0, b). By continuity

$$\dot{\gamma}_1(t_0) = \dot{\gamma}_2(t_0)$$

and since $\gamma_1(t_0 + t)$ and $\gamma_2(t_0 + t)$ are clearly geodesics with initial velocity $\dot{\gamma}_1(t_0) = \dot{\gamma}_2(t_0)$ the result follows from lemma 20.3. ■

A geodesic $\gamma : I \to M$ is called maximal if there is no other geodesic with domain J strictly larger $I \subsetneq J$ which agrees with γ in I.

Theorem 20.3 *For any $v \in TM$ there is a unique maximal geodesic γ_v with $\dot{\gamma}_v(0) = v$.*

Proof. Take the class $\mathcal{G}_v$ of all geodesics with initial velocity v. This is not empty by 20.3. If $\alpha, \beta \in \mathcal{G}_v$ and the respective domains I_α and I_β have nonempty intersection then α and β agree on this intersection by 20.4. From this we see that the geodesics in $\mathcal{G}_v$ fit together to form a manifestly maximal geodesic with domain $I = \cup_{\gamma \in \mathcal{G}_v} I_\gamma$.

It is clear this geodesic has initial velocity v. ■

Definition 20.23 *If the domain of every maximal geodesic emanating from a point $p \in T_pM$ is all of R then we say that M is geodesically complete at p. A semi-Riemannian manifold is said to be* ***geodesically complete*** *if and only if it is geodesically complete at each of its points.*

Exercise 20.7 *Let $\mathbb{R}^{n-\nu,\nu}, \eta$ be the semi-Euclidean space of index ν. Show that all geodesics are of the form $t \mapsto x_0 + tv$ for $v \in \mathbb{R}^{n-\nu,\nu}$. (don't confuse v with the index ν (a Greek letter).*

Exercise 20.8 *Show that if γ is a geodesic then a reparametrization $\gamma := \gamma \circ f$ is a geodesic if and only if $f(t) := at + b$ for some $a, b \in \mathbb{R}$ and $a \neq 0$.*

The existence of geodesics emanating from a point $p \in M$ in all possible directions and with all possible speeds allows us to define a very important map called the exponential map $\exp_p$. The domain of $\exp_p$ is the set $\widetilde{\mathcal{D}}_p$ of all $v \in T_pM$ such that the geodesic γ_v is defined at least on the interval $[0, 1]$. The map is define simply by $\exp_p v := \gamma_v(1)$. The map $s \mapsto \gamma_v(ts)$ is a geodesic with initial velocity $t\dot{\gamma}_v(0) = tv$ and the properties of geodesics that we have developed imply that $\exp_p(tv) = \gamma_{tv}(1) = \gamma_v(t)$. Now we have a very convenient situation. The geodesic through p with initial velocity v can always be written in the form $t \mapsto \exp_p tv$. Straight lines through $0_p \in T_pM$ are mapped by $\exp_p$ onto geodesics.

The exponential map is the bridge to comparing manifolds with each other as we shall see and provides a two types of special coordinates around any given p. The basic theorem is the following:

Theorem 20.4 *Let M, g be a Riemannian manifold and $p \in M$. There exists an open neighborhood $\widetilde{U}_p \subset \widetilde{\mathcal{D}}_p$ containing 0_p such that $\exp_p|_{\widetilde{U}_p} : \widetilde{U}_p \to \exp_p|_{\widetilde{U}_p}(\widetilde{U}_p) := U_p$ is a diffeomorphism.*

Proof. The tangent space T_pM is a vector space which is isomorphic to $\mathbb{R}^n$ (or a Hilbert space) and so has a standard differentiable structure. It is easy to see using the results about smooth dependence on initial conditions for differential equations that $\exp_p$ is well defined and smooth in some neighborhood of $0_p \in T_pM$. The main point is that the tangent map $T\exp_p : T_{0_p}(T_pM) \to T_pM$ is an isomorphism and so the inverse function theorem gives the result. To see that $T\exp_p$ is an isomorphism let $v_{0_p} \in T_{0_p}(T_pM)$ be the velocity of the curve $t \mapsto tv$ in T_pM. Under the canonically identification of $T_{0_p}(T_pM)$ with T_pM this velocity is just v. No unraveling definitions we have $v = v_{0_p} \mapsto T\exp_p \cdot v_{0_p} = \frac{d}{dt}_0 \exp_p \cdot tv = v$ so with this canonically identification $T\exp_p v = v$ so the tangent map is essentially the identity map. ■

Notation 20.2 *We will let the image set $\exp_p \widetilde{\mathcal{D}}_p$ be denoted by $\mathcal{D}_p$, the image $\exp_p(\widetilde{U}_p)$ is denoted by U_p.*

Definition 20.24 *An subset of a vector space* V *which contains* 0 *is called starshaped about* 0 *if $v \in \mathrm{V}$ implies $tv \in \mathrm{V}$ for all $t \in [0, 1]$.*

Definition 20.25 *If $\widetilde{U} \subset \widetilde{\mathcal{D}}_p$ is a starshaped open set about 0_p in T_pM such that $\exp_p|_{\widetilde{U}_p}$ is a diffeomorphism as in the theorem above then the image $\exp_p(\widetilde{U}) = U$ is called a* ***normal neighborhood of*** p. *(By convention such a set U is referred to as a starshaped also.)*

Theorem 20.5 *If $U \subset M$ is a normal neighborhood about p then for every point $q \in U$ there is a* unique *geodesic $\gamma : [0, 1] \to M$ such that $\gamma(0) = p$, $\gamma(1) = q$ and $\exp_p \dot{\gamma}(0) = q$.*

Proof. Let $p \in U$ and $\exp_p \widetilde{U} = U$. By assumption $\widetilde{U}$ is starshaped and so $\rho : t \mapsto tv$, $t \in [0, 1]$ has image in $\widetilde{U}$ and then the geodesic segment $\gamma : t \mapsto \exp_p tv$, $t \in [0, 1]$ has its image inside U. Clearly, $\gamma(0) = p$ and $\gamma(1) = q$. Since $\dot{\rho} = v$ we get

$$\dot{\gamma}(0) = T\exp_p \cdot \dot{\rho}(0) = T\exp_p \cdot v = v$$

under the usual identifications in T_pM.

Now assume that $\gamma_1 : [0, 1] \to M$ is some geodesic with $\gamma_1(0) = p$ and $\gamma_1(1) = q$. If $\dot{\gamma}_1(0) = w$ then $\gamma_1(t) = \exp_p \cdot tv$. We know that $\rho_1 : t \mapsto tw$, ($t \in [0, 1]$) is a ray that stays inside $\widetilde{U}$ and so $w = \rho_1(1) \in \widetilde{U}$. Also, $\exp_p w = \gamma_1(1) = q = \exp_p v$ and since $\exp_p|_{\widetilde{U}}$ is a diffeomorphism and hence 1-1 we conclude that $w = v$. Thus by the uniqueness theorems of differential equations the geodesics segments γ and γ_1 are both given by : $t \mapsto \exp_p tv$ and hence they are equal. ■

Definition 20.26 *A continuous curve $\gamma : I \to M$ is called **piecewise geodesic, broken geodesic** or even "**kinky geodesic**" if it is a piecewise smooth curve whose smooth segments are geodesics. If $t \in I$ (not an endpoint) is a point where γ is not smooth we call t or its image $\gamma(t)$ a **kink point**.*

Exercise 20.9 *Prove the following proposition:*

Proposition 20.4 *A semi-Riemannian manifold is connected if and only if every pair of its points can be joined by a broken geodesic $\gamma : [a, b] \to M$.*

We can gather the maps $\exp_p : \widetilde{\mathcal{D}}_p \subset T_pM \to M$ together to get a map $\exp : \widetilde{\mathcal{D}} \to M$ defined by $\exp(v) := \exp_{\pi(v)}(v)$ and where $\widetilde{\mathcal{D}} := \cup_p \widetilde{\mathcal{D}}_p$. To see this let $W \subset \mathbb{R} \times M$ be the domain of the geodesic flow $(s, v) \mapsto \gamma_v'(s)$. This is the flow of a vector field on TM and W is open. W is also the domain of the map $(s, v) \mapsto \pi \circ \gamma_v'(s) = \gamma_v(s)$. Now the map $(1, v) \mapsto \mathrm{Fl}_1^G(v) = \gamma_v(1) = v$ is a diffeomorphism and under this diffeomorphism $\widetilde{\mathcal{D}}$ corresponds to the set $W \cap (\{1\} \times TM)$ so must be open in TM. It also follows that $\widetilde{\mathcal{D}}_p$ is open in T_pM.

Definition 20.27 *A an open subset U of a semi-Riemannian manifold is **convex** if it is a normal neighborhood of each of its points.*

Notice that U being convex according to the above definition implies that for any two point p and q in U there is a unique geodesic segment $\gamma : [0, 1] \to U$ (image inside U) such that $\gamma(0) = p$ and $\gamma(1) = q$. Thinking about the sphere makes it clear that even if U is convex there may be geodesic segments connecting p and q whose images do not lie in U.

Lemma 20.5 *For each $p \in M$, $\mathcal{D}_p := \exp_p(\widetilde{\mathcal{D}}_p)$ is starshaped*

Proof. By definition $\mathcal{D}_p$ is starshaped if and only if $\widetilde{\mathcal{D}}_p$is starshaped. For $v \in \widetilde{\mathcal{D}}_p$ then γ_v is defined for all $t \in [0, 1]$. On the other hand, $\gamma_{tv}(1) = \gamma_v(t)$ and so $tv \in \widetilde{\mathcal{D}}_p$. ■

Now we take one more step and use the exponential map to get a map EXP from $\widetilde{\mathcal{D}}$ onto a subset containing the diagonal in $M \times M$. The diagonal is the set $\{(p, p) : p \in M\}$. The definition is simply $EXP : v \mapsto (\pi_{TM}(v), \exp_p v)$.

Theorem 20.6 *There is an open set $O \subset \widetilde{\mathcal{D}}$ which is mapped by EXP onto an open neighborhood of the diagonal $\triangle \subset M \times M$ and which is a diffeomorphism when restricted to a sufficiently small neighborhood of any point on $\triangle$.*

Proof. By the inverse mapping theorem and what we have shown about exp we only need to show that if $T_x \exp_p$ is nonsingular for some $x \in \widetilde{\mathcal{D}}_p \subset T_pM$ then T_xEXP is also nonsingular at x. So assume that $T_x \exp_p(v_x) = 0$ and, with an eye towards contradiction, suppose that $T_xEXP(v_x) = 0$. We have $\pi_{TM} = pr_1 \circ EXP$ and so $Tpr_1(v_x) = T\pi_{TM}(T_xEXP(v_x)) = 0$. This means that v_x is vertical (tangent to T_pM). On the other hand, it is easy to see that the following diagram commutes:

$$\begin{array}{ccc} T_pM & \overset{EXP_{T_pM}}{\rightarrow} & \{p\} \times M \\ id\downarrow & & pr_2\downarrow \\ T_pM & \overset{\exp_p}{\rightarrow} & M \end{array}$$

and hence so does

$$\begin{array}{ccc} T_x\left(T_pM\right) & \overset{T_xEXP_{T_pM}}{\rightarrow} & \{p\} \times M \\ id\downarrow & & pr_2\downarrow \\ T_x\left(T_pM\right) & \overset{T_x\exp_p}{\rightarrow} & M \end{array}$$

This implies that $T_x \exp_p(v) = 0$ and hence $v = 0$. ∎

Theorem 20.7 *Every $p \in M$ has a convex neighborhood.*

Proof. Let $p \in M$ and choose a neighborhood W of 0_p in TM such that $EXP|_W$ is a diffeomorphism onto a neighborhood of $(p,p) \in M \times M$. By a simple continuity argument we may assume that $EXP|_W(W)$ is of the form $U(\delta) \times U(\delta)$ for $U(\delta) := \{q : \sum_{i=1}^n (x^i(q))^2 < \delta\}$ and $\mathtt{x} = (x^i)$ some normal coordinate system. Now consider the tensor b on $U(\delta)$ whose components with respect to $\mathtt{x}$ are $b_{ij} = \delta_{ij} - \sum_k \Gamma_{ij}^k x^k$. This is clearly symmetric and positive definite at p and so by choosing δ smaller in necessary we may assume that this tensor is positive definite on $U(\delta)$. Let us show that $U(\delta)$ is a normal neighborhood of each of its points q. Let $W_q := W \cap T_qM$. We know that $EXP|_{W_q}$ is a diffeomorphism onto $\{q\} \times U(\delta)$ and it is easy to see that this means that $\left.\exp_q\right|_{W_q}$ is a diffeomorphism onto $U(\delta)$. We now show that W_q is star shaped about 0_q. Let $q' \in U(\delta)$, $q \neq q'$ and $v = EXP|_{W_q}^{-1}(q,q')$. This means that $\gamma_v : [0,1] \to M$ is a geodesic from q to q'. If $\gamma_v([0,1]) \subset U(\delta)$ then $tv \in W_q$ for all $t \in [0,1]$ and so we could conclude that W_q is starshaped. Let us assume that $\gamma_v([0,1])$ is not contained in $U(\delta)$ and work for a contradiction.

If in fact γ_v leaves $U(\delta)$ then the function $f : t \mapsto \sum_{i=1}^n (x^i(\gamma_v(t)))^2$ has a maximum at some $t_0 \in (0,1)$. We have

$$\frac{d^2}{dt^2} f = 2\sum_{i=1}^n \left(\frac{d\left(x^i \circ \gamma_v\right)}{dt} + x^i \circ \gamma_v \frac{d^2\left(x^i \circ \gamma_v\right)}{dt^2} \right).$$

But γ_v is a geodesic and so using the geodesic equations we get

$$\frac{d^2}{dt^2} f = 2\sum_{i,j} \left(\delta_{ij} - \sum_k \Gamma_{ij}^k x^k \right) \frac{d\left(x^i \circ \gamma_v\right)}{dt} \frac{d\left(x^i \circ \gamma_v\right)}{dt}.$$

Plugging in t_0 we get

$$\frac{d^2}{dt^2}f(t_0) = 2b(\gamma_v'(t_0), \gamma_v'(t_0)) > 0$$

which contradicts f having a maximum at t_0. ■

20.5.1 Normal coordinates

Let $M.g$ be finite a semi-Riemannian manifold of dimension n. Let and arbitrary $p \in M$ and pick any orthonormal basis $\{e_1, ..., e_n\}$ for the (semi-)Euclidean scalar product space $T_pM, \langle , \rangle_p$. Now this basis induces an isometry $I : \mathbb{R}^n_v \to T_pM$ by $(x^i) \mapsto x^i e_i$. Now if U is a normal neighborhood containing $p \in M$ then $\mathtt{x}_{Norm} := I \circ \exp_p|_{\tilde{U}}^{-1} : U \to \mathbb{R}^n_v = \mathbb{R}^n$ is a coordinate chart with domain U. These coordinates are referred to as normal coordinates centered at p. Normal coordinates have some very nice properties:

Theorem 20.8 *If $\mathtt{x}_{Norm} = (x^1, ..., x^n)$ are normal coordinates centered at $p \in U \subset M$ then*

$$g_{ij}(p) = \langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \rangle_p = \varepsilon_i \delta_{ij} \ (= \eta_{ij}) \text{ for all } i, j$$
$$\Gamma^i_{jk}(p) = 0 \text{ for all } i, j, k.$$

Proof. Let $v :=\in T_pM$ and let $\{e^i\}$ be the basis of T^*_pM dual to $\{e_i\}$. Then $e^i \circ \exp_p = x^i$. Now $\gamma_v(t) = \exp_p tv$ and so

$$x^i(\gamma_v(t)) = e^i(tv) = te^i(v) = ta^i$$

and so $a^i e_i = v = a^i \left.\frac{\partial}{\partial x^i}\right|_p$. In particular, if $a^i = \delta^i_j$ then $e_i = \left.\frac{\partial}{\partial x^i}\right|_p$ and so the coordinate vectors $\left.\frac{\partial}{\partial x^i}\right|_p \in T_pM$ are orthonormal; $\langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \rangle_p = \varepsilon_i \delta_{ij}$.

Now since γ_v is a geodesic and $x^i(\gamma_v(t)) = ta^i$ the coordinate expression for the geodesic equations reduces to $\Gamma^i_{jk}(\gamma_v(t))a^j a^k = 0$ for all i and this hold in particular at $p = \gamma_v(0)$. But (a^i) is arbitrary and so the quadratic form defined on R^n by $Q^k(\vec{a}) = \Gamma^i_{jk}(p)a^j a^k$ is identically zero an by polarization the bilinear form $Q^k : (\vec{a}, \vec{b}) \mapsto \Gamma^i_{jk}(p)a^j b^k$ is identically zero. Of course this means that $\Gamma^i_{jk}(p) = 0$ for all i, j and arbitrary k. ■

Exercise 20.10 *Sometimes the simplest situation becomes confusing because it is so special. For example, the identification $T_v(T_pM)$ with T_pM can have a confusing psychological effect. For practice with the "unbearably simple" determine $\exp_p$ for $p \in \mathbb{R}^{n-v,v}$ and find normal coordinates about p.*

Theorem 20.9 (Gauss Lemma) *Assume $0 < a < b$. Let $h : [a, b] \times (-\varepsilon, \varepsilon) \to M$ have the following properties:*

(i) $s \mapsto h(s, t)$ is a geodesic for all $t \in (-\varepsilon, \varepsilon)$;

(ii) $\langle \frac{\partial h}{\partial s}, \frac{\partial h}{\partial s} \rangle$ *is constant along each geodesic curve* $h_t : s \mapsto h(s,t)$.

Then $\langle \frac{\partial h}{\partial s}, \frac{\partial h}{\partial t} \rangle$ *is constant along each curve* h_t.

Proof. By definition we have

$$\frac{\partial h}{\partial s}(s,t) := T_{(s,t)}h \cdot \frac{\partial}{\partial s}$$
$$\frac{\partial h}{\partial t}(s,t) := T_{(s,t)}h \cdot \frac{\partial}{\partial t}$$

It is enough to show that $\frac{\partial}{\partial s}\langle \frac{\partial h}{\partial s}, \frac{\partial h}{\partial t} \rangle = 0$. We have

$$\begin{aligned}
\frac{\partial}{\partial s}\langle \frac{\partial h}{\partial s}, \frac{\partial h}{\partial t} \rangle &= \langle \nabla_{\frac{\partial}{\partial s}} \frac{\partial h}{\partial s}, \frac{\partial h}{\partial t} \rangle + \langle \frac{\partial h}{\partial s}, \nabla_{\frac{\partial}{\partial s}} \frac{\partial h}{\partial t} \rangle \\
&= \langle \frac{\partial h}{\partial s}, \nabla_{\frac{\partial}{\partial s}} \frac{\partial h}{\partial t} \rangle \text{ (since } s \mapsto h(s,t) \text{ is a geodesic)} \\
&= \langle \frac{\partial h}{\partial s}, \nabla_{\frac{\partial h}{\partial s}} \frac{\partial h}{\partial t} \rangle = \langle \frac{\partial h}{\partial s}, \nabla_{\frac{\partial h}{\partial t}} \frac{\partial h}{\partial s} \rangle \\
&= \frac{1}{2}\frac{\partial}{\partial s}\langle \frac{\partial h}{\partial s}, \frac{\partial h}{\partial s} \rangle = 0 \text{ by assumption (ii)}
\end{aligned}$$

■

Consider the following set up: $p \in M, x \in T_pM, x \neq 0_p, v_x, w_x \in T_x(T_pM)$ where v_x, w_x correspond to $v, w \in T_pM$. If v_x is radial, i.e. if v is a scalar multiple of x, then the Gauss lemma implies that

$$\langle T_x \exp_p v_x, T_x \exp_p w_x \rangle = \langle v_x, w_x \rangle := \langle v, w \rangle_p.$$

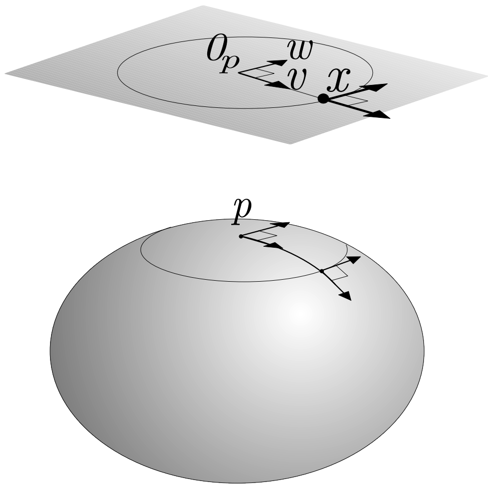
If v_x is not assumed to be radial then the above equality fails in general but we need to have the dimension of the manifold greater than 3 in order to see what can go wrong.

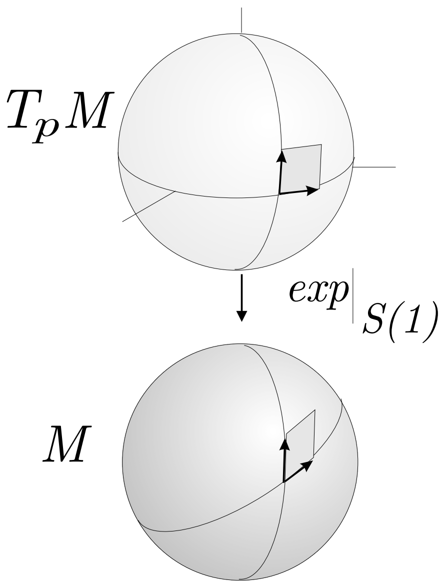
Exercise 20.11 *Show that if a geodesic* $\gamma : [a,b) \to M$ *is extendable to a continuous map* $\overline{\gamma} : [a,b] \to M$ *then there is an* $\varepsilon > 0$ *such that* $\gamma : [a,b) \to M$ *is extendable further to a geodesic* $\widetilde{\gamma} : [a, b+\varepsilon) \to M$ *such that* $\widetilde{\gamma}|_{[a,b]} = \overline{\gamma}$. *This is easy.*

Exercise 20.12 *Show that if* M, g *is a semi-Riemannian manifold then there exists a cover* $\{U_\alpha\}$ *such that each nonempty intersection* $U_{\alpha\beta} = U_\alpha \cap U_\beta$ *also convex.*

Let $v \in T_pM$ and consider the geodesic $\gamma_v : t \mapsto \exp_p tv$. Two rather easy facts that we often use without mention are

1. $\langle \gamma_v(t), \gamma_v(t) \rangle = \langle v, v \rangle$ for all t in the domain of γ_v.


0_p
w
v
x
p


T_pM
$exp|_{S(1)}$
M

2. If $r > 0$ is such that $\exp_p rv$ is defined then $\int_0^r |\langle \gamma_v(t), \gamma_v(t) \rangle|^{1/2} \, dt = r \, |\langle v, v \rangle|^{1/2} = r \, |v|$. In particular, if v is a unit vector then the length of the geodesic $\gamma_v|_{[0,r]}$ is r.

Under certain conditions geodesics can help us draw conclusions about maps. The following result is an example and a main ingredient in the proof of the Hadamard theorem proved later.

Theorem 20.10 *Let $f : M, \mathrm{g} \to N, \mathrm{h}$ be a local isometry of semi-Riemannian manifolds with N connected. Suppose that f has the property that given any geodesic $\gamma : [0, 1] \to N$ and a $p \in M$ with $f(p) = \gamma(0)$, there is a curve $\widetilde{\gamma} : [0, 1] \to M$ such that $p = \widetilde{\gamma}(0)$ and $\gamma = f \circ \widetilde{\gamma}$. Then ϕ is a semi-Riemannian covering.*

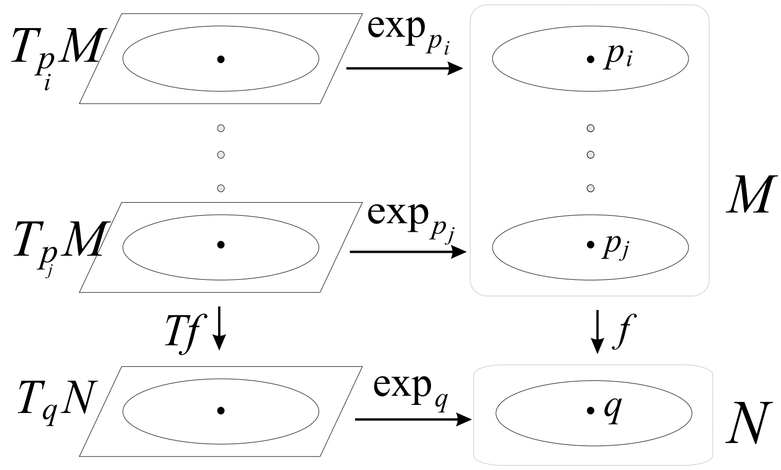
Proof. Since any two points of N can be joined by a broken geodesic it is easy to see that the hypotheses imply that f is onto.

Let U be a normal neighborhood of an arbitrary point $q \in N$ and $\widetilde{U} \subset T_q M$ the open set such that $\exp_q(\widetilde{U}) = U$. We will show that U is evenly covered by f. Choose $p \in f^{-1}\{q\}$. Observe that $T_p f : T_p M \to T_q N$ is a linear isometry (the metrics on the $T_p M$ and $T_q N$ are given by the scalar products $\mathrm{g}(p)$ and $\mathrm{h}(q)$). Thus $\widetilde{V_p} := T_p f^{-1}(\widetilde{U})$ is starshaped about $0_p \in T_p M$. Now if $v \in \widetilde{V}$ then by hypothesis the geodesic $\gamma(t) := \exp_q(t \, (T_p f v))$ has a lift to a curve $\widetilde{\gamma} : [0, 1] \to M$ with $\widetilde{\gamma}(0) = p$. But since f is a local isometry this curve must be a geodesic and it is also easy to see that $T_p(\widetilde{\gamma}'(0)) = \gamma'(0) = T_p f v$. It follows that $v = \widetilde{\gamma}'(0)$ and then $\exp_p(v) = \widetilde{\gamma}(1)$. Thus $\exp_p$ is defined on all of $\widetilde{V}$. In fact, it is clear that $f(\exp_p v) = \exp_{f(p)}(Tfv)$ and so we see that f maps $V_p := \exp_p(\widetilde{V}_p)$ onto the set $\exp_q(\widetilde{U}) = U$. We show that V is a normal neighborhood of p: We have $f \circ \exp_p = \exp_{f(p)} \circ Tf$ and it follows that $f \circ \exp_p$ is a diffeomorphism on $\widetilde{V}$. But then $\exp_p : \widetilde{V}_p \to V_p$ is 1-1 and onto and when combined with the fact that $Tf \circ T \exp_p$ is a linear isomorphism at each $v \in \widetilde{V}_p$ and the fact that Tf is a linear isomorphism it follows that $T_v \exp_p$ is a linear isomorphism. It follows that $\widetilde{U}$ is open and $\exp_p : \widetilde{V}_p \longrightarrow V_p$ is a diffeomorphism.

Now if we compose we obtain $f|_{V_p} = \exp_{f(p)}\big|_U \circ Tf \circ \exp_p\big|_{V_p}^{-1}$ which is a diffeomorphism taking V_p onto U.

Now we show that if $p_i, p_j \in f^{-1}\{q\}$ and $p_i \neq p_j$ then the sets V_{p_i} and V_{p_j} (obtained for these points as we did for a generic p above) are disjoint. Suppose to the contrary that $x \in V_{p_i} \cap V_{p_j}$ and let $\gamma_{p_i m}$ and $\gamma_{p_j m}$ be the reverse radial geodesics from m to p_i and p_j respectively. Then $f \circ \gamma_{p_i m}$ and $f \circ \gamma_{p_j m}$ are both reversed radial geodesics from $f(x)$ to q and so must be equal. But then $f \circ \gamma_{p_i m}$ and $f \circ \gamma_{p_j m}$ are equal since they are both lifts of the same curve an start at the same point. It follows that $p_i = p_j$ after all. It remains to prove that $f^{-1}(U) \supset \cup_{p \in f^{-1}(q)} V_p$ since the reverse inclusion is obvious. Let $x \in f^{-1}(U)$ and let $\alpha : [0, 1] \to U$ be the reverse radial geodesic from $f(x)$ to the center point q. Now let γ be the lift of α starting at x and let $p = \gamma(1)$.

Then $f(p) = \alpha(1) = q$ which means that $p \in f^{-1}(q)$. One the other hand, the image of γ must lie in $\widetilde{V}_p$ and so $x \in \widetilde{V}_p$. ■



20.6 Riemannian Manifolds and Distance

Once we have a notion of the length of a curve we can then define a distance function (metric in the sense of "metric space") as follow. Let $p, q \in M$. Consider the set path(p, q) of all smooth curves that begin at p and end at q. We define

$$\operatorname{dist}(p, q) = \inf\{l \in \mathbb{R} : l = L(c) \text{ and } c \in \operatorname{path}(p, q)\}. \tag{20.8}$$

or a general manifold just because $\operatorname{dist}(p, q) = r$ does **not necessarily** mean that there must be a curve connecting p to q having length r. To see this just consider the points $(-1, 0)$ and $(1, 0)$ on the punctured plane $\mathbb{R}^2 - 0$.

Definition 20.28 *If $p \in M$ is a point in a Riemannian manifold and $R > 0$ then the set $B_R(p)$ (also denoted $B(p, R)$) defined by $B_R(p) = \{q \in M : dist(p, q) < p\}$ is called a (geodesic)* ***ball*** *centered at p with radius R.*

It is important to notice that unless R is small enough $B_R(p)$ may not be homeomorphic to a ball in a Euclidean space.

Theorem 20.11 (distance topology) *Given a Riemannian manifold, define the distance function* dist *as above. Then M, dist is a metric space and the induced topology coincides with the manifold topology on M.*

Proof. That dist is true distance function (metric) we must show that
(1) dist is symmetric,
(2) dist satisfies the triangle inequality,

(3) $\operatorname{dist}(p,q) \geq 0$ and
(4) $\operatorname{dist}(p,q) = 0$ if and only if $p = q$.

Now (1) is obvious and (2) and (3) are clear from the properties of the integral and the metric tensor. To prove (4) we need only show that if $p \neq q$ then $\operatorname{dist}(p,\dot{q}) > 0$. Choose a chart ψ_α, U_α containing p but not q (M is Hausdorff). Now since $\psi_\alpha(U_\alpha) \subset \mathbb{R}^n$ we can transfer the Euclidean distance to U_α and define a small Euclidean ball $B_{Euc}(p,r)$ in this chart. Now any path from p to q must hit the boundary sphere $S(r) = \partial B_{Euc}(p,r)$. Now by compactness of $\bar{B}_{Euc}(p,r)$ we see that there are constants C_0 and C_1 such that $C_1\delta_{ij} \geq \mathrm{g}_{ij}(x) \geq C_0\delta_{ij}$ for all $x \in \bar{B}_{Euc}(p,r)$. Now any piecewise smooth curve $c : [a,b] \to M$ from p to q hits $S(r)$ at some parameter value $b_1 \leq b$ where we may assume this is the first hit (i.e. $c(t) \in B_{Euc}(p,r)$ for $a \leq t < b_0$). Now there is a curve that goes directly from p to q with respect to the Euclidean distance; i.e. a radial curve in the given Euclidean coordinates. This curve is given in coordinates as $\delta_{p,q}(t) = \frac{1}{b-1}(b-t)x(p) + \frac{1}{b-a}(t-a)x(q)$. Thus we have

$$\begin{aligned} L(c) &\geq \int_a^{b_0} \left(\mathrm{g}_{ij} \frac{d(x^i \circ c)}{dt} \frac{d(x^j \circ c)}{dt} \right)^{1/2} dt \geq C_0^{1/2} \int_a^{b_0} \left(\delta_{ij} \frac{d(x^i \circ c)}{dt} \right)^{1/2} dt \\ &= C_0^{1/2} \int_a^{b_0} |c'(t)|\, dt \geq C_0^{1/2} \int_a^{b_0} \left|\delta'_{p,q}(t)\right| dt = C_0^{1/2} r. \end{aligned}$$

Thus we have that $\operatorname{dist}(p,q) = \inf\{L(c) : c \text{ a curve from } p \text{ to } q\} \geq C_0^{1/2} r > 0$. This last argument also shows that if $\operatorname{dist}(p,x) < C_0^{1/2} r$ then $x \in B_{Euc}(p,r)$. This means that if $B(p, C_0^{1/2} r)$ is a ball with respect to dist then $B(p, C_0^{1/2} r) \subset B_{Euc}(p,r)$. Conversely, if $x \in B_{Euc}(p,r)$ then letting $\delta_{p,x}$ a "direct curve" analogous to the one above that connects p to x we have

$$\begin{aligned} \operatorname{dist}(p,x) &\leq L(\delta_{p,x}) \\ &= \int_a^{b_0} \left(\mathrm{g}_{ij} \frac{d(x^i \circ \delta)}{dt} \frac{d(x^j \circ \delta)}{dt} \right)^{1/2} dt \\ &\leq C_1^{1/2} \int_a^{b_0} \left|\delta'_{p,x}(t)\right| dt = C_1^{1/2} r \end{aligned}$$

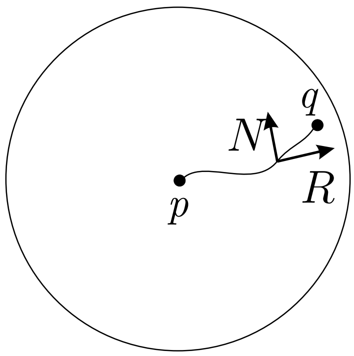
so we conclude that $B_{Euc}(p,r) \subset B(p, C_1^{1/2} r)$. Now we have that inside a chart, every dist-ball contains a Euclidean ball and visa versa. Thus since the manifold topology is generated by open subsets of charts we see that the two topologies coincide as promised. ∎

Lemma 20.6 *Let U be a normal neighborhood of a point p in a Riemannian manifold M, g. If $q \in U$ the radial geodesic $\gamma : [0,1] \to M$ such that $\gamma(0) = p$ and $\gamma(1) = q$, then γ is the unique shortest curve (up to reparameterization) connecting p to q.*

Proof. Let α be a curve connecting p to q. Without loss we may take the domain of α to be $[0,b]$. Let $\frac{\partial}{\partial r}$ be the radial unit vector field in U. Then if we define the vector field R along α by $t \mapsto \frac{\partial}{\partial r}\big|_{\alpha(t)}$ then we may write $\dot{\alpha} = \langle R, \dot{\alpha}\rangle R + N$ for some field N normal to R (but note $N(0) = 0$). We now have

$$\begin{aligned} L(\alpha) &= \int_0^b \langle \dot{\alpha}, \dot{\alpha}\rangle^{1/2} dt = \int_0^b \left[\langle R, \dot{\alpha}\rangle^2 + \langle N, N\rangle\right]^{1/2} dt \\ &\geq \int_0^b |\langle R, \dot{\alpha}\rangle|\, dt \geq \int_0^b \langle R, \dot{\alpha}\rangle dt = \int_0^b \frac{d}{dt}(r \circ \alpha)\, dt \\ &= r(\alpha(b)) = r(q) \end{aligned}$$

On the other hand, if $v = \dot{\gamma}(0)$ then $r(q) = \int_0^1 |v|\, dt = \int_0^1 \langle \dot{\gamma}, \dot{\gamma}\rangle^{1/2} dt$ so $L(\alpha) \geq L(\gamma)$. Now we show that if $L(\alpha) = L(\gamma)$ then α is a reparametrization of γ. Indeed, if $L(\alpha) = L(\gamma)$ then all of the above inequalities must be equalities so that N must be identically zero and $\frac{d}{dt}(r \circ \alpha) = \langle R, \dot{\alpha}\rangle = |\langle R, \dot{\alpha}\rangle|$. It follows that $\dot{\alpha} = \langle R, \dot{\alpha}\rangle R = \left(\frac{d}{dt}(r \circ \alpha)\right) R$ and so α travels radially from p to q and so must be a reparametrization of γ. ■



20.7 Covariant differentiation of Tensor Fields

Let ∇ be a natural covariant derivative on M. It is a consequence of proposition 8.4 that for each $X \in \mathfrak{X}(U)$ there is a unique tensor derivation ∇_X on $\mathfrak{T}^r_s(U)$ such that ∇_X commutes with contraction and coincides with the given covariant derivative on $\mathfrak{X}(U)$ (also denoted ∇_X) and with $\mathcal{L}_X f$ on $C^\infty(U)$.

To describe the covariant derivative on tensors more explicitly consider $\Upsilon \in \mathfrak{T}^1_1$ with a 1-form Since we have the contraction $Y \otimes \Upsilon \mapsto C(Y \otimes \Upsilon) = \Upsilon(Y)$ we should have

$$\begin{aligned} \nabla_X \Upsilon(Y) &= \nabla_X C(Y \otimes \Upsilon) \\ &= C(\nabla_X (Y \otimes \Upsilon)) \\ &= C(\nabla_X Y \otimes \Upsilon + Y \otimes \nabla_X \Upsilon) \\ &= \Upsilon(\nabla_X Y) + (\nabla_X \Upsilon)(Y) \end{aligned}$$

and so we should define $(\nabla_X \Upsilon)(Y) := \nabla_X(\Upsilon(Y)) - \Upsilon(\nabla_X Y)$. If $\Upsilon \in \mathfrak{T}^1_s$ then

$$(\nabla_X \Upsilon)(Y_1, ..., Y_s) = \nabla_X(\Upsilon(Y_1, ..., Y_s)) - \sum_{i=1}^{s} \Upsilon(..., \nabla_X Y_i, ...)$$

Now if $Z \in \mathfrak{T}^1_0$ we apply this to $\nabla Z \in \mathfrak{T}^1_1$ and get

$$\begin{aligned}(\nabla_X \nabla Z)(Y) &= X(\nabla Z(Y)) - \nabla Z(\nabla_X Y) \\ &= \nabla_X(\nabla_Y Z) - \nabla_{\nabla_X Y} Z\end{aligned}$$

from which we get the following definition:

Definition 20.29 *The second covariant derivative of a vector field $Z \in \mathfrak{T}^1_0$ is*

$$\nabla^2 Z : (X, Y) \mapsto \nabla^2_{X,Y}(Z) = \nabla_X(\nabla_Y Z) - \nabla_{\nabla_X Y} Z$$

Recall, that associated to any connection we have a curvature operator. In the case of the Levi-Civita Connection on a semi-Riemannian manifold the curvature operator is given by

$$R_{X,Y} Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z$$

and the associated $\mathfrak{T}^1_3$ tensor field $R(\alpha, X, Y, Z) := \alpha\,(R_{X,Y} Z)$ is called the Riemann curvature tensor. In a slightly different form we have the $\mathfrak{T}^0_4$ tensor defined by $R(W, X, Y, Z) := \langle W, R_{X,Y} Z\rangle$

Definition 20.30 *A tensor field Υ is said to be parallel if $\nabla_\xi \Upsilon = 0$ for all ξ. Similarly, if $\sigma : I \to T^r_s(M)$ is a tensor field along a curve $c : I \to M$ satisfies $\nabla_{\partial_t} \sigma = 0$ on I then we say that σ is parallel along c. Just as in the case of a general connection on a vector bundle we then have a parallel transport map $P(c)^t_{t_0} : T^r_s(M)_{c(t_0)} \to T^r_s(M)_{c(t)}$.*

Exercise 20.13 *Prove that*

$$\nabla_{\partial_t} \sigma(t) = \lim_{\epsilon \to 0} \frac{P(c)^t_{t+\epsilon} \sigma(t+\epsilon) - \sigma(t)}{\epsilon}.$$

Also, if $\Upsilon \in \mathfrak{T}^r_s$ then if c^X is the curve $t \mapsto Fl^X_t(p)$

$$\nabla_X \Upsilon(p) = \lim_{\epsilon \to 0} \frac{P(c^X)^t_{t+\epsilon}(\Upsilon \circ Fl^X_t(p)) - Y \circ Fl^X_t(p)}{\epsilon}.$$

The map $\nabla_X : \mathfrak{T}^r_s M \to \mathfrak{T}^r_s M$ just defined commutes with contraction. This means, for instance, that

$$\begin{aligned}\nabla_i(\Upsilon^{jk}{}_k) &= \nabla_i\ \Upsilon^{jk}{}_k \text{ and} \\ \nabla_i(\Upsilon^{ik}{}_j) &= \nabla_i\ \Upsilon^{ik}{}_l\,.\end{aligned}$$

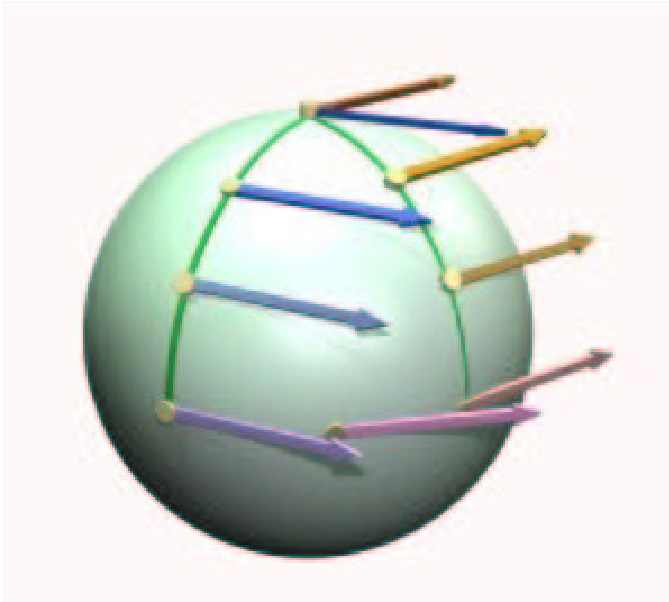


Figure 20.1: Parallel transport around path shows holonomy.

Furthermore, if the connection we are extending is the Levi-Civita connection for semi-Riemannian manifold M, g then

$$\nabla_\xi \mathrm{g} = 0 \text{ for all } \xi$$

To see this recall that

$$\nabla_\xi(\mathrm{g} \otimes Y \otimes W) = \nabla_\xi \mathrm{g} \otimes X \otimes Y + \mathrm{g} \otimes \nabla_\xi X \otimes Y + \mathrm{g} \otimes X \otimes \nabla_\xi Y$$

which upon contraction yields

$$\begin{aligned} \nabla_\xi(\mathrm{g}(X,Y)) &= (\nabla_\xi \mathrm{g})(X,Y) + \mathrm{g}(\nabla_\xi X, Y) + \mathrm{g}(X, \nabla_\xi Y) \\ \xi\langle X,Y\rangle &= (\nabla_\xi \mathrm{g})(X,Y) + \langle \nabla_\xi X, Y\rangle + \langle X, \nabla_\xi Y\rangle. \end{aligned}$$

We see that $\nabla_\xi \mathrm{g} \equiv 0$ for all ξ if and only if $\langle X,Y\rangle = \langle \nabla_\xi X, Y\rangle + \langle X, \nabla_\xi Y\rangle$ for all ξ, X, Y. In other words the statement that the metric tensor is parallel (constant) with respect to ∇ is the same as saying that the connection is a metric connection.

When ∇ is the Levi-Civita connection for the Riemannian manifold M, g we get the interesting formula

$$(\mathcal{L}_X \mathrm{g})(Y,Z) = \mathrm{g}(\nabla_X Y, Z) + \mathrm{g}(Y, \nabla_X Z) \tag{20.9}$$

for vector fields $X, Y, Z \in \mathfrak{X}(M)$.

20.8 Curvature

For M, g a Riemannian manifold with associated Levi-Civita connection ∇ we have the associated curvature which will now be called the Riemann curvature tensor: For $X, Y \in \mathfrak{X}(M)$ by $R_{X,Y} : \mathfrak{X}(M) \to \mathfrak{X}(M)$

$$R_{X,Y}Z := \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]}Z.$$

Traditionally two other "forms" of the curvature tensor are defined as

1. A $\mathfrak{T}^1_3$ tensor defined by $R(\alpha, Z, X, Y) = \alpha\,(R_{X,Y}Z)$.
2. A $\mathfrak{T}^0_4$ tensor defined by $R(W, Z, X, Y) = \langle R_{X,Y}Z, W\rangle$
   The seemly odd ordering of the variables is traditional and maybe a bit unfortunate.

Theorem 20.12 $R(W, Z, X, Y) = \langle R_{X,Y}Z, W\rangle$ *is tensorial in all variables and*

(i) $R_{X,Y} = -R_{Y,X}$

(ii) $\langle R(X,Y)Z, W\rangle = -\langle R(X,Y)W, Z\rangle$
(iii) $R_{X,Y}Z + R_{Y,Z}X + R_{Z,X}Y = 0$
(iv) $\langle R(X,Y)Z, W\rangle = \langle R(Z,W)X, Y\rangle$

Proof. (i) is immediate from the definition of R.
(ii) It is enough to show that $\langle R(X,Y)Z, Z\rangle = 0$. We have

$$\begin{aligned}\langle R(X,Y)Z, Z\rangle &= \langle \nabla_X \nabla_Y Z, Z\rangle - \langle Z, \nabla_X \nabla_Y Z\rangle \\ &= X\langle \nabla_Y Z, Z\rangle - \langle \nabla_X Z, \nabla_Y Z\rangle - Y\langle \nabla_X Z, Z\rangle \\ &+ \langle \nabla_X Z, \nabla_Y Z\rangle \\ &= \frac{1}{2}XY\langle Z, Z\rangle - \frac{1}{2}YX\langle Z, Z\rangle \\ &= \frac{1}{2}[X,Y]\langle Z, Z\rangle\end{aligned}$$

But since R is a tensor we may assume without loss that $[X,Y] = 0$. Thus $\langle R(X,Y)Z, Z\rangle = 0$ and the result follow by polarization.
(iii)

$$\begin{aligned}&R_{X,Y}Z + R_{Y,Z}X + R_{Z,X}Y \\ &= \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z + \nabla_Y \nabla_Z X - \nabla_Z \nabla_Y X + \nabla_Z \nabla_X Y - \nabla_X \nabla_Z Y = 0\end{aligned}$$

(iv) The proof of (iv) is rather unenlightening and is just some combinatorics which we omit: (doit?) ∎

Definition 20.31 *A semi-Riemannian manifold M, g is called* ***flat*** *if the curvature is identically zero.*

Theorem 20.13 *For $X, Y, Z \in \mathfrak{X}(M)$ we have*

$$(\nabla_Z R)(X, Y) + \nabla_X R(Y, Z) + \nabla_Y R(Z, X) = 0$$

Proof. *This is the second Bianchi identity for the Levi-Civita connection but we give another proof here. Since this is a tensor equation we only need to prove it under the assumption that all brackets among the X, Y, Z are zero. First we have*

$$\begin{aligned}(\nabla_Z R)(X, Y)W &= \nabla_Z(R(X, Y)W) - R(\nabla_Z X, Y)W \\ &\quad - R(X, \nabla_Z Y)W - R(X, Y)\nabla_Z W \\ &= [\nabla_Z, R_{X,Y}]W - R(\nabla_Z X, Y)W - R(X, \nabla_Z Y)W.\end{aligned}$$

Using this we calculate as follows:

$$\begin{aligned}&(\nabla_Z R)(X, Y)W + (\nabla_X R)(Y, Z)W + (\nabla_Y R)(Z, X)W \\ &= [\nabla_Z, R_{X,Y}]W + [\nabla_X, R_{Y,Z}]W + [\nabla_Y, R_{Z,X}]W \\ &- R(\nabla_Z X, Y)W - R(X, \nabla_Z Y)W \\ &- R(\nabla_X Y, Z)W - R(Y, \nabla_X Z)W \\ &- R(\nabla_Y Z, X)W - R(Z, \nabla_Y X)W \\ &= [\nabla_Z, R_{X,Y}]W + [\nabla_X, R_{Y,Z}]W + [\nabla_Y, R_{Z,X}]W \\ &+ R([X, Z], Y)W + R([Z, Y], X)W + R([Y, X], Z)W \\ &= [\nabla_Z, [\nabla_X, \nabla_Y]] + [\nabla_X, [\nabla_Y, \nabla_Z]] + [\nabla_Y, [\nabla_Z, \nabla_X]] = 0\end{aligned}$$

The last identity is the Jacobi identity for commutator (see exercise) ∎

Exercise 20.14 *Show that is $L_i, i = 1, 2, 3$ are linear operators $V \to V$ and the commutator is defined as usual ($[A, B] = AB - BA$) then we always have the Jacobi identity $[L_1, [L_2, L_3] + [L_2, [L_3, L_1] + [L_3, [L_1, L_2] = 0$.*

These several symmetry properties for the Riemann curvature tensor allow that we have a well defined map

$$\mathfrak{R} : \wedge^2(TM) \to \wedge^2(TM)$$

which is symmetric with respect the natural extension of g to $\wedge^2(TM)$. Recall that the natural extension is defined so that for an orthonormal $\{e_i\}$ the basis $\{e_i \wedge e_j\}$ is also orthonormal. We have

$$g(v_1 \wedge v_2, v_3 \wedge v_4) = \det \begin{pmatrix} g(v_1, v_3) & g(v_1, v_4) \\ g(v_2, v_3) & g(v_2, v_4) \end{pmatrix}$$

$\mathfrak{R}$ is defined implicitly as follows:

$$g(\mathfrak{R}(v_1 \wedge v_2), v_3 \wedge v_4) := \langle R(v_1, v_2)v_4, v_3\rangle.$$

20.8.1 Tidal Force and Sectional curvature

For each $v \in TM$ the tidal force operator $R_v : T_pM \to T_pM$ is defined by

$$R_v(w) := R_{v,w}v.$$

Another commonly used quantity is the **sectional curvature:**

$$\begin{aligned} K(v \wedge w) &:= -\frac{\langle R_v(w), w\rangle}{\langle v, v\rangle\langle w, w\rangle - \langle v, w\rangle^2} \\ &= \frac{\langle \mathfrak{R}(v \wedge w), v \wedge w\rangle}{\langle v \wedge w, v \wedge w\rangle} \end{aligned}$$

where $v, w \in T_pM$. The value $K(v \wedge w)$ only depends on the oriented plane spanned by the vectors v and w therefore if $P = \operatorname{span}\{v, w\}$ is such a plane we also write $K(P)$ instead of $K(v \wedge w)$. The set of all planes in T_pM is denoted $Gr_p(2)$.

In the following definition V is an $\mathsf{R}-$module. The two cases we have in mind are (1) where V is $\mathfrak{X}$, $\mathsf{R} = C^\infty(M)$ and (2) where V is T_pM, R=$\mathbb{R}$.

Definition 20.32 *A multilinear function $F : \mathrm{V} \times \mathrm{V} \times \mathrm{V} \times \mathrm{V} \to \mathsf{R}$ is said to be **curvature-like** $x, y, z, w \in \mathrm{V}$ it satisfies the symmetries proved for the curvature R above; namely,*

(i) $F(x, y, z, w) = -F(y, x, z, w)$
(ii) $F(x, y, z, w) = -F(x, y, w, z)$

(iii) $F(x, y, z, w) + F(y, z, x, w) + F(z, x, y, w) = 0$
(iv) $F(x, y, z, w) = F(w, z, x, y)$

As an example the tensor $C_g(X, Y, Z, W) := g(Y, Z)g(X, W) - g(X, Z)g(Y, W)$ is curvature-like.

Exercise 20.15 *Show that C_g is curvature-like.*

Proposition 20.5 *If F is curvature-like and $F(v, w, v, w) = 0$ for all $v, w \in \mathrm{V}$ then $F \equiv 0$.*

Proof. From (iv) it follows that F is symmetric in the second and forth variables so if $F(v, w, v, w) = 0$ for all $v, w \in \mathrm{V}$ then $F(v, w, v, z) = 0$ for all $v, w, z \in \mathrm{V}$. Now is a simple to show that (i) and (ii) imply that $F \equiv 0$. ■

Proposition 20.6 *If $K(v \wedge w)$ is know for all $v, w \in T_pM$ or if $\langle R_{v,w}v, w\rangle = g(R_{v,w}v, w)$ is known for all $v, w \in T_pM$ then R itself is determined at p.*

Proof. Using an orthonormal basis for T_pM we see that K and ϕ (where $\phi(v, w) := g(R_{v,w}v, w)$) contain the same information so we will just show that

ϕ determines R:

$$\begin{aligned}
&\left.\frac{\partial^2}{\partial s \partial t}\right|_{0,0} (\phi(v+tz, w+su) - \phi(v+tu, w+sz)) \\
&= \left.\frac{\partial^2}{\partial s \partial t}\right|_{0,0} \{g(R(v+tz, w+su)v+tz, w+su) \\
&\quad - g(R(v+tu, w+sz)v+tu, w+sz)\} \\
&= 6R(v, w, z, u)
\end{aligned}$$

■

We are now in a position to prove the following important theorem.

Theorem 20.14 *The following are all equivalent:*

(i) $K(P) = \kappa$ *for all* $P \in Gr_p(2)$

(ii) $g(R_{v_1,v_2}v_3, v_4) = \kappa C_g(v_1, v_2, v_3, v_4)$ *for all* $v_1, v_2, v_3, v_4 \in T_pM$

(iii) $R_v(w) = \kappa(w - g(w, v)v)$ *for all* $w, v \in T_pM$ *with* $|v| = 1$
(iv) $\mathfrak{R}(\omega) = \kappa\omega$ *and* $\omega \in \wedge^2 T_pM$.

Proof. Let $p \in M$. The proof that (ii)$\Longrightarrow$ (iii) and that (iii)$\Longrightarrow$ (i) is left as an easy exercise. We prove that (i) $\Longrightarrow$ (ii)$\Longrightarrow$ (iv)$\Longrightarrow$(i).
(i)$\Longrightarrow$ (ii): Let $T_g := F_\kappa - \kappa C_g$. Then T_g is curvature-like and $T_g(v, w, v, w) = 0$ for all $v, w \in T_pM$ by assumption. It follows from 20.5 that $T_g \equiv 0$.
(ii)$\Longrightarrow$ (iv): Let $\{e_1, ..., e_n\}$ be an orthonormal basis for T_pM. Then $\{e_i \wedge e_j\}_{i<j}$ is an orthonormal basis for $\wedge^2 T_pM$. Using (ii) we see that

$$\begin{aligned}
g(\mathfrak{R}(e_i \wedge e_j), e_k \wedge e_l) &= g(R_{e_i,e_j}, e_k, e_l) \\
&= g(R(e_i, e_j), e_k, e_l) \\
&= \kappa C_g(v_1, v_2, v_3, v_4) \\
&= \kappa g(e_i \wedge e_j, e_k \wedge e_l) \text{ for all } k, l
\end{aligned}$$

Since we are using a basis this implies that $\mathfrak{R}(e_i \wedge e_j) = \kappa e_i \wedge e_j$
(iv)$\Longrightarrow$(i): This follows because if v, w are orthonormal we have $\kappa = g(\mathfrak{R}(v \wedge w), v \wedge w) = K(v \wedge w)$. ■

20.8.2 Ricci Curvature

Definition 20.33 *Let* M, g *be a semi-Riemannian manifold. The Ricci curvature is the* $(1,1)-$*tensor Ric defined by*

$$Ric(v, w) := \sum_{i=1}^{n} \varepsilon_i \langle R_{v,e_i} e_i, w \rangle$$

where $\{e_1, ..., e_n\}$ *is any orthonormal basis of* T_pM *and* $\varepsilon_i := \langle e_i, e_i \rangle$.

We say that the Ricci curvature Ric is **bounded from below by** κ and write $Ric \geq k$ if $Ric(v,w) \geq k\langle v,w\rangle$ for all $v,w \in TM$. Similar and obvious definitions can be given for $Ric \leq k$ and the strict bounds $Ric > k$ and $Ric < k$. Actually, it is usually the case that the bound on Ricci curvature is given in the form $Ric \geq \kappa(n-1)$ where $n = \dim(M)$.

There is a very important an interesting class of manifolds Einstein manifolds. A semi-Riemannian manifold M, g is called an Einstein manifold with Einstein constant k if and only if $Ric(v,w) = k\langle v,w\rangle$ for all $v,w \in TM$. For example, if M, g has constant sectional curvature κ then M, g called an **Einstein manifold** with Einstein constant $k = \kappa(n-1)$. The effect of this condition depends on the signature of the metric. Particularly interesting is the case where the index is 0 (Riemannian) and also the case where the index is 1 (Lorentz). Perhaps the first question one should ask is whether there exists any Einstein manifolds that do not have constant curvature. It turns out that there are many interesting Einstein manifolds that do not have constant curvature.

Exercise 20.16 *Show that if M is connected and $\dim(M) > 2$ then $Ric = f\mathrm{g}$ where $f \in C^\infty(M)$ then $Ric = k\mathrm{g}$ for some $k \in \mathbb{R}$ (M, g is Einstein).*

20.9 Jacobi Fields

One again we consider a semi-Riemannian manifold M, g of arbitrary index. We shall be dealing with two parameter maps $h : [\epsilon_1, \epsilon_2] \times [a,b] \to M$. The partial maps $t \mapsto h_s(t) = h(s,t)$ are called the longitudinal curves and the curves $s \mapsto h(s,t)$ are called the transverse curves. Let α be the center longitudinal curve $t \mapsto h_0(t)$. The vector field along α defined by $V(t) = \frac{d}{ds}\big|_{s=0} h_s(t)$ is called the **variation vector field** along α. We will use the following important result more than once:

Lemma 20.7 *Let Y be a vector field along the map $h : [\epsilon_1, \epsilon_2] \times [a,b] \to M$. Then*

$$\nabla_{\partial_s}\nabla_{\partial_t}Y - \nabla_{\partial_s}\nabla_{\partial_t}Y = R(\partial_s h, \partial_t h)Y$$

Proof. This rather plausible formula takes a little case since h may not even be an immersion. Nevertheless, if one computes in a local chart the result falls out after a mildly tedious computation which we leave to the curious reader. ∎

Suppose we have the special situation that, for each s, the partial maps $t \mapsto h_s(t)$ are geodesics. In this case let us denote the center geodesic $t \mapsto h_0(t)$ by γ. We call h a variation of γ through geodesics. Let h be such a special variation and V the variation vector field. Using the previous lemma 20.7 the result of exercise 20.5 we compute

$$\begin{aligned}
\nabla_{\partial_t}\nabla_{\partial_t}V &= \nabla_{\partial_t}\nabla_{\partial_t}\partial_s h = \nabla_{\partial_t}\nabla_{\partial_s}\partial_t h \\
&= \nabla_{\partial_s}\nabla_{\partial_t}\partial_t h + R(\partial_t h, \partial_s h)\partial_t h \\
&= R(\partial_t h, \partial_s h)\partial_t h
\end{aligned}$$

and evaluating at $s = 0$ we get $\nabla_{\partial_t}\nabla_{\partial_t}V(t) = R(\dot{\gamma}(t), V(t))\dot{\gamma}(t)$. This equation is important and shows that V is a Jacobi field:

Definition 20.34 *Let $\gamma : [a, b] \to M$ be a geodesic and $J \in \mathfrak{X}_\gamma(M)$ be a vector field along γ. The field J is called a* ***Jacobi field*** *if*

$$\nabla_{\partial_t} J = R(\dot{\gamma}(t), J(t))\dot{\gamma}(t)$$

for all $t \in [a, b]$.

In local coordinates we recognize the above as a second order (system of) linear differential equations and we easily arrive at the following

Theorem 20.15 *Let M, g and $\gamma : [a, b] \to M$ be a geodesic as above. Given $w_1, w_2 \in T_{\gamma(a)}M$, there is a unique Jacobi field $J^{w_1,w_2} \in \mathfrak{X}_\gamma(M)$ such that $J(a) = w_1$ and $\frac{\nabla J}{dt}(a) = w_2$. The set $Jac(\gamma)$ of all Jacobi fields along γ is a vector space isomorphic to $T_{\gamma(a)}M \times T_{\gamma(a)}M$.*

If $\gamma_v(t) = \exp_p(tv)$ defined on $[0, b]$ then $J_\gamma^{0,w}$ denotes the unique Jacobi field along $\gamma_v(t)$ with initial conditions $J(0) = 0$ and $\nabla_{\partial_t} J(0) = w$.

Proposition 20.7 *If $w = rv$ for some $r \in \mathbb{R}$ then $J_\gamma^{0,rv}(t) = tr\gamma_v(t)$. If $w \perp v$ then $\langle J^{0,w}(t), \dot{\gamma}_v(t)\rangle = 0$ for all $t \in [0, b]$.*

Proof. First let $w = rv$. Let $J := rt\dot{\gamma}_v(t)$. Then clearly $J(0) = 0 \cdot \dot{\gamma}_v(0) = 0$ and $\nabla_{\partial_t} J(t) = rt\nabla_{\partial_t}\gamma_v(t) + r\gamma_v(t)$ so $\nabla_{\partial_t} J(0) = r\gamma_v(0) = rv$. Thus J satisfied the correct initial conditions. Now we have $\nabla^2_{\partial_t} J(t) = r\nabla^2_{\partial_t}\gamma_v(t) = r\nabla_{\partial_t}\dot{\gamma}_v(0) = 0$ since γ_v is a geodesic. One the other hand,

$$\begin{aligned}
&R(\dot{\gamma}_v(t), J(t))\dot{\gamma}_v(t)\\
&= R(\dot{\gamma}_v(t), r\dot{\gamma}_v(t))\dot{\gamma}_v(t)\\
&= rR(\dot{\gamma}_v(t), \dot{\gamma}_v(t))\dot{\gamma}_v(t) = 0
\end{aligned}$$

Thus $J = J^{0,rv}$.

Now for the case $w \perp v$ we have $\langle J^{0,w}(0), \dot{\gamma}_v(0)\rangle = 0$ and

$$\begin{aligned}
&\left.\frac{d}{dt}\right|_0 \langle J^{0,w}(t), \dot{\gamma}_v(t)\rangle\\
&= \langle \nabla_{\partial_t} J^{0,w}(0), \dot{\gamma}_v(0)\rangle + \langle J^{0,w}(t0), \nabla_{\partial_t}\dot{\gamma}_v(0)\rangle\\
&= \langle w, v\rangle = 0
\end{aligned}$$

Thus the function $f(t) = \langle J^{0,w}(t), \dot{\gamma}_v(t)\rangle$ satisfies $f'(0) = 0$ and $f(0) = 0$ and so $f \equiv 0$ on $[0, b]$. ■

Corollary 20.1 *Every Jacobi field $J^{0,w}$ along $\exp_v tv$ with $J^{0,w}(0) = 0$ has the form $J^{0,w} := rt\dot{\gamma}_v + J^{0,w_1}$ where $w = \nabla_{\partial_t} J(0) = rv + w_1$ and $w_1 \perp v$. Also, $J^{0,w}(t) \perp \dot{\gamma}_v(t)$ for all $t \in [0, b]$.*

We now examine the more general case of a Jacobi field J^{w_1,w_2} along a geodesic $\gamma : [a,b] \to M$. First notice that for any curve $\alpha : [a,b] \to M$ with $|\langle \dot{\alpha}(t), \dot{\alpha}(t)\rangle| > 0$ for all $t \in [a,b]$, any vector field Y along α decomposes into an orthogonal sum $Y^\top + Y^\perp$. This means that $Y^\top$ is a multiple of $\dot{\alpha}$ and $Y^\perp$ is normal to $\dot{\alpha}$. If $\gamma : [a,b] \to M$ is a geodesic then $\nabla_{\partial_t} Y^\perp$ is also normal to $\dot{\gamma}$ since $0 = \frac{d}{dt}\langle Y^\perp, \dot{\gamma}\rangle = \langle \nabla_{\partial_t} Y^\perp, \dot{\gamma}\rangle + \langle Y^\perp, \nabla_{\partial_t}\dot{\gamma}\rangle = \langle \nabla_{\partial_t} Y^\perp, \dot{\gamma}\rangle$. Similarly, $\nabla_{\partial_t} Y^\top$ is parallel to $\dot{\gamma}$ all along γ.

Proposition 20.8 *Let $\gamma : [a,b] \to M$ be a geodesic.*

(i) If $Y \in \mathfrak{X}_\gamma(M)$ is tangent to γ then Y is a Jacobi field if and only if $\nabla^2_{\partial_t} Y = 0$ along γ. In this case $Y(t) = (at+b)\dot{\gamma}(t)$.

(ii) If J is a Jacobi field along γ and there is some distinct $t_1, t_2 \in [a,b]$ with $J(t_1) \perp \dot{\gamma}(t_1)$ and $J(t_2) \perp \dot{\gamma}(t_2)$ then $J(t) \perp \dot{\gamma}(t)$ for all $t \in [a,b]$.

(iii) If J is a Jacobi field along γ and there is some $t_0 \in [a,b]$ with $J(t_0) \perp \dot{\gamma}(t_0)$ and $\nabla_{\partial_t} J(t_0) \perp \dot{\gamma}(t_0)$ then $J(t) \perp \dot{\gamma}(t)$ for all $t \in [a,b]$.

(iv) If γ is not a null geodesic then Y is a Jacobi field if and only if both $Y^\top$ and $Y^\perp$ are Jacobi fields.

Proof. (i) Let $Y = f\dot{\gamma}$. Then the Jacobi equation reads

$$\nabla^2_{\partial_t} f\dot{\gamma}(t) = R(\dot{\gamma}(t), f\dot{\gamma}(t))\dot{\gamma}(t) = 0$$

or

$$\nabla^2_{\partial_t} f\dot{\gamma}(t) = 0 \text{ or } f'' = 0$$

(ii) and (iii) $\frac{d^2}{dt^2}\langle Y, \dot{\gamma}\rangle = \langle R(\dot{\gamma}(t), Y(t))\dot{\gamma}(t), \dot{\gamma}(t)\rangle = 0$ (from the symmetries of the curvature tensor). Thus $\langle Y(t), \dot{\gamma}(t)\rangle = at + b$ for some $a, b \in \mathbb{R}$. The reader can now easily deduce both (ii) and (iii).

(iv) The operator $\nabla^2_{\partial_t}$ preserves the normal and tangential parts of Y. We now show that the same is true of the map $Y \mapsto R(\dot{\gamma}(t), Y)\dot{\gamma}(t)$. Since we assume that γ is not null we have $Y^\top = f\dot{\gamma}$ for some $\dot{\gamma}$. Thus $R(\dot{\gamma}(t), Y^\top)\dot{\gamma}(t) = R(\dot{\gamma}(t), f\dot{\gamma}(t))\dot{\gamma}(t) = 0$ which is trivially tangent to $\dot{\gamma}(t)$. On the other hand, $\langle R(\dot{\gamma}(t), Y^\perp(t))\dot{\gamma}(t), \dot{\gamma}(t)\rangle = 0$.

$$\begin{aligned}
\left(\nabla^2_{\partial_t} Y\right)^\top + \left(\nabla^2_{\partial_t} Y\right)^\perp &= \nabla^2_{\partial_t} Y = R(\dot{\gamma}(t), Y(t))\dot{\gamma}(t) \\
&= R(\dot{\gamma}(t), Y^\top(t))\dot{\gamma}(t) + R(\dot{\gamma}(t), Y^\perp(t))\dot{\gamma}(t) \\
&= 0 + R(\dot{\gamma}(t), Y^\perp(t))\dot{\gamma}(t)
\end{aligned}$$

So the Jacobi equation $\nabla^2_{\partial_t} J(t) = R(\dot{\gamma}(t), J(t))\dot{\gamma}(t)$ splits into two equations

$$\begin{aligned}
\nabla^2_{\partial_t} J^\top(t) &= 0 \\
\nabla^2_{\partial_t} J^\perp(t) &= R_{\dot{\gamma}(t), J^\perp(t)}\dot{\gamma}(t)
\end{aligned}$$

and the result follows from this. ■

The proof of the last result shows that a Jacobi field decomposes into a parallel vector field along γ which is just a multiple of the velocity $\dot{\gamma}$ and a "normal Jacobi field" $J^{\perp}$ which is normal to γ. Of course, the important part is the normal part and so we now restrict to that case. Thus we consider the Jacobi equation $\nabla^2_{\partial_t} J(t) = R(\dot{\gamma}(t), J(t))\dot{\gamma}(t)$ with initial conditions $J(a) = w_1$ with $\nabla_{\partial_t} J(a) = w_2$ and $w_1, w_2 \in (\dot{\gamma}(a))^{\perp}$. Notice that in terms of the tidal force operator the Jacobi equation is

$$\nabla^2_{\partial_t} J(t) = R_{\dot{\gamma}(t)}(J(t))$$

Exercise 20.17 *Prove that for $v \in T_{\dot{\gamma}(t)}M$ the operator R_v maps $(\dot{\gamma}(t))^{\perp}$ to itself.*

Definition 20.35 *Let $\gamma : [a, b] \to M$ be a geodesic. Let $\mathcal{J}_0(\gamma, a, b)$ denote the set of all Jacobi fields J such that $J(a) = J(b) = 0$.*

Definition 20.36 *If there exists a geodesic $\gamma : [a, b] \to M$ and a nonzero Jacobi field $J \in \mathcal{J}_0(\gamma, a, b)$ then we say that $\gamma(a)$ is conjugate to $\gamma(b)$ along γ.*

From standard considerations from the theory of linear differential equations the set $\mathcal{J}_0(\gamma, a, b)$ is a vector space. The dimension of the vector space $\mathcal{J}_0(\gamma, a, b)$ the order of the conjugacy. Since the Jacobi fields in $\mathcal{J}_0(\gamma, a, b)$ vanish twice and we have seen that this means that such fields are normal to $\dot{\gamma}$ all along γ it follows that the dimension of . $\mathcal{J}_0(\gamma, a, b)$ is at most $n-1$ where $n = \dim M$. We have seen that a variation through geodesics is a Jacobi field so if we can find a nontrivial variation h of a geodesic γ such that all of the longitudinal curves $t \mapsto h_s(t)$ begin and end at the same points $\gamma(a)$ and $\gamma(b)$ then the variation vector field will be a nonzero element of $\mathcal{J}_0(\gamma, a, b)$. Thus we conclude that $\gamma(a)$ is conjugate to $\gamma(b)$.

Let us get the exponential map into play. Let $\gamma : [0, b] \to M$ be a geodesic as above. Let $v = \dot{\gamma}(0) \in T_pM$. Then $\gamma : t \mapsto \exp_p tv$ is exactly our geodesic γ which begins at p and ends at q at time b. Now we create a variation of γ by

$$h(s, t) = \exp_p t(v + sw).$$

where $w \in T_pM$ and s ranges in $(-\epsilon, \epsilon)$ for some sufficiently small ϵ. Now we know that $J(t) = \frac{\partial}{\partial s}\big|_{s=0} h(s, t)$ is a Jacobi field. It is clear that $J(0) := \frac{\partial}{\partial s}\big|_{s=0} h(s, 0) = 0$. If w_{bv} is the vector tangent in $T_{bv}(T_pM)$ which canonically corresponds to w, in other words, if w_{bv} is the velocity vector at $s = 0$ for the curve $s \mapsto bv + sw$ in T_pM, then

$$\begin{aligned} J(b) &= \left.\frac{\partial}{\partial s}\right|_{s=0} h(s, b) \\ &= \left.\frac{\partial}{\partial s}\right|_{s=0} \exp_p t(bv + sw) = T_{bv} \exp_p(w_{bv}) \end{aligned}$$

Also,

$$\begin{aligned} \nabla_{\partial_t} J(0) &= \nabla_{\partial_t} \nabla_{\partial_s} \exp_p t(bv + sw)\big|_{s=0,t=0} \\ &= \nabla_{\partial_s}\big|_{s=0} \nabla_{\partial_t}\big|_{t=0} \exp_p t(bv + sw) \end{aligned}$$

But $X(s) := \frac{\partial}{\partial t}\big|_{t=0} \exp_p t(bv+sw) = bv+sw$ is a vector field along the constant curve $t \mapsto p$ and so by exercise 19.5 we have $\nabla_{\partial_s}\big|_{s=0} X(s) = X'(0) = w$. The equality $J(b) = T_{bv} \exp_p(v_{bv})$ is important because it shows that if $T_{bv} \exp_p : T_{bv}(T_pM) \to T_{\gamma(b)}M$ is not an isomorphism then we can find a vector $w_{bv} \in T_{bv}(T_pM)$ such that $T_{bv} \exp_p(w_{bv}) = 0$. But then if w is the vector in T_pM which corresponds to w_{bv} as above then for this choice of w the Jacobi field constructed above is such that $J(0) = J(b) = 0$ and so $\gamma(0)$ is conjugate to $\gamma(b)$ along γ. Also, if J is a Jacobi field with $J(0) = 0$ and $\nabla_{\partial_t} J(0) = w$ then this uniquely determines J and it must have the form $\frac{\partial}{\partial s}\big|_{s=0} \exp_p t(bv + sw)$ as above.

Theorem 20.16 *Let $\gamma : [0, b] \to M$ be a geodesic. Then the following are equivalent:*

(i) $\gamma(0)$ is conjugate to $\gamma(b)$ along γ.

(ii) There is a nontrivial variation h of γ through geodesics which all start at $p = \gamma(0)$ such that $J(b) := \frac{\partial h}{\partial s}(0, b) = 0$.

(iii) If $v = \gamma'(0)$ then $T_{bv} \exp_p$ is singular.

Proof. (ii)$\Longrightarrow$(i): We have already seen that a variation through geodesics is a Jacobi field J and if (ii) then by assumption $J(0) = J(b) = 0$ and so we have (i).

(i)$\Longrightarrow$(iii): If (i) is true then there is a nonzero Jacobi field J with $J(0) = J(b) = 0$. Now let $w = \nabla_{\partial_t} J(0)$ and $h(s, t) = \exp_p t(bv + sw)$. Then $h(s, t)$ is a variation through geodesics and $0 = J(b) = \frac{\partial}{\partial s}\big|_{s=0} \exp_p t(bv + sw) = T_{bv} \exp_p(w_{bv})$ so that $T_{bv} \exp_p$ is singular.

(iii)$\Longrightarrow$(ii): Let $v = \gamma'(0)$. If $T_{bv} \exp_p$ is singular then there is a w with $T_{bv} \exp_p \cdot w_{bv} = 0$. Thus the variation $h(s, t) = \exp_p t(bv + sw)$ does the job. ∎

20.10 First and Second Variation of Arc Length

Let us restrict attention to the case where α is either spacelike of timelike and let $\varepsilon = +1$ if α is spacelike and $\varepsilon = -1$ if α is timelike. This is just the condition that $\left|\langle \dot{h}_0(t), \dot{h}_0(t)\rangle\right| > 0$. By a simple continuity argument we may choose $\epsilon > 0$ small enough that $\left|\langle \dot{h}_s(t), \dot{h}_s(t)\rangle\right| > 0$ for all $s \in (-\epsilon, \epsilon)$. Consider the arc length functional defined by

$$L(\alpha) = \int_a^b \varepsilon\langle\dot{\alpha}(t), \dot{\alpha}(t)\rangle^{1/2} dt.$$

Now if $h : (-\epsilon, \epsilon) \times [a, b] \to M$ is a variation of α as above with variation vector field V then formally V is a tangent vector at α in the space of curves $[a, b] \to M$. Then we have the variation of the arc length functional defined by

$$\delta L|_\alpha (V) := \frac{d}{ds}_{s=0} L(h_s) := \frac{d}{ds}_{s=0} \int_a^b \varepsilon\langle\dot{h}_s(t), \dot{h}_s(t)\rangle^{1/2} dt.$$

So we are interesting in studying the critical points of $L(s) := L(h_s)$ and so we need to find $L'(0)$ and $L''(0)$. For the proof of the following proposition we use the result of exercise 20.5 to the effect that $\nabla_{\partial_s} \partial_t h = \nabla_{\partial_t} \partial_s h$.

Proposition 20.9 *Let $h : (-\epsilon, \epsilon) \times [a, b] \to M$ be a variation of a curve $\alpha := h_0$ such that $\left|\langle\dot{h}_s(t), \dot{h}_s(t)\rangle\right| > 0$ for all $s \in (-\epsilon, \epsilon)$. Then*

$$L'(s) = \int_a^b \varepsilon\langle\nabla_{\partial_s} \partial_t h(s, t), \partial_t h(s, t)\rangle \left(\varepsilon\langle\partial_t h(s, t), \partial_t h(s, t)\rangle\right)^{-1/2} dt$$

Proof.

$$\begin{aligned}
L'(s) &= \frac{d}{ds} \int_a^b \varepsilon\langle\dot{h}_s(t), \dot{h}_s(t)\rangle dt \\
&= \int_a^b \frac{d}{ds} \varepsilon\langle\dot{h}_s(t), \dot{h}_s(t)\rangle dt \\
&= \int_a^b 2\varepsilon\langle\nabla_{\partial_s} \dot{h}_s(t), \dot{h}_s(t)\rangle \frac{1}{2} \left(\varepsilon\langle\dot{h}_s(t), \dot{h}_s(t)\rangle\right)^{-1/2} dt \\
&= \int_a^b \varepsilon\langle\nabla_{\partial_s} \partial_t h(s, t), \partial_t h(s, t)\rangle \left(\varepsilon\langle\partial_t h(s, t), \partial_t h(s, t)\rangle\right)^{-1/2} dt \\
&= \varepsilon \int_a^b \langle\nabla_{\partial_t} \partial_s h(s, t), \partial_t h(s, t)\rangle \left(\varepsilon\langle\partial_t h(s, t), \partial_t h(s, t)\rangle\right)^{-1/2} dt
\end{aligned}$$

■

Corollary 20.2

$$\delta L|_\alpha (V) = L'(0) = \varepsilon \int_a^b \langle\nabla_{\partial_t} V(t), \dot{\alpha}(t)\rangle \left(\varepsilon\langle\dot{\alpha}(t), \dot{\alpha}(t)\rangle\right)^{-1/2} dt.$$

Let us now consider a more general situation where $\alpha : [a, b] \to M$ is only piecewise smooth (but still continuous). Let us be specific by saying that there is a partition $a < t_1 < ... < t_k < b$ so that α is smooth on each $[t_i, t_{i+1}]$. A variation appropriate to this situation is a continuous map $h : (-\epsilon, \epsilon) \times [a, b] \to M$ with $h(0, t) = \alpha(t)$ such that h is smooth on each set of the form $(-\epsilon, \epsilon) \times [t_i, t_{i+1}]$.

This is what we shall mean by a piecewise smooth variation of a piecewise smooth curve. The velocity $\dot{\alpha}$ and the variation vector field $V(t) := \frac{\partial h(0,t)}{\partial s}$ are only piecewise smooth. At each "kink" point t_i we have the jump vector $\triangle\dot{\alpha}(t_i) := V(t_i+) - V(t_i-)$ which measure this discontinuity of $\dot{\alpha}$ at t_i. Using this notation we have the following theorem which gives the **first variation formula**:

Theorem 20.17 *Let $h : (-\epsilon, \epsilon) \times [a, b] \to M$ be a piecewise smooth variation of a piecewise smooth curve $\alpha : [a, b]$. If α has constant speed $c = (\varepsilon\langle\dot{\alpha}, \dot{\alpha}\rangle)^{1/2}$ and variation vector field V then*

$$\delta L|_{\alpha}(V) = L'(0) = -\frac{\varepsilon}{c}\int_a^b \langle\nabla_{\partial_t}\dot{\alpha}, V\rangle dt - \frac{\varepsilon}{c}\sum_{i=1}^{k}\langle\triangle V(t_i), V(t_i)\rangle + \frac{\varepsilon}{c}\left.\langle\dot{\alpha}, V\rangle\right|_a^b$$

Proof. Since $c = (\varepsilon\langle\dot{\alpha}, \dot{\alpha}\rangle)^{1/2}$ the proposition 20.2 gives $L'(0) = \frac{\varepsilon}{c}\int_a^b \langle\nabla_{\partial_t} V(t), \dot{\alpha}(t)\rangle dt$. Now since we have $\langle\dot{\alpha}, \nabla_{\partial_t} V\rangle = \frac{d}{dt}\langle\dot{\alpha}, V\rangle - \langle\nabla_{\partial_t}\dot{\alpha}, V\rangle$ we can employ integration by parts: On each interval $[t_i, t_{i+1}]$ we have

$$\frac{\varepsilon}{c}\int_{t_i}^{t_{i+1}} \langle\nabla_{\partial_t} V, \dot{\alpha}\rangle dt = \frac{\varepsilon}{c}\left.\langle\dot{\alpha}, V\rangle\right|_{t_i}^{t_{i+1}} - \frac{\varepsilon}{c}\int_{t_i}^{t_{i+1}} \langle\nabla_{\partial_t}\dot{\alpha}, V\rangle dt.$$

Taking the convention that $t_0 = a$ and $t_{k+1} = b$ we sum from $i = 0$ to $i = k$ to get

$$\frac{\varepsilon}{c}\int_a^b \langle\dot{\alpha}, \nabla_{\partial_t} V\rangle dt = \frac{\varepsilon}{c}\langle\dot{\alpha}, V\rangle - \frac{\varepsilon}{c}\sum_{i=1}^{k}\langle\triangle\dot{\alpha}(t_i), V(t_i)\rangle$$

which equivalent to the result. ■

A variation $h : (-\epsilon, \epsilon) \times [a, b] \to M$ of α is called a **fixed endpoint variation** if $h(s, a) = \alpha(a)$ and $h(s, b) = \alpha(b)$ for all $s \in (-\epsilon, \epsilon)$. In this situation the variation vector field V is zero at a and b.

Corollary 20.3 *A piecewise smooth curve $\alpha : [a, b] \to M$ with constant speed $c > 0$ on each subinterval where α is smooth is a (nonnull) geodesic if and only if $\delta L|_{\alpha}(V) = 0$ for all fixed end point variations of α. In particular, if M is a Riemannian manifold and $\alpha : [a, b] \to M$ minimizes length among nearby curves then α is an (unbroken) geodesic.*

Proof. If α is a geodesic then it is smooth and so $\triangle\dot{\alpha}(t_i) = 0$ for all t_i (even though α is smooth the variation still only need to be piecewise smooth). It follows that $L'(0) = 0$.

Now if we suppose that α is a piecewise smooth curve and that $L'(0) = 0$ for any variation then we can conclude that α is a geodesic by picking some clever variations. As a first step we show that $\alpha|_{[t_i, t_{i+1}]}$ is a geodesic for each segment $[t_i, t_{i+1}]$. Let $t \in (t_i, t_{i+1})$ be arbitrary and let v be any nonzero vector in $T_{\alpha(t)}M$. Let b be a cut-off function on $[a, b]$ with support in $(t - \delta, t + \delta)$ and δ chosen small and then let $V(t) := b(t)Y(t)$ where Y is the parallel translation of

y along α. We can now easily produce a fixed end point variation with variation vector field V by the formula

$$h(s,t) := \exp_{\alpha(t)} sV(t).$$

With this variation the last theorem gives

$$L'(0) = -\frac{\varepsilon}{c}\int_a^b \langle \nabla_{\partial_t}\dot{\alpha}, V\rangle dt = -\frac{\varepsilon}{c}\int_{t-\delta}^{t+\delta} \langle \nabla_{\partial_t}\dot{\alpha}, b(t)Y(t)\rangle dt$$

which must hold no matter what our choice of y and for any $\delta > 0$. From this it is straightforward to show that $\nabla_{\partial_t}\dot{\alpha}(t) = 0$ and since t was an arbitrary element of (t_i, t_{i+1}) we conclude that $\alpha|_{[t_i,t_{i+1}]}$ is a geodesic. All that is left is to show that there can be no discontinuities of $\dot{\alpha}$ (recall exercise 20.12). One again we choose a vector y but this time $y \in T_{\alpha(t_i)}M$ where t_i is a potential kink point. take another cut-off function b with $\operatorname{supp} b \subset [t_{i-1}, t_{i+1}] = [t_{i-1}, t_i] \cup [t_i, t_{i+1}]$, $b(t_i) = 1$, and i a fixed but arbitrary element of $\{1, 2, ..., k\}$. Extend y to a field Y as before and let $V = bY$. Since we now have that α is a geodesic on each segment and we are assume the variation is zero, the first variation formula for any variation with variation vector field V reduces to

$$0 = L'(0) = -\frac{\varepsilon}{c}\langle \triangle\dot{\alpha}(t_i), y\rangle$$

for all y. This means that $\triangle\dot{\alpha}(t_i) = 0$ and since i was arbitrary we are done. ■

We now see that for fixed endpoint variations $L'(0) = 0$ implies that α is a geodesic. The geodesics are the critical "points" (or curves) of the arc length functional restricted to all curves with fixed endpoints. In order to classify the critical curves we look at the second variation but we only need the formula for variations of geodesics. For a variation h of a geodesic γ we have the variation vector field V as before but we also now consider the transverse acceleration vector field $A(t) := \nabla_{\partial_s}\partial_s h(0,t)$. Recall that for a curve γ with $|\langle\dot{\gamma},\dot{\gamma}\rangle| > 0$ a vector field Y along γ has an orthogonal decomposition $Y = Y^{\top} + Y^{\perp}$ (tangent and normal to γ). Also we have $(\nabla_{\partial_t}Y)^{\perp} = (\nabla_{\partial_t}Y)^{\perp}$ and so we can use $\nabla_{\partial_t}Y^{\perp}$ to denote either of these without ambiguity.

We now have the **second variation formula of Synge**:

Theorem 20.18 *Let $\gamma : [a,b] \to M$ be a (nonnull) geodesic of speed $c > 0$. Let $\varepsilon = sgn\langle\dot{\gamma},\dot{\gamma}\rangle$ as before. If $h : (-\epsilon,\epsilon) \times [a,b]$ is a variation of γ with variation vector field V and acceleration vector field A then the second variation of $L(s) := L(h_s(t))$ at $s = 0$ is*

$$\begin{aligned} L''(0) &= \frac{\varepsilon}{c}\int_a^b \left(\langle\nabla_{\partial_t}Y^{\perp}, \nabla_{\partial_t}Y^{\perp}\rangle + \langle R_{\dot{\gamma},V}\dot{\gamma}, V\rangle\right) dt + \frac{\varepsilon}{c}\left.\langle\dot{\gamma}, A\rangle\right|_a^b \\ &= \frac{\varepsilon}{c}\int_a^b \left(\langle\nabla_{\partial_t}Y^{\perp}, \nabla_{\partial_t}Y^{\perp}\rangle + \langle R_{\dot{\gamma}}(V), V\rangle\right) dt + \frac{\varepsilon}{c}\left.\langle\dot{\gamma}, A\rangle\right|_a^b \end{aligned}$$

Proof. Let $H(s,t) := \left|\langle \frac{\partial h}{\partial s}(s,t), \frac{\partial h}{\partial s}(s,t)\rangle\right|^{1/2} = \left(\varepsilon\langle \frac{\partial h}{\partial s}(s,t), \frac{\partial h}{\partial s}(s,t)\rangle\right)^{1/2}$. We have $L'(s) = \int_a^b \frac{\partial}{\partial s^2} H(s,t)dt$. Now computing as before we see that $\frac{\partial H(s,t)}{\partial s} = \frac{\varepsilon}{H}\langle \frac{\partial \gamma}{\partial s}(s,t), \nabla_{\partial_s}\frac{\partial \gamma}{\partial t}(s,t)\rangle$. Taking another derivative we have

$$\begin{aligned}
\frac{\partial^2 H(s,t)}{\partial s^2} &= \frac{\varepsilon}{H^2}\left(H\frac{\partial}{\partial s}\langle \frac{\partial h}{\partial t}, \nabla_{\partial_s}\frac{\partial h}{\partial t}\rangle - \langle \frac{\partial h}{\partial t}, \nabla_{\partial_s}\frac{\partial h}{\partial t}\rangle \frac{\partial H}{\partial s}\right) \\
&= \frac{\varepsilon}{H}\left(\langle \nabla_{\partial_s}\frac{\partial h}{\partial t}, \nabla_{\partial_s}\frac{\partial h}{\partial t}\rangle + \langle \frac{\partial h}{\partial t}, \nabla^2_{\partial_s}\frac{\partial h}{\partial t}\rangle - \frac{1}{H}\langle \frac{\partial h}{\partial t}, \nabla_{\partial_s}\frac{\partial h}{\partial t}\rangle \frac{\partial H}{\partial s}\right) \\
&= \frac{\varepsilon}{H}\left(\langle \nabla_{\partial_s}\frac{\partial h}{\partial t}, \nabla_{\partial_s}\frac{\partial h}{\partial t}\rangle + \langle \frac{\partial h}{\partial t}, \nabla^2_{\partial_s}\frac{\partial h}{\partial t}\rangle - \frac{\varepsilon}{H}\langle \frac{\partial h}{\partial t}, \nabla_{\partial_s}\frac{\partial h}{\partial t}\rangle^2\right).
\end{aligned}$$

Now using $\nabla_{\partial_t}\partial_s h = \nabla_{\partial_s}\partial_t h$ and lemma 20.7 we obtain

$$\nabla_{\partial_s}\nabla_{\partial_s}\frac{\partial h}{\partial t} = \nabla_{\partial_s}\nabla_{\partial_t}\frac{\partial h}{\partial s} = R(\frac{\partial h}{\partial s}, \frac{\partial h}{\partial t})\frac{\partial h}{\partial s} + \nabla_{\partial_t}\nabla_{\partial_s}\frac{\partial h}{\partial s}$$

and then

$$\begin{aligned}
\frac{\partial^2 H}{\partial s^2} = \frac{\varepsilon}{H}\Bigg\{&\langle \nabla_{\partial_t}\frac{\partial \gamma}{\partial s}, \nabla_{\partial_t}\frac{\partial \gamma}{\partial s}\rangle + \langle \frac{\partial h}{\partial t}, R(\frac{\partial h}{\partial s}, \frac{\partial h}{\partial t})\frac{\partial h}{\partial s}\rangle \\
&+\langle \frac{\partial h}{\partial t}, \nabla_{\partial_t}\nabla_{\partial_s}\frac{\partial h}{\partial s}\rangle - \frac{\varepsilon}{H^2}\langle \frac{\partial h}{\partial t}, \nabla_{\partial_t}\frac{\partial h}{\partial s}\rangle^2\Bigg\}
\end{aligned}$$

Now we let $s = 0$ and get

$$\begin{aligned}
\frac{\partial^2 H}{\partial s^2}(0,t) = \frac{\varepsilon}{c}\Big\{&\langle \nabla_{\partial_t}V, \nabla_{\partial_t}V\rangle + \langle \dot\gamma, R(V,\dot\gamma)V\rangle \\
&+\langle \dot\gamma, \nabla_{\partial_t}A\rangle - \frac{\varepsilon}{c^2}\langle \dot\gamma, \nabla_{\partial_t}V\rangle^2\Big\}
\end{aligned}$$

Now before we integrate the above expression we use the fact that $\langle \dot\gamma, \nabla_{\partial_t}A\rangle = \frac{d}{dt}\langle \dot\gamma, A\rangle$ (γ is a geodesic) and the fact that the orthogonal decomposition of $\frac{\nabla V}{dt}$ is

$$\nabla_{\partial_t}V = \frac{\varepsilon}{c^2}\langle \dot\gamma, \nabla_{\partial_t}V\rangle\dot\gamma + \nabla_{\partial_t}V^{\perp}$$

so that $\langle \nabla_{\partial_t}V, \nabla_{\partial_t}V\rangle = \frac{\varepsilon}{c^2}\langle \dot\gamma, \nabla_{\partial_t}V\rangle^2 + \langle \nabla_{\partial_t}V^{\perp}, \nabla_{\partial_t}V^{\perp}\rangle$. Plugging these identities in, observing the cancellation, and integrating we get

$$\begin{aligned}
L''(0) &= \int_a^b \frac{\partial^2 H}{\partial s^2}(0,t)dt \\
&= \frac{\varepsilon}{c}\int_a^b \left(\langle \nabla_{\partial_t}V^{\perp}, \nabla_{\partial_t}V^{\perp}\rangle + \langle \dot\gamma, R(V,\dot\gamma)V\right) \\
&+ \frac{\varepsilon}{c}\langle \dot\gamma, A\rangle\Big|_a^b
\end{aligned}$$

■

It is important to notice that the term $\frac{\varepsilon}{c}\langle \dot\gamma, A\rangle\big|_a^b$ depends not just on A but also on the curve itself. Thus the right hand side of the main equation of the second variation formula just proved depends only on V and A except for the last term. But if the variation is a fixed endpoint variation then this dependence drops out.

For our present purposes we will not loose anything by assume that $a = 0$. On the other hand it will pay to refrain from assuming $b = 1$.It is traditional to think of the set $\Omega_{0,b}(p,q)$ of all piecewise smooth curves $\alpha : [0,b] \to M$ from p to q as an infinite dimensional manifold. Then a variation vector field V along a curve $\alpha \in \Omega(p,q)$ which is zero at the endpoints is the tangent at α to the curve in $\Omega_{0,b}(p,q)$ given by the corresponding fixed endpoint variation h. Thus the "tangent space" $T(\Omega) = T_\alpha(\Omega_{0,b}(p,q))$ at α is the set of all piecewise smooth vector fields V along α such that $V(0) = V(b) = 0$. We then think of L as being a function on $\Omega_{0,b}(p,q)$ whose (nonnull[2]) critical points we have discovered to be nonnull geodesics beginning at p and ending at q at times 0 and b respectively. Further thinking along these lines leads to the idea of the index form.

Definition 20.37 *For a given nonnull geodesic $\gamma : [0,b] \to M$, the index form $I_\gamma : T_\gamma\Omega_{0,b} \times T_\gamma\Omega_{0,b} \to \mathbb{R}$ is defined by $I_\gamma(V,V) = L''_\gamma(0)$ where $L_\gamma(s) = \int_0^b \left|\langle \dot{h}_s(t), \dot{h}_s(t)\rangle\right| dt$ and $\frac{\nabla h}{\partial s}(0,t) = V$.*

Of course this definition makes sense because $L''_\gamma(0)$ only depends on V and not on h itself. Also, we have defined the quadratic form $I_\gamma(V,V)$ but not directly $I_\gamma(V,W)$. On the other hand, polarizations gives a but if $V, W \in T_\gamma\Omega_{0,b}$ then it is not hard to see from the second variation formula that

$$I_\gamma(V,W) = \frac{\varepsilon}{c}\int_0^b \left\{\langle \nabla_{\partial_t}V^\perp, \nabla_{\partial_t}W^\perp\rangle + \langle R(V,\dot\gamma)W\rangle, \dot\gamma\right\} dt \tag{20.10}$$

It is important to notice that the right hand side of the above equation is in fact symmetric in V and W. It is also not hard to prove that if even one of V or W is tangent to $\dot\gamma$ then $I_\gamma(V,W) = 0$ and so $I_\gamma(V,W) = I_\gamma(V^\perp, W^\perp)$ and so we may as well restrict I_γ to

$$T_\gamma^\perp\Omega_{0,b} = \{V \in T_\gamma\Omega_{0,b} : V \perp \dot\gamma\}.$$

This restriction will be denoted by $I_\gamma^\perp$.

It is important to remember that the variations and variation vector fields we are dealing with are allowed to by only piecewise smooth even if the center curve is smooth. So let $0 = t_0 < t_1 < ... < t_k < t_{k+1} = b$ as before and let V and W be vector fields along a geodesic γ. We now derive another formula for $I_\gamma(V,W)$. Rewrite formula20.10 as

$$I_\gamma(V,W) = \frac{\varepsilon}{c}\sum_{i=0}^{k}\int_{t_i}^{t_{i+1}} \left\{\langle \nabla_{\partial_t}V^\perp, \nabla_{\partial_t}W^\perp\rangle + \langle R(V,\dot\gamma)W, \dot\gamma\rangle\right\} dt$$

On each interval $[t_i, t_{i+1}]$ we have

$$\langle \nabla_{\partial_t}V^\perp, \nabla_{\partial_t}W^\perp\rangle = \nabla_{\partial_t}\langle \nabla_{\partial_t}V^\perp, W^\perp\rangle - \langle \nabla^2_{\partial_t}V^\perp, W^\perp\rangle$$

[2]By nonnull we just mean that the geodesic is nonnull.

and substituting this into the above formula we have

$$I_\gamma(V,W) = \frac{\varepsilon}{c}\sum_{i=0}^{k}\int_{t_i}^{t_{i+1}} \left\{\nabla_{\partial_t}\langle\nabla_{\partial_t}V^\perp, W^\perp\rangle - \langle\nabla^2_{\partial_t}V^\perp, W^\perp\rangle + \langle R(V,\dot\gamma)W,\dot\gamma\rangle\right\}dt.$$

As for the curvature term we use

$$\begin{aligned}\langle R(V,\dot\gamma)W,\dot\gamma\rangle &= \langle R(\dot\gamma,V)\dot\gamma,W\rangle \\ &= \langle R(\dot\gamma,V^\perp)\dot\gamma,W^\perp\rangle.\end{aligned}$$

Substituting we get

$$I_\gamma(V,W) = \frac{\varepsilon}{c}\sum_{i=0}^{k}\int_{t_i}^{t_{i+1}} \left\{\nabla_{\partial_t}\langle\nabla_{\partial_t}V^\perp, W^\perp\rangle - \langle\nabla^2_{\partial_t}V^\perp, W^\perp\rangle + \langle R(\dot\gamma,V^\perp)\dot\gamma,W^\perp\rangle\right\}dt.$$

Using the fundamental theorem of calculus on each interval $[t_i, t_{i+1}]$ and the fact that W vanishes at a and b we obtain alternate formula:

Proposition 20.10 (Formula for Index Form) *Let $\gamma : [0,b] \to M$ be a nonnull geodesic. Then*

$$I_\gamma(V,W) = -\frac{\varepsilon}{c}\int_0^b \langle\nabla^2_{\partial_t}V^\perp + R(\dot\gamma,V^\perp)\dot\gamma, W^\perp\rangle dt - \frac{\varepsilon}{c}\sum_{i=1}^{k}\langle\triangle\nabla_{\partial_t}V^\perp(t_i), W^\perp(t_i)\rangle$$

where $\triangle\nabla_{\partial_t}V^\perp(t_i) = \nabla_{\partial_t}V^\perp(t_i+) - \nabla_{\partial_t}V^\perp(t_i-)$.

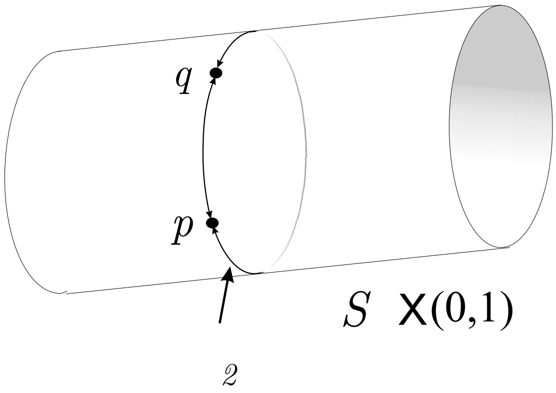
Letting $V = W$ we have

$$I_\gamma(V,V) = -\frac{\varepsilon}{c}\int_0^b \langle\nabla^2_{\partial_t}V^\perp + R(\dot\gamma,V^\perp)\dot\gamma, V^\perp\rangle dt - \frac{\varepsilon}{c}\sum_{i=1}^{k}\langle\triangle\nabla_{\partial_t}V^\perp(t_i), V^\perp(t_i)\rangle$$

and the presence of the term $R(\dot\gamma,V^\perp)\dot\gamma,V^\perp\rangle$ indicates a connection with Jacobi fields.

Definition 20.38 *A geodesic segment $\gamma : [a,b] \to M$ is said to be **relatively length minimizing** if for all piecewise smooth fixed endpoint variations h of γ the function $L(s) := \int_a^b \left|\langle\dot h_s(t),\dot h_s(t)\rangle\right| dt$ has a local minimum at $s = 0$ (where $\gamma = h_0(t) := h(0,t)$).*

If $\gamma : [a,b] \to M$ is a relatively length minimizing nonnull geodesic then L is at least twice differentiable at $s = 0$ and clearly $L''(0) = 0$ which means that $I_\gamma(V,V) = 0$ for any $V \in T_\gamma\Omega_{a,b}$. The adjective "relatively" is included in the terminology because of the possibility that there may be curve in $\Omega_{a,b}$ which are "far away from γ" which has smaller length than γ. A simple example of this is depicted in the figure below where γ_2 has greater length than γ even though γ_2 is relatively length minimizing. We are assume that the metric on $(0,1)\times S^1$ is the



usual definite metric dx^2+dy^2 induced from that on $\mathbb{R}\times(0,1)$ where we identify $S^1\times(0,1)$ with the quotient $\mathbb{R}\times(0,1)/((x,y)\sim(x+2\pi,y))$. On the other hand, one sometimes hears the statement that geodesics in a Riemannian manifold are **locally length minimizing**. This means that for any geodesic $\gamma:[a,b]\to M$ the restrictions to small intervals $[a,\epsilon]$ is always relatively length minimizing but this is only true for Riemannian manifolds. For a semi-Riemannian manifold with indefinite metric a small geodesic segment can have nearby curves that increase the length. To see an example of this consider the metric $-dx^2+dy^2$ on $\mathbb{R}\times(0,1)$ and the induced metric on the quotient $S^1\times(0,1)=\mathbb{R}\times(0,1)/\sim$. In this case, the geodesic γ has length less than all nearby geodesics; the index form I_γ is now negative semidefinite.

Exercise 20.18 *Prove the above statement concerning I_γ for $S^1\times(0,1)$ with the index 1 metric $-dx^2+dy^2$.*

Exercise 20.19 *Let M be a Riemannian manifold. Show that if $I_\gamma(V,V)=0$ for all $V\in T_\gamma^\perp\Omega_{a,b}$ then a nonnull geodesic $\gamma$is relatively length minimizing.*

Theorem 20.19 *Let $\gamma:[0,b]\to M$ be a nonnull geodesic. The nullspace $\mathcal{N}(I_\gamma^\perp)$(defffine!!) of $I_\gamma^\perp:T_\gamma^\perp\Omega_{0,b}\to R$ is exactly the space $\mathcal{J}_0(\gamma,0,b)$ of Jacobi fields vanishing at $\gamma(0)$ and $\gamma(b)$.*

Proof. It follow from the formula of proposition 20.10 makes it clear that $\mathcal{J}_0(\gamma,0,b)\subset\mathcal{N}(I_\gamma^\perp)$.

Suppose that $V\in\mathcal{N}(I_\gamma^\perp)$. Let $t\in(t_i,t_{i+1})$ where the t_i is the partition of V. Pick an arbitrary nonzero element $y\in T_{\gamma(t)}M$ and let Y be the unique parallel field along $\gamma|_{[t_i,t_{i+1}]}$ such that $Y(t)=y$. Picking a cut-off function β with support in $[t+\delta,t_i-\delta]\subset(t_i,t_{i+1})$ as before we extend βY to a field W along γ with $W(t)=y$. Now V is normal to the geodesic and so $I_\gamma(V,W)=I_\gamma^\perp(V,W)$ and

$$I_\gamma(V,W)=-\frac{\varepsilon}{c}\int_{t-\delta}^{t+\delta}\langle\nabla_{\partial_t}^2V+R(\dot\gamma,V)\dot\gamma,\beta Y^\perp\rangle dt$$

for small δ, $\beta Y^{\perp}$ is approximately the arbitrary nonzero y and it follows that $\nabla^2_{\partial_t} V + R(\dot{\gamma}, V)\dot{\gamma}$ is zero at t since t was arbitrary it is identically zero on (t_i, t_{i+1}). Thus V is a Jacobi field on each interval (t_i, t_{i+1}) and since $V$is continuous on $[0, b]$ it follows from standard theory of differential equations that V is a smooth Jacobi field along all of γ and since $V \in T_{\gamma(t)}M$ we already have $V(0) = V(b) = 0$ so $V \in \mathcal{J}_0(\gamma, 0, b)$. ■

Proposition 20.11 *Let M, g be a semi-Riemannian manifold of index $\nu = \operatorname{ind}(M)$ and $\gamma : [a, b] \to M$ a nonnull geodesic. If the index I_γ is positive semidefinite then $\nu = 0$ or n (thus the metric is definite and so, up to sign of g, the manifold is Riemannian). On the other hand, if I_γ is negative semidefinite then $\nu = 1$ or $n - 1$ (so that up to sign convention M is a Lorentz manifold).*

Proof. Let I_γ be positive semi-definite and assume that $0 < \nu < n$. In this case there must be a unit vector u in $T_{\gamma(a)}M$ which is normal to $\dot{\gamma}(0)$ and has the opposite causal character of $\dot{\gamma}(a)$. This means that if $\varepsilon = \langle \dot{\gamma}(a), \dot{\gamma}(a)\rangle$ then $\varepsilon\langle u, u\rangle = -1$. Let U be the field along γ which is the parallel translation of u. By choosing $\delta > 0$ appropriately we can arrange that δ is as small as we like and simultaneously that $\sin(t/\delta)$ is zero at $t = a$ and $t = b$. Let $V := \delta \sin(t/\delta)U$ and make the harmless assumption that $|\dot{\gamma}| = 1$. Notice that by construction $V \perp \dot{\gamma}$. We compute:

$$
\begin{aligned}
I_\gamma(V, V) &= \varepsilon \int_a^b \{\langle \nabla_{\partial_t} V, \nabla_{\partial_t} V\rangle + \langle R(\dot{\gamma}, V)\dot{\gamma}, V\rangle\}\, dt \\
&= \varepsilon \int_a^b \{\langle \nabla_{\partial_t} V, \nabla_{\partial_t} V\rangle + \langle R(\dot{\gamma}, V)V, \dot{\gamma}\rangle\}\, dt \\
&\varepsilon \int_a^b \{\langle \nabla_{\partial_t} V, \nabla_{\partial_t} V\rangle + K(V \wedge \dot{\gamma})\langle V \wedge \dot{\gamma}, V \wedge \dot{\gamma}\rangle\}\, dt \\
&= \varepsilon \int_a^b \{\langle \nabla_{\partial_t} V, \nabla_{\partial_t} V\rangle + K(V \wedge \dot{\gamma})\langle V, V\rangle\varepsilon\}\, dt
\end{aligned}
$$

where $K(V \wedge \dot{\gamma}) := \frac{\langle \mathfrak{R}(V\wedge\dot{\gamma}), V\wedge\dot{\gamma}\rangle}{\langle V\wedge\dot{\gamma}, V\wedge\dot{\gamma}\rangle} = \frac{\langle \mathfrak{R}(V\wedge\dot{\gamma}), V\wedge\dot{\gamma}\rangle}{\langle V, V\rangle^2}$ defined earlier. Continuing the computation we have

$$
\begin{aligned}
I_\gamma(V, V) &= \varepsilon \int_a^b \left\{\langle u, u\rangle \cos^2(t/\delta) + K(V \wedge \dot{\gamma})\delta^2 \sin^2(t/\delta)\right\} dt \\
&= \int_a^b \left\{-\cos^2(t/\delta) + \varepsilon K(V \wedge \dot{\gamma})\delta^2 \sin^2(t/\delta)\right\}
\end{aligned}
$$

Not as we said, we can choose δ as small as we like and since $K(V(t) \wedge \dot{\gamma}(t))$ is bounded on the (compact) interval this clearly means that $I_\gamma(V, V) < 0$ which contradicts the fact that I_γ is positive semidefinite. Thus our assumption that $0 < \nu < n$ is impossible.

Now let I_γ be negative semi-definite. Suppose that we assume that contrary to what we wish to show, ν is not 1 or $n - 1$. In this case one can find a unit vector $u \in T_{\gamma(a)}M$ normal to $\dot{\gamma}(a)$ such that $\varepsilon\langle u, u\rangle = +1$. The the same sort of calculation as we just did shows that I_γ cannot be semi-definite; again a contradiction. ■

By changing the sign of the metric the cases handled by this last theorem boil down to the two important cases 1) where M, g is Riemannian, γ is arbitrary and 2) where M, g is Lorentz and γ is timelike. We consolidate these two cases by a definition:

Definition 20.39 *A geodesic $\gamma : [a, b] \to M$ is cospacelike if the subspace $\dot{\gamma}(s)^\perp \subset T_{\gamma(s)}M$ is spacelike for some (and consequently all) $s \in [a, b]$.*

Exercise 20.20 *Show that if $\gamma : [a, b] \to M$ is cospacelike then γ is nonnull, $\dot{\gamma}(s)^\perp \subset T_{\gamma(s)}M$ is spacelike for all $s \in [a, b]$ and also show that M, g is either Riemannian of Lorentz.*

A useful observation about Jacobi fields along a geodesic is the following:

Lemma 20.8 *If we have two Jacobi fields, say J_1 and J_2 along a geodesic γ then $\langle \nabla_{\partial_t} J_1, J_2\rangle - \langle J_1, \nabla_{\partial_t} J_2\rangle$ is constant along γ.*

To see this we note that

$$\begin{aligned}
\nabla_{\partial_t}\langle \nabla_{\partial_t} J_1, J_2\rangle &= \langle \nabla^2_{\partial_t} J_1, J_2\rangle + \langle \nabla_{\partial_t} J_1, \nabla_{\partial_t} J_2\rangle \\
&= \langle R(J_{1,}\dot{\gamma})J_2, \dot{\gamma}\rangle + \langle \nabla_{\partial_t} J_1, \nabla_{\partial_t} J_2\rangle \\
&= \langle R(J_2\dot{\gamma})J_{1,}, \dot{\gamma}\rangle + \langle \nabla_{\partial_t} J_2, \nabla_{\partial_t} J_1\rangle \\
&= \nabla_{\partial_t}\langle \nabla_{\partial_t} J_2, J_1\rangle.
\end{aligned}$$

In particular, if $\langle \nabla_{\partial_t} J_1, J_2\rangle = \langle J_1, \nabla_{\partial_t} J_2\rangle$ at $t = 0$ then $\langle \nabla_{\partial_t} J_1, J_2\rangle - \langle J_1, \nabla_{\partial_t} J_2\rangle = 0$ for all t.

We now need another simple but rather technical lemma.

Lemma 20.9 *If $J_1, ..., J_k$ are Jacobi fields along a geodesic γ such that $\langle \nabla_{\partial_t} J_i, J_j\rangle = \langle J_i, \nabla_{\partial_t} J_j\rangle$ for all $i, j \in \{1, ..., k\}$ then and field Y which can be written $Y = \varphi^i J_i$ has the property that*

$$\langle \nabla_{\partial_t} Y, \nabla_{\partial_t} Y\rangle + \langle R(Y, \dot{\gamma})Y, \dot{\gamma}\rangle = \langle \left(\partial_t \varphi^i\right) J_i, \left(\partial_t \varphi^i\right) J_i\rangle + \partial_t\langle Y, \varphi^r \left(\nabla_{\partial_t} J_r\right)\rangle$$

Proof. $\nabla_{\partial_t} Y = \left(\partial_t \varphi^i\right) J_i + \varphi^r \left(\nabla_{\partial_t} J_r\right)$ and so

$$\begin{aligned}
\partial_t\langle Y, \varphi^r \left(\nabla_{\partial_t} J_r\right)\rangle &= \langle \left(\nabla_{\partial_t} Y\right), \varphi^r \left(\nabla_{\partial_t} J_r\right)\rangle + \langle \left(\partial_t \varphi^i\right) J_i, \nabla_{\partial_t} \left[\varphi^r \left(\nabla_{\partial_t} J_r\right)\right]\rangle \\
&= \langle \left(\partial_t \varphi^i\right) J_i, \varphi^r \left(\nabla_{\partial_t} J_r\right)\rangle + \langle \varphi^r \left(\nabla_{\partial_t} J_r\right), \varphi^r \left(\nabla_{\partial_t} J_r\right)\rangle \\
&\quad + \langle Y, \partial_t \varphi^r \nabla_{\partial_t} J_r\rangle + \langle Y, \varphi^r \nabla^2_{\partial_t} J_r\rangle
\end{aligned}$$

The last term $\langle Y, \varphi^r \nabla^2_{\partial_t} J_r\rangle$ equals $\langle R(Y, \dot{\gamma})Y, \dot{\gamma}\rangle$ by the Jacobi equation. Using this and the fact that $\langle Y, \partial_t \varphi^r \nabla_{\partial_t} J_r\rangle = \langle \left(\partial_t \varphi^i\right) J_i, \varphi^r \left(\nabla_{\partial_t} J_r\right)\rangle$ which follows from a short calculation using the hypotheses on the J_i we arrive at

$$\begin{aligned}
\partial_t\langle Y, \varphi^r \left(\nabla_{\partial_t} J_r\right)\rangle &= 2\langle \left(\partial_t \varphi^i\right) J_i, \varphi^r \left(\nabla_{\partial_t} J_r\right)\rangle + \langle \varphi^r \left(\nabla_{\partial_t} J_r\right), \varphi^r \left(\nabla_{\partial_t} J_r\right)\rangle \\
&\quad + \langle R(Y, \dot{\gamma})Y, \dot{\gamma}\rangle.
\end{aligned}$$

Using the last equation together with $\nabla_{\partial_t} Y = \left(\partial_t \varphi^i\right) J_i + \varphi^r \left(\nabla_{\partial_t} J_r\right)$ gives the result (check it!). ∎

Exercise 20.21 *Work through the details of the proof of the lemma above.*

Through out the following discussion $\gamma : [0, b] \to M$ will be a cospacelike geodesic with sign $\varepsilon := \langle \dot\gamma, \dot\gamma \rangle$.

Suppose that, for whatever reason, there are no conjugate points of $p = \gamma(0)$ along γ. There exist Jacobi fields $J_1, ..., J_{n-1}$ along γ which vanish at $t = 0$ and such that the vectors $\nabla_{\partial_t} J_1(0), ..., \nabla_{\partial_t} J_{n-1}(0) \in T_p M$ are a basis for the space $\dot\gamma(0)^\perp \subset T_{\gamma(0)} M$. We know that these J_i are all normal to the geodesic and since we are assuming that there are no conjugate points to p along γ it follows that the fields remain linearly independent and at each t with $0 < t \leq b$ form a basis of $\dot\gamma(t)^\perp \subset T_{\gamma(t)} M$. Now let $Y \in T_\gamma(\Omega)$ be a piecewise smooth variation vector field along γ and write $Y = \varphi^i J_i$ for some piecewise smooth functions φ^i on $(0, b]$ which can be show to extend continuously to $[0, b]$. Since $\langle \nabla_{\partial_t} J_i, J_j \rangle = \langle J_i, \nabla_{\partial_t} J_j \rangle = 0$ at $t = 0$ we have $\langle \nabla_{\partial_t} J_i, J_j \rangle - \langle J_i, \nabla_{\partial_t} J_j \rangle = 0$ for all t by lemma 20.8. This allows the use of the lemma 20.9 to arrive at

$$\langle \nabla_{\partial_t} Y, \nabla_{\partial_t} Y \rangle + \langle R(Y, \dot\gamma) Y, \dot\gamma \rangle = \langle \left(\partial_t \varphi^i\right) J_i, \left(\partial_t \varphi^i\right) J_i \rangle + \partial_t \langle Y, \varphi^r \left(\nabla_{\partial_t} J_r\right) \rangle$$

and then

$$\varepsilon I_\gamma(Y, Y) = \frac{1}{c} \int_0^b \langle \left(\partial_t \varphi^i\right) J_i, \left(\partial_t \varphi^i\right) J_i \rangle dt + \frac{1}{c} \langle Y, \varphi^r \left(\nabla_{\partial_t} J_r\right) \rangle \Big|_0^b . \tag{20.11}$$

On the other hand, Y is zero at a and b and so the last term above vanishes. Now we notice that since γ is cospacelike and the acceleration field $\nabla_{\partial_t} \dot\gamma$ is normal to the geodesic we must have $\langle \nabla_{\partial_t} \dot\gamma, \nabla_{\partial_t} \dot\gamma \rangle \geq 0$ (Exercise: prove this last statement). Now by equation 20.11 above we conclude that $\varepsilon I_\gamma(Y, Y) \geq 0$. On the other hand, if $I_\gamma(Y, Y) = 0$ identically then $\int_0^b \langle \left(\partial_t \varphi^i\right) J_i, \left(\partial_t \varphi^i\right) J_i \rangle dt = 0$ and $\langle \left(\partial_t \varphi^i\right) J_i, \left(\partial_t \varphi^i\right) J_i \rangle = 0$. In turn this implies that $A = 0$ and that each φ^i is zero and finally that Y itself is identically zero along γ. A moments thought shows that all we have assumed about Y is that it is in the domain of the restricted index $I_\gamma^\perp$ and so we have proved the following:

Proposition 20.12 *If $\gamma \in \Omega_{0,b}(\gamma)$ is cospacelike and there is no conjugate points to $p = \gamma(0)$ along γ then $\varepsilon I_\gamma^\perp(Y, Y) \geq 0$ and $Y = 0$ along γ if and only if $I_\gamma^\perp(Y, Y) = 0$.*

We may paraphrase the above result as follows: For a cospacelike geodesic γ without conjugate points, the restricted index for $I_\gamma^\perp$ is definite; positive definite if $\varepsilon = +1$ and negative definite if $\varepsilon = -1$. The first case ($\varepsilon = +1$) is exactly the case where M, g is Riemannian (exercise 20.20).

Next we consider the situation where the cospacelike geodesic $\gamma : [0, b] \to M$ is such that $\gamma(b)$ is the only point conjugate to $p = \gamma(0)$ along γ. In this case, theorem 20.19 tells use that $I_\gamma^\perp$ has a nontrivial nullspace and so I_γ cannot be definite. Claim: I_γ is semidefinite. To see this let $Y \in T_\gamma\Omega_{0,b}(\gamma)$ and write Y is the form $(b-t)Z(t)$ for some (continuous) piecewise smooth Z. Let $b_i \to b$ and define Y_i to be $(b_i - t)Z(t)$ on $[0, b_i]$. Our last proposition applied to $\gamma_i := \gamma|_{[0,b_i]}$ shows that $\varepsilon I_{\gamma_i}(Y_i, Y_i) \geq 0$. Now $\varepsilon I_{\gamma_i}(Y_i, Y_i) \to \varepsilon I_\gamma(Y, Y)$ (some uninteresting details are omitted) and so the claim is true.

Now we consider the case where there is a conjugate point to p before $\gamma(b)$. Suppose that J is a nonzero Jacobi field along $\gamma|_{[0,r]}$ with $0 < r < b$ such that $J(0) = J(r) = 0$. We can extend J to a field J_{ext} on $[0, b]$ by defining it to be 0 on $[r, b]$. Notice that $\nabla_{\partial_t} J_{ext}(r-)$ is equal to $\nabla_{\partial_t} J(r)$ which is not 0 since otherwise J would be identically zero (over determination). On the other hand, $\nabla_{\partial_t} J_{ext}(r+) = 0$ and so the "kink" $\triangle J'_{ext}(r) := \nabla_{\partial_t} J_{ext}(r+) - \nabla_{\partial_t} J_{ext}(r-)$ is not zero. We will now show that if $W \in T_\gamma(\Omega)$ such that $W(r) = \triangle J'_{ext}(r)$ (and there are plenty of such fields), then $\varepsilon I_\gamma(J_{ext} + \delta W, J_{ext} + \delta W) < 0$ for small enough $\delta > 0$. This will allow us to conclude that I_γ *cannot be definite* since by 20.11 we can always find a Z with $\varepsilon I_\gamma(Z, Z) > 0$. We have

$$\varepsilon I_\gamma(J_{ext} + \delta W, J_{ext} + \delta W) = \varepsilon I_\gamma(J_{ext}, J_{ext}) + 2\delta\varepsilon I_\gamma(J_{ext}, W) + \varepsilon\delta^2 I_\gamma(W, W)$$

Now it is not hard to see from the formula of theorem 20.10 that $I_\gamma(J_{ext}, J_{ext})$ is zero since it is Piecewise Jacobi and is zero at the single kink point r. But suing the formula again, $\varepsilon I_\gamma(J_{ext}(r), W(r))$ reduces to

$$-\frac{1}{c}\langle \triangle J'_{ext}(r), W(r)\rangle = -\frac{1}{c}\left|\triangle J'_{ext}(r)\right|^2 < 0$$

and so taking δ small enough gives the desired conclusion.

Summarizing the conclusion the above discussion (together with the result of proposition 20.12) we obtain the following nice theorem:

Theorem 20.20 *If $\gamma : [0, b] \to M$ is a cospacelike geodesic of sign $\varepsilon := \langle \dot\gamma, \dot\gamma\rangle$ then M, g is either Riemannian of Lorentz and we have the following three cases:*

(i) If there are no conjugate to $\gamma(0)$ along γ then $\varepsilon I_\gamma^\perp$ is definite (positive if $\varepsilon = 1$ and negative if $\varepsilon = -1$)

(ii) If $\gamma(b)$ is the only conjugate point to $\gamma(0)$ along γ then I_γ is not definite but must be semidefinite.

(iii) If there is a conjugate $\gamma(r)$ to $\gamma(0)$ with $0 < r < b$ then I_γ is not semidefinite (or definite).

Corollary 20.4 *If $\gamma : [0, b] \to M$ is a cospacelike geodesic of sign $\varepsilon := \langle \dot\gamma, \dot\gamma\rangle$ and suppose that Y is a vector field along γ and that J is a Jacobi field along γ such that*

$$\begin{aligned} Y(0) &= J(0) \\ Y(b) &= J(b) \\ (Y - J) &\perp \dot\gamma \end{aligned}$$

Then $\varepsilon I_\gamma(J,J) \leq \varepsilon I_\gamma(Y,Y)$.

Proof of Corollary. From the previous theorem we have $0 \leq \varepsilon I_\gamma^\perp(Y - J, Y - J) = \varepsilon I_\gamma(Y - J, Y - J)$ and so

$$0 \leq \varepsilon I_\gamma(Y,Y) - 2\varepsilon I_\gamma(J,Y) + \varepsilon I_\gamma(J,J)$$

On the other hand,

$$\begin{aligned}
\varepsilon I_\gamma(J,Y) &= \varepsilon \left.\langle \nabla_{\partial_t} J, Y\rangle\right|_0^b - \varepsilon \int_0^b \langle \nabla^2_{\partial_t} J, Y\rangle - \langle R_{\dot\gamma,J}\dot\gamma, J\rangle \\
&= \varepsilon \left.\langle \nabla_{\partial_t} J, Y\rangle\right|_0^b = \varepsilon \left.\langle \nabla_{\partial_t} J, J\rangle\right|_0^b \\
&= \varepsilon I_\gamma(J,J) \text{ (since } J \text{ is a Jacobi field)}
\end{aligned}$$

Thus $0 \leq \varepsilon I \varepsilon_\gamma(Y,Y) - 2\varepsilon I_\gamma(J,Y) + \varepsilon I_\gamma(J,J) = \varepsilon I_\gamma(Y,Y) - \varepsilon I_\gamma(J,J)$. ∎

As we mentioned the Jacobi equation can be written in terms of the tidal force operator: $R_v : T_pM \to T_pM$ as

$$\nabla^2_{\partial_t} J(t) = R_{\dot\gamma(t)}(J(t))$$

The meaning of the term force here is that $R_{\dot\gamma(t)}$ controls the way nearby families of geodesics attract or repel each other. Attraction tends to create conjugate points while repulsion tends to prevent conjugate points. If γ is cospacelike then if we take any unit vector u normal to $\dot\gamma(t)$ then we can look at the component of $R_{\dot\gamma(t)}(u)$ in the u direction; $\langle R_{\dot\gamma(t)}(u), u\rangle u = \langle R_{\dot\gamma(t),u}(\dot\gamma(t)), u\rangle u = -\langle \Re(\dot\gamma(t) \wedge u), \dot\gamma(t) \wedge u\rangle u$. In terms of sectional curvature

$$\langle R_{\dot\gamma(t)}(u), u\rangle u = K(\dot\gamma(t) \wedge u)\langle \dot\gamma(t), \langle \dot\gamma(t)\rangle.$$

It follows from the Jacobi equation that if $\langle R_{\dot\gamma(t)}(u), u\rangle \geq 0$, i.e. if $K(\dot\gamma(t) \wedge u)\langle \dot\gamma(t), \dot\gamma(t)\rangle \leq 0$ then we have repulsion and if this always happens anywhere along γ we expect that $\gamma(0)$ has no conjugate point along γ. This intuition is indeed correct:

Proposition 20.13 *Let $\gamma : [0,b] \to M$ be a cospacelike geodesic. If for every t and every vector $v \in \gamma(t)^\perp$ we have $\langle R_{\dot\gamma(t)}(v), v\rangle \geq 0$ (i.e. if $K(\dot\gamma(t) \wedge v)\langle \dot\gamma(t), \dot\gamma(t)\rangle \leq 0$) then $\gamma(0)$ has no conjugate point along γ.*

In particular, a Riemannian manifold with sectional curvature $K \leq 0$ has no conjugate pairs of points. Similarly, a Lorentz manifold with sectional curvature $K \geq 0$ has no conjugate pairs along any timelike geodesics.

Proof. Take J to be a Jacobi field along γ such that $J(0)$ and $J \perp \dot\gamma$. We have $\frac{d}{dt}\langle J, J\rangle = 2\langle \nabla_{\partial_t} J, J\rangle$ and

$$\begin{aligned}
\frac{d^2}{dt^2}\langle J, J\rangle &= 2\langle \nabla_{\partial_t} J, \nabla_{\partial_t} J\rangle + 2\langle \nabla^2_{\partial_t} J, J\rangle \\
&= 2\langle \nabla_{\partial_t} J, \nabla_{\partial_t} J\rangle + 2\langle R_{\dot\gamma(t),J}(\dot\gamma(t)), J\rangle \\
&= 2\langle \nabla_{\partial_t} J, \nabla_{\partial_t} J\rangle + 2\langle R_{\dot\gamma(t)}(J), J\rangle
\end{aligned}$$

and by the hypotheses $\frac{d^2}{dt^2}\langle J, J\rangle \geq 0$. On the other hand, we have $\langle J(0), J(0)\rangle = 0$ and $\left.\frac{d}{dt}\right|_0 \langle J, J\rangle = 0$. It follows that since $\langle J, J\rangle$ is not identically zero we must have $\langle J, J\rangle > 0$ for all $t \in (0, b]$. ∎

20.11 More Riemannian Geometry

Recall that a manifold is geodesically complete at p if and only if $\exp_p$ is defined on all of T_pM. The following lemma is the essential ingredient in the proof of the Hopf-Rinow theorem stated and proved below. In fact, this lemma itself is sometimes referred to as the Hopf-Rinow theorem. Note that this is a theorem about Riemannian manifolds.

Lemma 20.10 *Let M, g be a (finite dimensional) connected Riemannian manifold. Suppose that $\exp_p$ is defined on the ball $B_\rho(p)$ for $\rho > 0$. Then each point $q \in B_\rho(p)$ can be connected to p by an absolutely minimizing geodesic. In particular, if M is geodesically complete at $p \in M$ then each point $q \in M$ can be connected to p by an absolutely minimizing geodesic.*

Proof. Let $q \in B_\rho(p)$ with $p \neq q$ and let $R = dist(p, q)$. Choose $\epsilon > 0$ small enough that $B_{2\epsilon}(p)$ is the domain of a normal coordinate system. By lemma 20.6 we already know the theorem is true if $B_\rho(p) \subset B_\epsilon(p)$ so we will assume that $\epsilon < R < \rho$. Because $\partial B_\epsilon(p)$ is diffeomorphic to $S^{n-1} \subset \mathbb{R}^n$ it is compact and so there is a point $p_\epsilon \in \partial B_\epsilon(p)$ such that $x \mapsto dist(x, q)$ achieves its minimum at p_ϵ. This means that

$$\begin{aligned} dist(p,q) &= dist(p,p_\epsilon) + dist(p_\epsilon, q) \\ &= \epsilon + dist(p_\epsilon, q). \end{aligned}$$

Let $\gamma : [0, \rho] \to M$ be the constant speed geodesic with $|\dot\gamma| = 1$, $\gamma(0) = p$, and $\gamma(\epsilon) = p_\epsilon$. It is not difficult to see that the set

$$T = \{t \in [0, R] : dist(p, \gamma(t)) + dist(\gamma(t), q)$$

is closed in $[0, R]$ and is nonempty since $\epsilon \in T$. Let $t_{\sup} = \sup T > 0$. We will show that $t_{\max} = R$ from which it will follow that $\gamma|_{[0,R]}$ is a minimizing geodesic from p to q. With an eye toward a contradiction, assume that $t_{\sup} < R$. Let $x := \gamma(t_{\sup})$ and choose ϵ_1 with $0 < \epsilon_1 < R - t_{\sup}$ and small enough that $B_{2\epsilon_1}(x)$ is the domain of normal coordinates about x. Arguing as before we see that there must be a point $x_{\epsilon_1} \in \partial B_{\epsilon_1}(x)$ such that

$$dist(x, q) = dist(x, x_{\epsilon_1}) + dist(x_{\epsilon_1}, q) = \epsilon_1 + dist(x_{\epsilon_1}, q).$$

Now let γ_1 be the unit speed geodesic such that $\gamma_1(0) = x$ and $\gamma_1(\epsilon_1) = x_{\epsilon_1}$. Combining, we now have

$$dist(p, q) = dist(p, x) + dist(x, x_{\epsilon_1}) + dist(x_{\epsilon_1}, q).$$

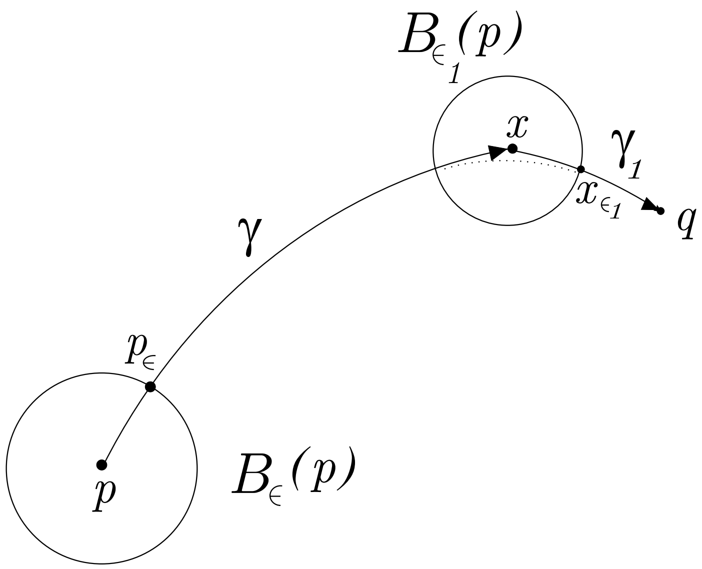
By the triangle inequality $dist(p, q) \leq dist(p, x_{\epsilon_1}) + dist(x_{\epsilon_1}, q)$ and so

$$dist(p, x) + dist(x, x_{\epsilon_1}) \leq dist(p, x_{\epsilon_1}).$$

But also $dist(p, x_{\epsilon_1}) \leq dist(p, x) + dist(x, x_{\epsilon_1})$ and so

$$dist(p, x_{\epsilon_1}) = dist(p, x) + dist(x, x_{\epsilon_1})$$

Now examining the implications of this last equality we see that the concatenation of $\gamma|_{[0,t_{\sup}]}$ with γ_1 forms a curve from p to x_{ϵ_1} of length $dist(p, x_{\epsilon_1})$ which must therefore be a minimizing curve. By exercise 20.22 below, this potentially broken geodesic must in fact be unbroken and so must actually be the geodesic $\gamma|_{[0,t_{\sup}+\epsilon_1]}$ and so $t_{\sup} + \epsilon_1 \in T$ which contradicts the definition of $t_{\sup}$. This contradiction forces us to conclude that $t_{\max} = R$ and we are done. ■



Exercise 20.22 *Show that a piecewise smooth curve connecting two points p_0 and p_1 in a Riemannian manifold must be smooth if it is length minimizing. Hint: Suppose the curve has a corner and look in a small normal neighborhood of the corner. Show that the curves can be shortened by rounding off the corner.*

Theorem 20.21 (Hopf-Rinow) *If M, g is a connected Riemannian manifold then the following statements are equivalent:*

(i) The metric space M, dist is complete. That is every Cauchy sequence is convergent.

(ii) There is a point $p \in M$ such that M is geodesically complete at p.

(iii) M is geodesically complete.

(iv) Every closed and bounded subset of M is compact.

Proof. (i)⇔(iv) is the famous Heine-Borel theorem and we shall not reproduce the proof here.

(i)⇒(iii): Let p be arbitrary and let $\gamma_v(t)$ be the geodesic with $\dot{\gamma}_v(0) = v$ and J its maximal domain of definition. We can assume without loss that $\langle v, v \rangle = 1$ so that $L(\gamma_v|_{[t_1,t_2]}) = t_2 - t_1$ for all relevant t_1, t_2. We want to show that there can be no upper bound for the set J. We argue by contradiction: Assume that $t_+ = \sup J$ is finite. Let $\{t_n\} \subset J$ be a Cauchy sequence such that $t_n \to t_+ < \infty$. Since $dist(\gamma_v(t), \gamma_v(s)) \leq |t - s|$ it follows that $\gamma_v(t_n)$ is a Cauchy sequence in M which by assumption must converge. Let $q := \lim_{n\to\infty} \gamma_v(t_n)$ and choose a small ball $B_\epsilon(q)$ which is small enough to be a normal neighborhood. Take t_1 with $0 < t_+ - t_1 < \epsilon/2$ and let γ_1 be the (maximal) geodesic with initial velocity $\dot{\gamma}_v(t_1)$. Then in fact $\gamma_1(t) = \gamma_v(t_1 + t)$ and so γ_1 is defined $t_1 + \epsilon/2 > t_+$ and this is a contradiction.

(iii)⇒(ii) is a trivial implication.

(ii)⇒(i): Suppose M is geodesically complete at p. Now let $\{x_n\}$ be any Cauchy sequence in M. For each x_n there is (by assumption) a minimizing geodesic from p to x_n which we denote by γ_{px_n}. We may assume that each γ_{px_n} is unit speed. It is easy to see that the sequence $\{l_n\}$, where $l_n := L(\gamma_{px_n}) = dist(p, x_n)$, is a Cauchy sequence in $\mathbb{R}$ with some limit, say l. The key fact is that the vectors $\dot{\gamma}_{px_n}$ are all unit vectors in T_pM and so form a sequence in the (compact) unit sphere in T_pM. Replacing$\{\dot{\gamma}_{px_n}\}$ by a subsequence if necessary we have $\dot{\gamma}_{px_n} \to u \in T_pM$ for some unit vector u. Continuous dependence on initial velocities implies that $\{x_n\} = \{\gamma_{px_n}(l_n)\}$ has the limit $\gamma_u(l)$. ■

If M, g is a complete connected Riemannian manifold with sectional curvature $K \leq 0$ then for each point $p \in M$ the geodesics emanating from p have no conjugate points and so $T_{v_p} \exp_p : T_{v_p}T_pM \to M$ is nonsingular for each $v_p \in \mathrm{dom}(\exp_p) \subset T_pM$. This means that $\exp_p$ is a local diffeomorphism. If we give T_pM the metric $\exp_p^*(\mathrm{g})$ then $\exp_p$ is a local isometry. Also, since the rays $t \mapsto tv$ in T_pM map to geodesics we see that M is complete at p and then by the Hopf-Rinow theorem M is complete. It now follows from theorem 20.10 that $\exp_p : T_pM \to M$ is a Riemannian covering. Thus we arrive at the Hadamard theorem

Theorem 20.22 (Hadamard) *If M, g is a complete* simply connected *Riemannian manifold with sectional curvature $K \leq 0$ then $\exp_p : T_pM \to M$ is a diffeomorphism and each two points of M can be connected by unique geodesic segment.*

Definition 20.40 *If M, g is a Riemannian manifold then the diameter of M is defined to be*

$$\mathrm{diam}(M) := \sup\{\mathrm{dist}(p, q) : p, q \in M\}$$

The injectivity radius at $p \in M$, denoted $inj(p)$, is the supremum over all $\epsilon > 0$ such that $\exp_p : \widetilde{B}(0_p, \epsilon) \to B(p, \epsilon)$ is a diffeomorphism. The injectivity radius of M is $\mathrm{inj}(M) := \inf_{p\in M}\{\mathrm{inj}(p)\}$.

The Hadamard theorem above has as a hypothesis that the sectional curvature is nonpositive. A bound on the sectional curvature is stronger that a bound on Ricci curvature since the latter is a sort of average sectional curvature. In the sequel, statement like $Ric \geq C$ should be interpreted to mean $Ric(v, v) \geq C\langle v, v\rangle$ for all $v \in TM$.

Lemma 20.11 *Let M, g be an n-dimensional Riemannian manifold and Let $\gamma : [0, L] \to M$ be a unit speed geodesic. Suppose that $Ric \geq (n-1)\,\kappa > 0$ for some constant $\kappa > 0$ (at least along γ). If the length L of γ is greater than or equal to $\pi/\sqrt{\kappa}$ then there is a point conjugate to $\gamma(0)$ along γ.*

Proof. Suppose $b = \pi/\sqrt{\kappa}$ for $0 < b \leq L$. Letting $\varepsilon = \langle\dot{\gamma}, \dot{\gamma}\rangle$, if we can show that $\varepsilon I_{\perp}$ is not positive definite then 19.6 implies the result. To show that $I_{\perp}$ is not positive definite we find an appropriate vector field $V \neq 0$ along γ such that $I(V, V) \leq 0$. Choose orthonormal fields $E_2, ..., E_n$ so that $\dot{\gamma}, E_2, ..., E_n$ is an orthonormal frame along γ. Now for a function $f : [0, \pi/\sqrt{\kappa}] \to \mathbb{R}$ which vanishes we form the fields fE_i using 20.10 we have

$$I(fE_j, fE_j) = \int_0^{\pi/\sqrt{\kappa}} \left\{ f'(s)^2 + f(s)^2\langle R_{E_j,\dot{\gamma}}(E_j(s)), \dot{\gamma}(s)\rangle \right\} ds$$

and then

$$\begin{aligned} \sum_{j=2}^{n} I(fE_j, fE_j) &= \int_0^{\pi/\sqrt{\kappa}} \left\{ (n-1)f'^2 - f^2 Ric(\dot{\gamma}, \dot{\gamma}) \right\} ds \\ &\leq (n-1)\int_0^{\pi/\sqrt{\kappa}} \left(f'^2 - \kappa f^2 \right) ds \end{aligned}$$

Letting $f(s) = \sin(\sqrt{\kappa} s)$ we get

$$\sum_{j=2}^{n} I(fE_j, fE_j) \leq (n-1)\int_0^{\pi/\sqrt{\kappa}} \kappa \left(\cos^2(\sqrt{\kappa} s) - \sin(\sqrt{\kappa} s) \right) ds = 0$$

and so $I(fE_j, fE_j) \leq 0$ for some j. ■

The next theorem also assumes only a bound on the Ricci curvature and is one of the most celebrated theorems of Riemannian geometry.

Theorem 20.23 (Myers) *Let M, g be a complete Riemannian manifold of dimension n. If $Ric \geq (n-1)\,\kappa > 0$ then*

(i) $\operatorname{diam}(M) \leq \pi/\sqrt{\kappa}$, *$M$ is compact and*

(ii) $\pi_1(M)$ is finite.

Proof. Since M complete there is always a shortest geodesic γ_{pq} between any two given points p and q. We can assume that γ_{pq} is parameterized by arc length:

$$\gamma_{pq} : [0, dist(p, q)] \to M$$

It follows that $\gamma_{pq}|_{[0,a]}$ is arc length minimizing for all $a \in [0, dist(p,q)]$. From 20.11we see that the only possible conjugate to p along γ_{pq} is q. From preceding lemma we see that $\pi/\sqrt{\kappa} > 0$ is impossible.

Since the point p and q were arbitrary we must have $\mathrm{diam}(M) \leq \pi/\sqrt{\kappa}$. It follows from the Hopf-Rinow theorem that M is compact.

For (ii) we consider the simply connected covering $\wp : \widetilde{M} \to M$ (which is a local isometry). Since $\widetilde{M}$ is also complete, and has the same Ricci curvature bound as M so $\widetilde{M}$ is also compact. It follows easily that $\wp^{-1}(p)$ is finite for any $p \in M$ from which (ii) follows. ∎

20.12 Cut Locus

Related to the notion of conjugate point is the notion of a cut point. For a point $p \in M$ and a geodesic γ emanating from $p = \gamma(0)$, a **cut point of** p along γ is the first point $q = \gamma(t')$ along γ such that for any point $r = \gamma(t'')$ beyond p, (i.e. $t'' > t'$) there is a geodesic shorter that $\gamma|_{[0,t']}$ which connects p with r. To see the difference between this notion and that of a point conjugate to p it suffices to consider the example of a cylinder $S^1 \times \mathbb{R}$ with the obvious flat metric. If $p = (e^{i\theta}, 0) \in S^1 \times \mathbb{R}$ then for any$x \in \mathbb{R}$, the point $(e^{i(\theta+\pi)}, x)$ is a cut point of p along the geodesic $\gamma(t) := ((e^{i(\theta+t\pi)}, x))$. We know that beyond a conjugate point, a geodesic is not (locally) minimizing but the last example shows that a cut point need not be a conjugate point. In fact, $S^1 \times \mathbb{R}$ has no conjugate points along any geodesic. Let us agree that all geodesics referred to in this section are parameterized by arc length unless otherwise indicated.

Definition 20.41 *Let M, g be a complete Riemannian manifold and let $p \in M$. The set $C(p)$ of all cut points to p along geodesics emanating from p is called the **cut locus** of p.*

For a point $p \in M$, the situations is summarized by the fact that if $q = \gamma(t')$ is a cut point of p along a geodesic γ then for any $t'' > t'$ there is a geodesic connecting p with q which is shorter than $\gamma|_{[0,t']}$ while if $t'' < t'$ then not only is there geodesic connecting p and $\gamma(t'')$ with shorter length but there is no geodesic connecting p and $\gamma(t'')$ whose length is even equal to that of $\gamma|_{[0,t'']}$.

Consider the following two conditions:

a $\gamma(t_0)$ is the first conjugate point of $p = \gamma(0)$ along γ.

b There is a geodesic α different from $\gamma|_{[0,t_0]}$ such that $L(\alpha) = L(\gamma|_{[0,t_0]})$.

Proposition 20.14 *Let M be a complete Riemannian manifold.*
(i) If for a given unit speed geodesic γ, either condition (a) or (b) holds, then there is a $t_1 \in (0, t_0]$ such that $\gamma(t_1)$ is the cut point of p along γ.

(ii) If $\gamma(t_0)$ is the cut point of $p = \gamma(0)$ along the unit speed geodesic ray γ. Then either condition (a) or (b) holds.

Proof. (i) This is already clear from our discussion: for suppose (a) holds, then $\gamma|_{[0,t']}$ cannot minimize for $t' > t_0$ and so the cut point must be $\gamma(t_1)$ for some $t_1 \in (0, t_0]$. Now if (b) hold then choose $\epsilon > 0$ small enough that $\alpha(t_0 - \epsilon)$ and $\gamma(t_0 + \epsilon)$ are both contained in a convex neighborhood of $\gamma(t_0)$. The concatenation of $\alpha|_{[0,t_0]}$ and $\gamma|_{[t_0,t_0+\epsilon]}$ is a curve c that has a kink at $\gamma(t_0)$. But there is a unique minimizing geodesic τ joining $\alpha(t_0 - \epsilon)$ to $\gamma(t_0 + \epsilon)$ and we can concatenate the geodesic $\alpha|_{[0,t_0-\epsilon]}$ with τ to get a curve with arc length strictly less than $L(c) = t_0 + \epsilon$. It follows that the cut point to p along γ must occur at $\gamma(t')$ for some $t' \leq t_0 + \epsilon$. But ϵ can be taken arbitrarily small and so the result (i) follows.

Now suppose that $\gamma(t_0)$ is the cut point of $p = \gamma(0)$ along a unit speed geodesic ray γ. We let $\epsilon_i \to 0$ and consider a sequence $\{\alpha_i\}$ of minimizing geodesics with α_i connecting p to $\gamma(t_0 + \epsilon_i)$. We have a corresponding sequence of initial velocities $u_i := \dot{\alpha}_i(0) \in S^1 \subset T_pM$. The unit sphere in T_pM is compact so replacing u_i by a subsequence we may assume that $u_i \to u \in S^1 \subset T_pM$. Let α be the unit speed segment joining p to $\gamma(t_0 + \epsilon_i)$ with initial velocity u. Arguing from continuity, we see that α is also a minimizing and $L(\alpha) = L(\gamma|_{[0,t_0+\epsilon]})$. If $\alpha \neq \gamma|_{[0,t_0+\epsilon]}$ then we are done. If $\alpha = \gamma|_{[0,t_0]}$ then since $\gamma|_{[0,t_0+\epsilon]}$ is minimizing it will suffice to show that $T_{t_0\dot{\gamma}(0)} \exp_p$ is singular since that would imply that condition (a) holds. The proof of this last statement is by contradiction: Suppose that $\alpha = \gamma|_{[0,t_0]}$ (so that $\dot{\gamma}(0) = u$) and that $T_{t_0\dot{\gamma}(0)} \exp_p$ is not singular. Take U to be an open neighborhood of $t_0\dot{\gamma}(0)$ in T_pM such that $\exp_p\big|_U$ is a diffeomorphism. Now $\alpha_i(t_0+\epsilon_i') = \gamma(t_0+\epsilon_i)$ for $0 < \epsilon_i' \leq \epsilon_i$ since the α_i are minimizing. We now restrict attention to i such that ϵ_i is small enough that $(t_0 + \epsilon_i')u_i$ and $(t_0 + \epsilon_i)u$ are in U. Then we have

$$\begin{aligned}\exp_p(t_0 + \epsilon_i)u = \gamma(t_0 + \epsilon_i) = \\ \alpha_i(t_0 + \epsilon_i') = \exp_p(t_0 + \epsilon_i')u_i\end{aligned}$$

and so $(t_0 + \epsilon_i)u = (t_0 + \epsilon_i')u_i$ and then since $\epsilon_i \to 0$ we have $\dot{\gamma}(0) = u = u_i$ for sufficiently large i. But then $\alpha_i = \gamma$ on $[0, t_0]$ which contradicts the fact that $\gamma|_{[0,t_0]}$ is not minimizing. ∎

Exercise 20.23 *Show that if q is the cut point of p along γ then p is the cut point of q along $\gamma^{\leftarrow}$ (where $\gamma^{\leftarrow}(t) := \gamma(L - t)$ and $L = L(\gamma)$).*

It follows from the development so far that if $q \in M \backslash C(p)$ then there is a unique minimizing geodesic joining p to q and that if $B(p, R)$ is the ball of radius R centered at p then $\exp_p$ is a diffeomorphism on $B(p, R)$ if $R \leq d(p, C(p))$. In fact, an alternative definition of the injectivity radius at p is $d(p, C(p))$ and the injectivity radius of M is

$$inj(M) = \inf_{p \in M} \{d(p, C(p))\}$$

Intuitively, the complexities of the topology of M begin at the cut locus of a given point.

Let T_1M denote the unit tangent bundle of the Riemannian manifold:

$$T_1M = \{u \in TM : \|u\| = 1\}.$$

Define a function $c_M : T_1M \to (0, \infty]$ by

$$c_M(u) := \begin{cases} t_0 & \text{if } \gamma_u(t_0) \text{ is the cut point of } \pi_{TM}(u) \text{ along } \gamma_u \\ \infty & \text{if there is no cut point in the direction } u \end{cases}$$

Recall that the topology on $(0, \infty]$ is such that a sequence t_k converges to the point ∞ if $\lim_{k\to\infty} t_k = \infty$ in the usual sense. We now have

Theorem 20.24 *If M, g is complete Riemannian manifold then the function $c_M : T_1M \to (0, \infty]$ is continuous.*

Proof. If u_i is a sequence in T_1M converging to $u \in T_1M$ then $\pi_{TM}(u_i) = p_i$ converges to $p = \pi_{TM}(u)$. Let u_i be such a sequence and let γ_i be the unit speed geodesic connecting p_i to the corresponding cut point $\gamma_i(t_{0i})$ in the direction u_i and where $t_{0i} = \infty$ if there is no such cut point (in this case just let $\gamma_i(t_{0i})$ be arbitrary). Note that $u_i = \dot{\gamma}_i(0)$. Also let γ be the geodesic with initial velocity u. Let $t_0 \in (0, \infty]$ be the distance to the cut point in the direction u. Our task is to show that t_{0i} converges to t_0.

Claim 1: $\limsup t_{0i} \leq t_0$. If $t_0 = \infty$ then this claim is trivially true so assume that $t_0 < \infty$. Given any $\epsilon > 0$ there is only a finite number of i such that $t_0 + \epsilon < t_{0i}$ for otherwise we would have a sequence i_k such that $d(p_{i_k}, \gamma_{i_k}(t_0 + \epsilon)) = t_0 + \epsilon$ which would give $d(p, \gamma(t_0 + \epsilon)) = t_0 + \epsilon$. But this last equality contradicts the fact that $\gamma(t_0)$ is the cut point of p along γ. Thus we must have $\limsup t_{0i} \leq t_0 + \epsilon$ and since ϵ was arbitrarily small we deduce the truth of the claim.

Claim 2: $\liminf t_{0i} \geq t$. The theorem is proved once we prove this claim. For the proof of this claim we suppose that $\liminf t_{0i} < \infty$ since otherwise there is nothing to prove. Once again let t_{0i} be a sequence which (after a reduction to a subsequence we may assume) converges to $t_{\inf} := \liminf t_{0i}$. The reader may easily prove that if (after another reduction to a subsequence) the points $\gamma_i(t_{0i})$ are conjugate to p_i along γ_i then $\gamma(t_{\inf})$ is conjugate to p along γ. If this is the case then $t_{\inf} := \liminf t_{0i} \geq t_0$ and the claim is true. Suppose therefore that there is a sequence of indices i_k so that $t_{0i_k} \to t_{\inf}$ and such that $\gamma_{i_k}(t_{0i_k})$ is *not* conjugate to p_i along γ_{i_k}. In this case there must be a (sub)sequence of geodesics α_i different from γ_i such that $\alpha_i(0) = \gamma_i(0) = p_i$, $\alpha_{i_k}(t_{0i_k}) = \gamma_{i_k}(t_{0i_k})$ and $L(\alpha_{i_k}) = L(\gamma_{i_k})$. After another reduction to a subsequence we may assume that $\dot{\alpha}_i(0) \to v \in T_1M$ and that a geodesic α in with $\dot{\alpha}(0) = v$ connects p to $\gamma(t_{\inf})$. If α is different than γ then by 20.14 $t_0 \leq t_{\inf}$ and we are done. If, on the other hand, $\alpha = \gamma$ then an argument along the lines of proposition 20.14(see exercise 20.24 below) $\gamma(t_{\inf})$ is conjugate to p along γ which again implies $t_0 \leq t_{\inf}$. ∎

Exercise 20.24 *Fill in the final details of the proof of theorem 20.24.*

Chapter 21

Geometry of Submanifolds

21.1 Definitions

Let M be a d dimensional submanifold of a semi-Riemannian manifold $\overline{M}$ of dimension n where $d < n$. The metric $\mathrm{g}(.,.) = \langle .,. \rangle$ on $\overline{M}$ restricts to tensor on M which we denote by h. Since h is a restriction of g we shall also use the notation $\langle .,. \rangle$ for h. If the restriction h of is nondegenerate on each space T_pM then h is a metric tensor on M and we say that M is a semi-Riemannian submanifold of $\bar{M}$. If $\bar{M}$ is Riemannian then this nondegeneracy condition is automatic and the metric h is automatically Riemannian. More generally, if $\phi : M \to \bar{M}, \mathrm{g}$ is an immersion we can consider the pull-back tensor $\phi^*\mathrm{g}$ defined by

$$\phi^*\mathrm{g}(X, Y) = \mathrm{g}(T\phi \cdot X, T\phi \cdot Y).$$

If $\phi^*\mathrm{g}$ is nondegenerate on each tangent space then it is a metric on M called the pull-back metric and we call ϕ a semi-Riemannian immersion. If M is already endowed with a metric g_M then if $\phi^*\mathrm{g} = \mathrm{g}_M$ then we say that $\phi : M, \mathrm{g}_M \to \bar{M}, g$ is an isometric immersion. Of course, if $\phi^*\mathrm{g}$ is a metric at all, as it always is if $\bar{M}, g$ is Riemannian, then the map $\phi : M, \phi^*\mathrm{g} \to \bar{M}, \mathrm{g}$ is an isometric immersion. Since every immersion restricts locally to an embedding we may, for many purposes, assume that M is a submanifold and that ϕ is just the inclusion map.

Definition 21.1 *Let $\overline{M}, \mathrm{g}$ be a Lorentz manifold. A submanifold M is said to be spacelike (resp. timelike, lightlike) if $T_pM \subset T_p\overline{M}$ is spacelike (resp. timelike, lightlike).*

There is an obvious bundle on M which is the restriction of $T\overline{M}$ to M. This is the bundle $T\overline{M}\big|_M = \bigsqcup_{p \in M} T_p\overline{M}$. Each tangent space $T_p\overline{M}$ decomposes as

$$T_p\overline{M} = T_pM \oplus (T_pM)^{\perp}$$

where $(T_pM)^{\perp} = \{v \in T_p\overline{M} : \langle v, w \rangle = 0 \text{ for all } w \in T_pM\}$. Then $TM^{\perp} = \bigsqcup_p (T_pM)^{\perp}$ its natural structure as a smooth vector bundle called the **normal**

bundle to M in $\overline{M}$. The smooth sections of the normal bundle will be denoted by $\Gamma\left(TM^{\perp}\right)$ or $\mathfrak{X}(M)^{\perp}$. Now the orthogonal decomposition above is globalizes as

$$T\overline{M}\big|_M = TM \oplus TM^{\perp}$$

A vector field on M is always the restriction of some (not unique) vector field on a neighborhood of $\overline{M}$. The same is true of any not necessarily tangent vector field along M. The set of all vector fields along M will be denoted by $\mathfrak{X}(\overline{M})\big|_M$. Since any function on M is also the restriction of some function on $\overline{M}$ we may consider $\mathfrak{X}(M)$ as a submodule of $\mathfrak{X}(\overline{M})\big|_M$. If $\overline{X} \in \mathfrak{X}(\overline{M})$ then we denote its restriction to M by $\overline{X}\big|_M$ or sometimes just X. Notice that $\mathfrak{X}(M)^{\perp}$ is a submodule of $\mathfrak{X}(\overline{M})\big|_M$. We have two projection maps : $T_p\overline{M} \to N_pM$ and tan : $T_p\overline{M} \to T_pM$ which in turn give module projections nor : $\mathfrak{X}(\overline{M})\big|_M \to \mathfrak{X}(M)^{\perp}$ and : $\mathfrak{X}(\overline{M})\big|_M \to \mathfrak{X}(M)$. The reader should contemplate the following diagrams:

$$\begin{array}{ccccc} C^{\infty}(\overline{M}) & \overset{restr}{\to} & C^{\infty}(M) & = & C^{\infty}(M) \\ \times\downarrow & & \times\downarrow & & \times\downarrow \\ \mathfrak{X}(\overline{M}) & \overset{restr}{\to} & \mathfrak{X}(\overline{M})\big|_M & \overset{\tan}{\to} & \mathfrak{X}(M) \end{array}$$

$$\text{and} \quad \begin{array}{ccccc} C^{\infty}(\overline{M}) & \overset{restr}{\to} & C^{\infty}(M) & = & C^{\infty}(M) \\ \times\downarrow & & \times\downarrow & & \times\downarrow \\ \mathfrak{X}(\overline{M}) & \overset{restr}{\to} & \mathfrak{X}(\overline{M})\big|_M & \overset{\text{nor}}{\to} & \mathfrak{X}(M)^{\perp} \end{array}.$$

Now we also have an exact sequence of modules

$$0 \to \mathfrak{X}(M)^{\perp} \to \mathfrak{X}(\overline{M})\big|_M \overset{\tan}{\to} \mathfrak{X}(M) \to 0$$

which is in fact a split exact sequence since we also have

$$0 \longleftarrow \mathfrak{X}(M)^{\perp} \overset{\text{nor}}{\longleftarrow} \mathfrak{X}(\overline{M})\big|_M \longleftarrow \mathfrak{X}(M) \longleftarrow 0$$

The extension map $\mathfrak{X}(\overline{M})\big|_M \longleftarrow \mathfrak{X}(M)$ is not canonical but in the presence of a connection it is almost so: If $U_\epsilon(M)$ is the open tubular neighborhood of M given, for sufficiently small ϵ by

$$U_\epsilon(M) = \{p \in \overline{M} : \text{dist}(p, M) < \epsilon\}$$

then we can use the following trick to extend any $X \in \mathfrak{X}(M)$ to $\mathfrak{X}(U_\epsilon(M))|_M$. First choose a smooth frame field $E_1, ..., E_n$ defined along M so that $E_i \in \mathfrak{X}(\overline{M})\big|_M$. We may arrange if need to have the first d of these tangent to M. Now parallel translate each frame radially outward a distance ϵ to obtain a smooth frame field $\overline{E}_1, ..., \overline{E}_n$ on $U_\epsilon(M)$.

Now we shall obtain a sort of splitting of the Levi-Civita connection of $\overline{M}$ along the submanifold M. The reader should recognize a familiar theme here especially if elementary surface theory is fresh in his or her mind. First we notice that the Levi-Civita connection $\bar{\nabla}$ on $\overline{M}$ restrict nicely to a connection on the bundle $T\overline{M}_M \to M$. The reader should be sure to realize that the space of sections of this bundle is exactly $\mathfrak{X}(\overline{M})\big|_M$ and so the restricted connection is a map $\bar{\nabla}\big|_M : \mathfrak{X}(M) \times \mathfrak{X}(\overline{M})\big|_M \to \mathfrak{X}(\overline{M})\big|_M$. The point is that if $X \in \mathfrak{X}(M)$ and $W \in \mathfrak{X}(\overline{M})\big|_M$ then $\bar{\nabla}_X W$ doesn't seem to be defined since X and W are not elements of $\mathfrak{X}(\overline{M})$. But we may extend X and W to elements of $\mathfrak{X}(\overline{M})$ and then restrict again to get an element of $\mathfrak{X}(\overline{M})\big|_M$.Then recalling the local properties of a connection we see that the result does not depend on the extension.

Exercise 21.1 *Use local coordinates to prove the claimed independence on the extension.*

We shall write simply $\bar{\nabla}$ in place of $\bar{\nabla}\big|_M$ since the context make it clear when the later is meant. Thus $\bar{\nabla}_X W := \bar{\nabla}_{\bar{X}} \bar{W}$ where $\bar{X}$ and $\bar{W}$ are any extensions of X and W respectively.

Clearly we have $\bar{\nabla}_X \langle Y_1, Y_2\rangle = \langle \bar{\nabla}_X Y_1, Y_2\rangle + \langle Y_1, \bar{\nabla}_X Y_2\rangle$ and so $\bar{\nabla}$ is a metric connection on $T\overline{M}\big|_M$. For a fixed $X, Y \in \mathfrak{X}(M)$ we have the decomposition of $\bar{\nabla}_X Y$ into tangent and normal parts. Similarly, for $V \in \mathfrak{X}(M)^\perp$ we can consider the decomposition of $\bar{\nabla}_X V$ into tangent and normal parts. Thus we have

$$\begin{aligned}
\bar{\nabla}_X Y &= (\bar{\nabla}_X Y)^{\tan} + (\bar{\nabla}_X Y)^\perp \\
\bar{\nabla}_X V &= (\bar{\nabla}_X V)^{\tan} + (\bar{\nabla}_X V)^\perp
\end{aligned}$$

We make the following definitions:

$$\begin{aligned}
\nabla_X Y &:= (\bar{\nabla}_X Y)^{\tan} \text{ for all } X, Y \in \mathfrak{X}(M) \\
b_{12}(X, Y) &:= (\bar{\nabla}_X Y)^\perp \text{ for all } X, Y \in \mathfrak{X}(M) \\
b_{21}(X, V) &:= (\bar{\nabla}_X V)^{\tan} \text{ for all } X \in \mathfrak{X}(M),\ V \in \mathfrak{X}(M)^\perp \\
\nabla^\perp_X V &:= (\bar{\nabla}_X V)^\perp \text{ for all } X \in \mathfrak{X}(M),\ V \in \mathfrak{X}(M)^\perp
\end{aligned}$$

Now if $X, Y \in \mathfrak{X}(M)$, $V \in \mathfrak{X}(M)^\perp$ then $0 = \langle Y, V\rangle$ and so

$$\begin{aligned}
0 &= \bar{\nabla}_X \langle Y, V\rangle \\
&= \langle \bar{\nabla}_X Y, V\rangle + \langle Y, \bar{\nabla}_X V\rangle \\
&= \langle \left(\bar{\nabla}_X Y\right)^\perp, V\rangle + \langle Y, \left(\bar{\nabla}_X V\right)^{\tan}\rangle \\
&= \langle b_{12}(X, Y), V\rangle + \langle Y, b_{21}(X, V)\rangle.
\end{aligned}$$

It follows that $\langle b_{12}(X, Y), V\rangle = -\langle Y, b_{21}(X, V)\rangle$. Now we from this that $b_{12}(X, Y)$ is not only $C^\infty(M)$ linear in X but also in Y. Thus b_{12} is tensorial and so for each fixed $p \in M$, $b_{12}(X_p, Y_p)$ is a well defined element of $T_pM^\perp$ for each fixed $X_p, Y_p \in T_pM$. Also, for any $X_1, X_2 \in \mathfrak{X}(M)$ we have

$$\begin{aligned}
&b_{12}(X_1, X_2) - b_{12}(X_2, X_1) \\
&= \left(\bar{\nabla}_{X_1} X_2 - \bar{\nabla}_{X_2} X_1\right)^\perp \\
&= ([X_1, X_2])^\perp = 0.
\end{aligned}$$

So b_{12} is symmetric. The classical notation for b_{12} is II and the form is called the **second fundamental tensor** or the **second fundamental form**. For $\xi \in T_pM$ we define the linear map $B_\xi(\cdot) := b_{12}(\xi, \cdot)$. With this in mind we can easily deduce the following facts which we list as a theorem:

1. $\nabla_X Y := (\bar{\nabla}_X Y)^{\tan}$ defines a connection on M which is identical with the Levi-Civita connection for the induced metric on M.
2. $(\bar{\nabla}_X V)^{\perp} := \nabla_X^{\perp} V$ defines a metric connection on the vector bundle $TM^{\perp}$.
3. $(\bar{\nabla}_X Y)^{\perp} := b_{12}(X, Y)$ defines a symmetric $C^\infty(M)$-bilinear form with values in $\mathfrak{X}(M)^{\perp}$.
4. $(\bar{\nabla}_X V)^{\tan} := b_{21}(X, V)$ defines a symmetric $C^\infty(M)$-bilinear form with values in $\mathfrak{X}(M)$.

Corollary 21.1 *b_{21} is tensorial and so for each fixed $p \in M$, $b_{21}(X_p, Y_p)$ is a well defined element of $T_pM^{\perp}$ for each fixed $X_p, Y_p \in T_pM$ and we have a bilinear form $b_{21} : T_pM \times T_pM^{\perp} \to T_pM$.*

Corollary 21.2 *The map $b_{21}(\xi, \cdot) : T_pM^{\perp} \to T_pM$ is equal to $-B_\xi^t : T_pM \to T_pM^{\perp}$.*

Writing any $Y \in \mathfrak{X}(\overline{M})\big|_M$ as $Y = (Y^{\tan}, Y^{\perp})$ we can write the map $\bar{\nabla}_X : \mathfrak{X}(\overline{M})\big|_M \to \mathfrak{X}(\overline{M})\big|_M$ as a matrix of operators:

$$\begin{bmatrix} \nabla_X & B_X \\ -B_X^t & \nabla_X^{\perp} \end{bmatrix}$$

Next we define the shape operator which is also called the Weingarten map. There is not a perfect agreement on sign conventions; some author's shape operator is the negative of the shape operator as defined by other others. We shall define S^+ and S^- which only differ in sign. This way we can handle both conventions simultaneously and the reader can see where the sign convention does or does not make a difference in any given formula.

Definition 21.2 *Let $p \in M$. For each unit vector u normal to M at p we have a map called the ($\pm$) **shape operator** $S_u^{\pm}$ associated to u. defined by $S_u^{\pm}(v) := \left(\pm\bar{\nabla}_v U\right)^{\tan}$ where U is any unit normal field defined near p such that $U(p) = u$.*

The shape operators $\{S_u^{\pm}\}_{u \text{ a unit normal}}$ contain essentially the same information as the second fundamental tensor $II = b_{12}$. This is because for any $X, Y \in \mathfrak{X}(M)$ and $U \in \mathfrak{X}(M)^{\perp}$ we have

$$\begin{aligned} \langle S_U^{\pm} X, Y \rangle &= \langle \left(\pm\bar{\nabla}_X U\right)^{\tan}, Y \rangle = \langle U, \pm\bar{\nabla}_X Y \rangle \\ &= \langle U, \left(\pm\bar{\nabla}_X Y\right)^{\perp} \rangle = \langle U, \pm b_{12}(X, Y) \rangle \\ &= \langle U, \pm II(X, Y) \rangle. \end{aligned}$$

Note: $\langle U, b_{12}(X, Y) \rangle$ is tensorial in U, X and Y. Of course , $S_{-U}^{\pm} X = -S_U^{\pm} X$.

Theorem 21.1 *Let M be a semi-Riemannian submanifold of $\bar{M}$. We have the* ***Gauss equation***

$$\begin{aligned}\langle R_{VW}X,Y\rangle &= \langle \bar{R}_{VW}X,Y\rangle \\ &\quad - \langle II(V,X), II(W,Y)\rangle + \langle II(V,Y), II(W,X)\rangle\end{aligned}$$

Proof. Since this is clearly a tensor equation we may assume that $[V,W]=0$. With this assumption we have we have $\langle \bar{R}_{VW}X,Y\rangle = (VW)-(WV)$ where $(VW) = \langle \bar{\nabla}_V\bar{\nabla}_W X,Y\rangle$

$$\begin{aligned}\langle \bar{\nabla}_V\bar{\nabla}_W X,Y\rangle &= \langle \bar{\nabla}_V\nabla_W X,Y\rangle + \langle \bar{\nabla}_V(II(W,X)),Y\rangle \\ &= \langle \nabla_V\nabla_W X,Y\rangle + \langle \bar{\nabla}_V(II(W,X)),Y\rangle \\ &= \langle \nabla_V\nabla_W X,Y\rangle + V\langle II(W,X),Y\rangle - \langle II(W,X),\bar{\nabla}_V Y\rangle \\ &= \langle \nabla_V\nabla_W X,Y\rangle + V\langle II(W,X),Y\rangle - \langle II(W,X),\bar{\nabla}_V Y\rangle\end{aligned}$$

Since

$$\begin{aligned}&\langle II(W,X),\bar{\nabla}_V Y\rangle \\ &= \langle II(W,X),\left(\bar{\nabla}_V Y\right)^{\perp}\rangle = \langle II(W,X), II(V,Y)\rangle\end{aligned}$$

we have $(VW) = \langle \nabla_V\nabla_W X,Y\rangle - \langle II(W,X), II(V,Y)\rangle$. Interchanging the roles of V and W and subtracting we get the desired conclusion. ∎

Another formula that follows easily from the Gauss equation is the following formula (also called the Gauss formula):

$$K(v\wedge w) = \bar{K}(v\wedge w) + \frac{\langle II(v,v), II(w,w)\rangle - \langle II(v,w), II(v,w)\rangle}{\langle v,v\rangle\langle w,w\rangle - \langle v,w\rangle^2}$$

Exercise 21.2 *Prove the last formula (second version of the Gauss equation).*

From this last version of the Gauss equation we can show that a sphere $S^n(r)$ of radius r in $\mathbb{R}^{n+1}$ has constant sectional curvature $1/r^2$ for $n>1$. If (u_i) is the standard coordinates on $\mathbb{R}^{n+1}$ then the position vector field in $\mathbb{R}^{n+1}$ is $\mathbf{r} = \sum_{i=1}^{n+1} u_i\partial_i$. Recalling that the standard (flat connection) D on $\mathbb{R}^{n+1}$ is just the Lie derivative we see that $D_X\mathbf{r} = \sum_{i=1}^{n+1} Xu_i\partial_i = X$. Now using the usual identifications, the unit vector field $U = \mathbf{r}/r$ is the outward unit vector field on $S^n(r)$. We have

$$\begin{aligned}\langle II(X,Y),U\rangle &= \langle D_X Y,U\rangle \\ &= \frac{1}{r}\langle D_X Y,\mathbf{r}\rangle = -\frac{1}{r}\langle Y, D_X\mathbf{r}\rangle \\ &= -\frac{1}{r}\langle Y,X\rangle = -\frac{1}{r}\langle X,Y\rangle.\end{aligned}$$

Now letting $\bar{M}$ be $\mathbb{R}^{n+1}$ and M be $S^n(r)$ and using the fact that the Gauss curvature of $\mathbb{R}^{n+1}$ is identically zero, the Gauss equation gives $K = 1/r$.

The second fundamental form contains information about how the semi-Riemannian submanifold M bends about in $\bar{M}$. First we need a definition:

Definition 21.3 *Let M be semi-Riemannian submanifold of $\bar{M}$ and N a semi-Riemannian submanifold of $\bar{N}$. A pair isometry $\Phi : (\bar{M}, M) \to (\bar{N}, N)$ consists of an isometry $\Phi : \bar{M} \to \bar{N}$ such that $\Phi(M) = N$ and such that $\Phi|_M : M \to N$ is an isometry.*

Proposition 21.1 *A pair isometry $\Phi : (\bar{M}, M) \to (\bar{N}, N)$ preserves the second fundamental tensor:*

$$T_p\Phi \cdot II(v, w) = II(T_p\Phi \cdot v, T_p\Phi \cdot w)$$

for all $v, w \in T_pM$ and all $p \in M$.

Proof. Let $p \in M$ and extend $v, w \in T_pM$ to smooth vector fields V and W. Since an isometry respects the Levi-Civita connections we have $\Phi_*\bar{\nabla}_V W = \bar{\nabla}_{\Phi_* V}\Phi_* W$. Now since Φ is a pair isometry we have $T_p\Phi(T_pM) \subset T_{\Phi(p)}N$ and $T_p\Phi(T_pM^{\perp}) \subset \left(T_{\Phi(p)}N\right)^{\perp}$. This means that $\Phi_* : \mathfrak{X}(\bar{M})\big|_M \to \mathfrak{X}(\bar{N})\big|_N$ preserves normal and tangential components $\Phi_*(\mathfrak{X}(M)) \subset \mathfrak{X}(N)$ and $\Phi_*(\mathfrak{X}(M)^{\perp}) \subset \mathfrak{X}(N)^{\perp}$. We have

$$\begin{aligned}
T_p\Phi \cdot II(v, w) &= \Phi_* II(V, W)(\Phi(p)) \\
&= \Phi_*\left(\bar{\nabla}_V W\right)^{\perp}(\Phi(p)) \\
&= \left(\Phi_*\bar{\nabla}_V W\right)^{\perp}(\Phi(p)) \\
&= \left(\bar{\nabla}_{\Phi_* V}\Phi_* W\right)^{\perp}(\Phi(p)) \\
&= II(\Phi_* V, \Phi_* W)(\Phi(p)) \\
&= II(\Phi_* V, \Phi_* W)(\Phi(p)) \\
&= II(T_p\Phi \cdot v, T_p\Phi \cdot w)
\end{aligned}$$

■

The following example is simple but conceptually very important.

Example 21.1 *Let M be the strip 2 dimensional strip $\{(x, y, 0) : -\pi < x < \pi\}$ considered as submanifold of $\mathbb{R}^3$ (with the canonical Riemannian metric). Let N be the subset of $\mathbb{R}^3$ given by $\{(x, y, \sqrt{1 - x^2}) : -1 < x < 1\}$.* ***Exercise:*** *Show that M is isometric to M. Show that there is no pair isometry $(\mathbb{R}^3, M) \to (\mathbb{R}^3, N)$.*

21.2 Curves in Submanifolds

If $\gamma : I \to M$ is a curve in M and M is a semi-Riemannian submanifold of $\bar{M}$ then we have $\bar{\nabla}_{\partial_t} Y = \nabla_{\partial_t} Y + II(\dot{\gamma}, Y)$ for any vector field Y **along** γ. If Y is a vector field in $\mathfrak{X}(\bar{M})\big|_M$ or in $\mathfrak{X}(\bar{M})$ then $Y \circ \gamma$ is a vector field along γ. In this case we shall still write $\bar{\nabla}_{\partial_t} Y = \nabla_{\partial_t} Y + II(\dot{\gamma}, Y)$ rather than $\bar{\nabla}_{\partial_t}(Y \circ \gamma) = \nabla_{\partial_t}(Y \circ \gamma) + II(\dot{\gamma}, Y \circ \gamma)$.

Recall that $\dot{\gamma}$ is a vector field along γ. We also have $\ddot{\gamma} := \bar{\nabla}_{\partial_t}\dot{\gamma}$ which in this context will be called the **extrinsic acceleration** (or acceleration in $\bar{M}$. By definition we have $\nabla_{\partial_t} Y = \left(\bar{\nabla}_{\partial_t} Y\right)^{\perp}$. The **intrinsic acceleration** (acceleration in M) is $\nabla_{\partial_t}\dot{\gamma}$. Thus we have

$$\ddot{\gamma} = \nabla_{\partial_t}\dot{\gamma} + II(\dot{\gamma}, \dot{\gamma}).$$

From this definitions we can immediately see the truth of the following

Proposition 21.2 *If $\gamma : I \to M$ is a curve in M and M is a semi-Riemannian submanifold of $\bar{M}$ then γ is a geodesic in M if and only if $\ddot{\gamma}(t)$ is normal to M for every $t \in I$.*

Exercise 21.3 *A constant speed parameterization of a great circle in $S^n(r)$ is a geodesic. Every geodesic in $S^n(r)$ is of this form.*

Definition 21.4 *A semi-Riemannian manifold $M \subset \bar{M}$ is called totally geodesic if every geodesic in M is a geodesic in $\bar{M}$.*

Theorem 21.2 *For a semi-Riemannian manifold $M \subset \bar{M}$ the following conditions are equivalent*

i) M is totally geodesic

ii) $II \equiv 0$

iii) For all $v \in TM$ the $\bar{M}$ geodesic γ_v with initial velocity v is such that $\gamma_v[0, \epsilon] \subset M$ for $\epsilon > 0$ sufficiently small.

iv) For any curve $\alpha : I \to M$, parallel translation along α induced by $\bar{\nabla}$ in $\bar{M}$ is equal to parallel translation along α induced by ∇ in M.

Proof. (i)$\Longrightarrow$(iii) follows from the uniqueness of geodesics with a given initial velocity.

iii)$\Longrightarrow$(ii); Let $v \in TM$. Applying 21.2 to γ_v we see that $II(v, v) = 0$. Since v was arbitrary we conclude that $II \equiv 0$.

(ii)$\Longrightarrow$(iv); Suppose $v \in T_pM$. If V is a parallel vector field with respect to ∇ that is defined near p such that $V(p) = v$. Then $\bar{\nabla}_{\partial_t} V = \nabla_{\partial_t} V + II(\dot{\gamma}, V) = 0+0$ for any γ with $\gamma(0) = p$ so that V is a parallel vector field with respect to $\bar{\nabla}$.

(iv)$\Longrightarrow$(i); Assume (iv). If γ is a geodesic in M then γ' is parallel along γ with respect to ∇. Then by assumption γ' is parallel along γ with respect to $\bar{\nabla}$. Thus γ is also a $\bar{M}$ geodesic. ∎

21.3 Hypersurfaces

If the codimension of M in $\bar{M}$ is equal to 1 then we say that M is a **hypersurface**. If M is a semi-Riemannian hypersurface in $\bar{M}$ and $\langle u, u\rangle > 0$ for every $u \in (T_pM)^{\perp}$ we call M a positive hypersurface. If $\langle u, u\rangle < 0$ for every $u \in (T_pM)^{\perp}$ we call M a negative hypersurface. Of course, if $\bar{M}$ is Riemannian then every hypersurface in $\bar{M}$ is positive. The sign of M, in $\bar{M}$ denoted $\operatorname{sgn} M$ is $\operatorname{sgn}\langle u, u\rangle$.

Exercise 21.4 *Suppose that c is a regular value of $f \in C^{\infty}(\bar{M})$ then $M = f^{-1}(c)$ is a semi-Riemannian hypersurface if $\langle df, df\rangle > 0$ on all of M or if $\langle df, df\rangle < 0$. $\operatorname{sgn}\langle df, df\rangle = \operatorname{sgn}\langle u, u\rangle$.*

From the preceding exercise it follows if $M = f^{-1}(c)$ is a semi-Riemannian hypersurface then $U = \nabla f / \|\nabla f\|$ is a unit normal for M and $\langle U, U\rangle = \operatorname{sgn} M$. Notice that this implies that $M = f^{-1}(c)$ is orientable if $\bar{M}$ is orientable. Thus not every semi-Riemannian hypersurface is of the form $f^{-1}(c)$. On the other hand every hypersurface is locally of this form.

In the case of a hypersurface we have (locally) only two choices of unit normal. Once we have chosen a unit normal u the shape operator is denoted simply by S rather than S_u.

We are already familiar with the sphere $S^n(r)$ which is $f^{-1}(r^2)$ where $f(x) = \langle x, x\rangle = \sum_{i=1}^{n} x^i x^i$. A similar example arises when we consider the semi-Euclidean space $\mathbb{R}^{n+1-\nu,\nu}$ where $\nu \neq 0$. In this case, the metric is $\langle x, y\rangle_{\nu} = -\sum_{i=1}^{\nu} x^i y^i + \sum_{i=\nu+1}^{n} x^i y^i$. We let $f(x) := -\sum_{i=1}^{\nu} x^i x^i + \sum_{i=\nu+1}^{n} x^i x^i$ and then for $r > 0$ we have that $f^{-1}(\epsilon r^2)$ is a semi-Riemannian hypersurface in $\mathbb{R}^{n+1-\nu,\nu}$ with sign ε and unit normal $U = \mathbf{r}/r$. We shall divide these hypersurfaces into two classes according to sign.

Definition 21.5 *For $n > 1$ and $0 \leq v \leq n$, we define*

$$S_{\nu}^{n}(r) = \{x \in \mathbb{R}^{n+1-\nu,\nu} : \langle x, x\rangle_{\nu} = r^2.$$

$S_{\nu}^{n}(r)$ is called the pseudo-sphere of index ν.

Definition 21.6 *For $n > 1$ and $0 \leq \nu \leq n$, we define*

$$H_{\nu}^{n}(r) = \{x \in \mathbb{R}^{n+1-(\nu+1),\nu+1} : \langle x, x\rangle_{\nu} = -r^2.$$

$H_{\nu}^{n}(r)$ is called the pseudo-hyperbolic space of radius r and index ν.

Chapter 22

Killing Fields and Symmetric Spaces

Following the excellent treatment given by Lang [L1], we will treat Killing fields in a more general setting than is traditional. We focus attention on pairs (M, ∇) where M is a smooth manifold and ∇ is a (not necessarily metric) torsion free connection. Now let (M, ∇^M) and (N, ∇^N) be two such pairs. For a diffeomorphism $\varphi : M \to N$ the pullback connection $\varphi^*\nabla$ is defined so that

$$(\varphi^*\nabla^N)_{\varphi^*X}\varphi^*Y = \varphi^*(\nabla^M_X Y)$$

for $X, Y \in \mathfrak{X}(N)$ and φ^*X is defined as before by $\varphi^*X = T\varphi^{-1} \circ X \circ \varphi$. An isomorphism of pairs $(M, \nabla^M) \to (N, \nabla^N)$ is a diffeomorphism $\varphi : M \to M$ such that $\varphi^*\nabla = \nabla$. Automorphism of a pair (M, ∇) is defined in the obvious way. Because we will be dealing with the flows of vector fields which are not necessarily complete we will have to deal with maps $\varphi :: M \to M$ with domain a proper open subset $O \subset M$. In this case we say that φ is a local isomorphism if $\varphi^* \nabla|_{\varphi^{-1}(O)} = \nabla|_O$.

Definition 22.1 *A vector field $X \in \mathfrak{X}(M)$ is called a $\nabla-$**Killing field** if Fl^X_t is a local isomorphism of the pair (M, ∇) for all sufficiently small t.*

Definition 22.2 *Let M, g be a semi-Riemannian manifold. A vector field $X \in \mathfrak{X}(M)$ is called a $\mathrm{g}-$**Killing field** if X is a local isometry.*

It is clear that if ∇^{g} is the Levi-Civita connection for M, g then any isometry of M, g is an automorphism of (M, ∇^{g}) and any $\mathrm{g}-$Killing field is a $\nabla^{\mathrm{g}}-$Killing field. The reverse statements are not always true. Letting $Kill_g(M)$ denote the $\mathrm{g}-$Killing fields and $Kill_\nabla(M)$ the $\nabla-$Killing fields we have

$$Kill_{\mathrm{g}}(M) \subset Kill_{\nabla^{\mathrm{g}}}(M)$$

Lemma 22.1 *If (M, ∇) is as above then for any $X, Y, Z \in \mathfrak{X}(M)$ we have*

$$[X, \nabla_Z Y] = \nabla_{Y,Z} - R_{Y,Z}X + \nabla_{[X,Z]}Y + \nabla_Z[X, Y]$$

Proof. The proof is a straightforward calculation left to the reader. ■

Theorem 22.1 *Let M be a smooth manifold and ∇ a torsion free connection. For $X \in \mathfrak{X}(M)$, the following three conditions are equivalent:*

(i) X is a $\nabla-$Killing field

(ii) $[X, \nabla_Z Y] = \nabla_{[X,Z]}Y + \nabla_Z[X, Y]$ for all $Y, Z \in \mathfrak{X}(M)$

(iii) $\nabla_{Y,Z}X = R_{Y,Z}X$ for all $Y, Z \in \mathfrak{X}(M)$.

Proof. The equivalence of (ii) and (iii) follows from the previous lemma. Let $\phi_t := Fl_t^X$. If X is Killing (so (i) is true) then locally we have $\frac{d}{dt}Fl_t^*Y = Fl_t^*\mathcal{L}_X Y = Fl_t^*[X, Y]$ for all $Y \in \mathfrak{X}(M)$. We also have $\phi_t^*X = X$. One calculates that

$$\begin{aligned}\frac{d}{dt}\phi_t^*(\nabla_Z Y) &= \phi_t^*[X, \nabla_Z Y] = [\phi_t^*X, \phi_t^*\nabla_Z Y] \\ &= [X, \phi_t^*\nabla_Z Y]\end{aligned}$$

and on the other hand

$$\frac{d}{dt}\nabla_{\phi_t^*Z}\phi_t^*Y = \nabla_{\phi_t^*[X,Z]}\phi_t^*Y + \nabla_{\phi_t^*Z}(\phi_t^*[X, Y]).$$

Setting $t = 0$ and comparing we get (ii).

Now assume (ii). We would like to show that $\phi_{-t}^*\nabla_{\phi_t^*Z}\phi_t^*Y = \nabla_Z Y$. We show that $\frac{d}{dt}\phi_{-t}^*\nabla_{\phi_t^*Z}\phi_t^*Y = 0$ for all sufficiently small t. The thing to notice here is that since the difference of connections is tensorial $\tau(Y, Z) = \frac{d}{dt}\phi_{-t}^*\nabla_{\phi_t^*Z}\phi_t^*Y$ is tensorial being the limit of the a difference quotient. Thus we can assume that $[X, Y] = [X, Z] = 0$. Thus $\phi_t^*Z = Z$ and $\phi_t^*Y = Y$. We now have

$$\begin{aligned}&\frac{d}{dt}\phi_{-t}^*\nabla_{\phi_t^*Z}\phi_t^*Y \\ &= \frac{d}{dt}\phi_{-t}^*(\nabla_Z Y) = \phi_{-t}^*[X, \nabla_Z Y] \\ &= \phi_{-t}^*(\nabla_{[X,Z]}Y + \nabla_Z[X, Y]) = 0\end{aligned}$$

But since $\phi_{-t}^*\nabla_{\phi_t^*Z}\phi_t^*Y$ is equal to $\nabla_Z Y$ when $t = 0$, we have that $\phi_{-t}^*\nabla_{\phi_t^*Z}\phi_t^*Y = \nabla_Z Y$ for all t which is (i). ■

Clearly the notion of a Jacobi field makes sense in the context of a general (torsion free) connection on M. Notice also that an automorphism ϕ of a pair (M, ∇) has the property that $\gamma \circ \phi$ is a geodesic if and only if ϕ is a geodesic.

Proposition 22.1 *X is a $\nabla-$Killing field if and only if $X \circ \gamma$ is a Jacobi field along γ for every geodesic γ.*

Proof. If X is Killing then $(s,t) \mapsto Fl_s^X(\gamma(t))$ is a variation of γ through geodesics and so $t \mapsto X \circ \gamma(t) = \frac{\partial}{\partial s}\big|_{s=0} Fl_s^X(\gamma(t))$ is a Jacobi field. The proof of the converse (Karcher) is as follows: Suppose that X is such that its restriction to any geodesic is a Jacobi field. Then for γ a geodesic, we have

$$\begin{aligned}\nabla_{\dot\gamma}^2 (X \circ \gamma) &= R(\dot\gamma, X \circ \gamma)\dot\gamma \\ &= \nabla_{\dot\gamma}\nabla_X \dot\gamma - \nabla_X \nabla_{\dot\gamma}\dot\gamma = \nabla_{\dot\gamma}\nabla_X \dot\gamma\end{aligned}$$

where we have used $\dot\gamma$ to denote not only a field along γ but also an extension to a neighborhood with $[\dot\gamma, X] = 0$. But

$$\begin{aligned}\nabla_{\dot\gamma}\nabla_X \dot\gamma &= \nabla_{\dot\gamma}\nabla_{\dot\gamma} X - \nabla_{\dot\gamma}[\dot\gamma, X] = \nabla_{\dot\gamma}\nabla_{\dot\gamma} X \\ &= \nabla_{\dot\gamma}\nabla_{\dot\gamma} X - \nabla_{\nabla_{\dot\gamma}\dot\gamma} X = \nabla_{\dot\gamma,\dot\gamma} X\end{aligned}$$

and so $\nabla_{\dot\gamma}^2 (X \circ \gamma) = \nabla_{\dot\gamma,\dot\gamma} X$. Now let $v, w \in T_pM$ for $p \in M$. Then there is a geodesic γ with $\dot\gamma(0) = v + w$ and so

$$\begin{aligned}&R(v,X)v + R(w,X)w + R(v,X)w + R(w,X)v \\ &= R(v+w,X)(v+w) = \nabla_{\dot\gamma,\dot\gamma} X = \nabla_{v+w,v+w} X \\ &= \nabla_{v,v} X + \nabla_{w,w} X + \nabla_{v,w} X + \nabla_{w,v} X\end{aligned}$$

Now replace w with $-w$ and subtract (polarization) to get

$$\nabla_{v,w} X + \nabla_{w,v} X = R(v,X)w + R(w,X)v.$$

On the other hand, $\nabla_{v,w} X - \nabla_{w,v} X = R(v,w)X$ and adding this equation to the previous one we get $2\nabla_{v,w} X = R(v,X)w - R(X,w)v - R(w,v)X$ and then by the Bianchi identity $2\nabla_{v,w} X = 2R(v,w)X$. The result now follows from theorem 22.1. ■

We now give equivalent conditions for $X \in \mathfrak{X}(M)$ to be a $g-$Killing field for a semi-Riemannian (M,g). First we need a lemma.

Lemma 22.2 *For any vector fields $X, Y, Z \in \mathfrak{X}(M)$ we have*

$$\begin{aligned}\mathcal{L}_X \langle Y, Z\rangle &= \langle \mathcal{L}_X Y, Z\rangle + \langle Y, \mathcal{L}_X Z\rangle + (\mathcal{L}_X g)(Y,Z) \\ &= \langle \mathcal{L}_X Y, Z\rangle + \langle Y, \mathcal{L}_X Z\rangle + \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle\end{aligned}$$

Proof. This is again a straightforward calculation using for the second equality

$$\begin{aligned}(\mathcal{L}_X g)(Y,Z) &= \mathcal{L}_X \langle Y, Z\rangle - \langle \mathcal{L}_X Y, Z\rangle - \langle Y, \mathcal{L}_X Z\rangle \\ &= \mathcal{L}_X \langle Y, Z\rangle + \langle \nabla_X Y - \nabla_Y X, Z\rangle + \langle Y, \nabla_X Z - \nabla_Z X\rangle \\ &= \mathcal{L}_X \langle Y, Z\rangle + \langle \nabla_X Y - \nabla_Y X, Z\rangle + \langle Y, \nabla_X Z - \nabla_Z X\rangle \\ &= \mathcal{L}_X \langle Y, Z\rangle - \langle \nabla_Y X, Z\rangle - \langle Y, \nabla_Z X\rangle + \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle \\ &0 + \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle\end{aligned}$$

■

Theorem 22.2 *Let (M, g) be semi-Riemannian. $X \in \mathfrak{X}(M)$ is a $g-$Killing field if and only if any one of the following conditions hold:*

(i) $\mathcal{L}_X g = 0$.

(ii) $\mathcal{L}_X \langle Y, Z \rangle = \langle \mathcal{L}_X Y, Z \rangle + \langle Y, \mathcal{L}_X Z \rangle$ *for all* $Y, Z \in \mathfrak{X}(M)$.

(iii) For each $p \in M$, the map $T_pM \to T_pM$ given by $v \longmapsto \nabla_v X$ is skewsymmetric for the inner product $\langle ., . \rangle_p$.

Proof. Since if X is Killing then $Fl_t^{X*} g = g$ and in general for a vector field X we have

$$\frac{d}{dt} Fl_t^{X*} g = Fl_t^{X*} \mathcal{L}_X g$$

the equivalence of (i) with the statement that X is $g-$Killing is immediate.

If (ii) holds then by the previous lemma, (i) holds and conversely. The equivalence of (ii) and (i) also follows from the lemma and is left to the reader. ■

Chapter 23

Comparison Theorems

23.0.1 Rauch's Comparison Theorem

In this section we deal strictly with Riemannian manifolds. Recall that for a (semi-) Riemannian manifold M, the sectional curvature $K^M(P)$ of a $2-$plane $P \subset T_pM$ is

$$\langle \mathfrak{R}\left(e_1 \wedge e_2\right), e_1 \wedge e_2 \rangle$$

for any orthonormal pair e_1, e_2 that span P.

Lemma 23.1 *If Y is a vector field along a curve $\alpha : [a, b] \to M$ then if $Y^{(k)}$ is the parallel transport of $\nabla_{\partial_t}^k Y(a)$ along α we have the following Taylor expansion:*

$$Y(t) = \sum_{k=0}^{m} \frac{Y^{(k)}(t)}{k!}(t-a)^k + O(|t-a|^{m+1})$$

Proof. Exercise. ∎

Definition 23.1 *If M, g and N, h are Riemannian manifolds and $\gamma^M : [a, b] \to M$ and $\gamma^N : [a, b] \to N$ are unit speed geodesics defined on the same interval $[a, b]$ then we say that $K^M \geq K^N$ along the pair (γ^M, γ^N) if $K^M(Q_{\gamma^M(t)}) \geq K^N(P_{\gamma^N(t)})$ for every all $t \in [a, b]$ and every pair of 2-planes $Q_{\gamma^M(t)} \in T_{\gamma^M(t)}M$, $P_{\gamma^N(t)} \in T_{\gamma^N(t)}N$.*

We develop some notation to be used in the proof of Rauch's theorem. Let M be a given Riemannian manifold. If Y is a vector field along a unit speed geodesic γ^M such that $Y(a) = 0$ then let

$$I_s^M(Y, Y) := I_{\gamma^M|[a,s]}(Y, Y) = \int_a^s \langle \nabla_{\partial_t} Y(t), \nabla_{\partial_t} Y(t) \rangle + \langle R_{\dot{\gamma}^M, Y} \dot{\gamma}^M, Y \rangle(t) dt.$$

If Y is an orthogonal Jacobi field then $I_s^M(Y, Y) = \langle \nabla_{\partial_t} Y, Y \rangle(s)$ by theorem 20.10.

Theorem 23.1 (Rauch) *Let M, g and N, h be Riemannian manifolds of the same dimension and let $\gamma^M : [a, b] \to M$ and $\gamma^N : [a, b] \to N$ unit speed geodesics defined on the same interval $[a, b]$. Let J^M and J^N be Jacobi fields along γ^M and γ^N respectively and orthogonal to their respective curves. Suppose that the following four conditions hold:*

(i) $J^M(a) = J^N(a)$ and neither of $J^M(t)$ or $J^N(t)$ is zero for $t \in (a, b]$
(ii) $\left\| \nabla_{\partial_t} J^M(a) \right\| = \left\| \nabla_{\partial_t} J^N(a) \right\|$
(iii) $L(\gamma^M) = dist(\gamma^M(a), \gamma^M(b))$
(iv) $K^M \geq K^N$ along the pair (γ^M, γ^N)

Then $\left\| J^M(t) \right\| \leq \left\| J^N(t) \right\|$ for all $t \in [a, b]$.

Proof. Let f_M be defined by $f_M(s) := \left\| J^M(s) \right\|$ and h_M by $h_M(s) := I^M_s(J^M, J^M)/\left\| J^M(s) \right\|^2$ for $s \in (a, b]$. Define f_N and h_N analogously. We have

$$f'_M(s) = 2I^M_s(J^M, J^M) \text{ and } f'_M/f_M = 2h_M$$

and the analogous equalities for f_N and h_N. If $c \in (a, b)$ then

$$\ln(\left\| J^M(s) \right\|^2) = \ln(\left\| J^M(c) \right\|^2) + 2\int_c^s h_M(s')ds'$$

with the analogous equation for N. Thus

$$\ln\left(\frac{\left\| J^M(s) \right\|^2}{\left\| J^N(s) \right\|^2}\right) = \ln\left(\frac{\left\| J^M(c) \right\|^2}{\left\| J^N(c) \right\|^2}\right) + 2\int_c^s [h_M(s') - h_N(s')]ds'$$

From the assumptions (i) and (ii) and the Taylor expansions for J^M and J^N we have

$$\lim_{c \to a} \frac{\left\| J^M(c) \right\|^2}{\left\| J^N(c) \right\|^2} = 0$$

and so

$$\ln\left(\frac{\left\| J^M(s) \right\|^2}{\left\| J^N(s) \right\|^2}\right) = 2\lim_{c \to a}\int_c^s [h_M(s') - h_N(s')]ds'$$

If we can show that $h_M(s) - h_N(s) \leq 0$ for $s \in (a, b]$ then the result will follow. So fix $s_0 \in (a, b]$ let $Z^M(s) := J^M(s)/\left\| J^M(r) \right\|$ and $Z^N(s) := J^N(s)/\left\| J^N(r) \right\|$. We now define a parameterized families of sub-tangent spaces along γ^M by $W_M(s) := \dot{\gamma}^M(s)^\perp \subset T_{\gamma^M(s)}M$ and similarly for $W_N(s)$. We can choose a linear isometry $L_r : W_N(r) \to W_M(r)$ such that $L_r(J^N(r)) = J^M(r)$. We now want to extend L_r to a family of linear isometries $L_s : W_N(s) \to W_M(s)$. We do this using parallel transport by

$$L_s := P(\gamma^M)^s_r \circ L_r \circ P(\gamma^N)^r_s.$$

Define a vector field Y along γ^M by $Y(s) := L_s(J^M(s))$. Check that

$$\begin{aligned} Y(a) &= J^M(a) \\ Y(r) &= J^M(r) \\ \|Y\|^2 &= \left\|J^N\right\|^2 \\ \|\nabla_{\partial_t} Y\|^2 &= \left\|\nabla_{\partial_t} J^N\right\|^2 \end{aligned}$$

The last equality is a result of exercise 19.10 where in the notation of that exercise $\beta(t) := P(\gamma^M)_t^r \circ Y(t)$. Since (iii) holds there can be no conjugates along γ^M up to r and so I_r^M is positive definite. Now $Y - J^M$ is orthogonal to the geodesic γ^M and so by corollary 20.4we have $I_r^M(J^M, J^M) \leq I_r^M(Y, Y)$ and in fact

$$\begin{aligned} I_r^M(J^M, J^M) \leq I_r^M(Y, Y) &= \int_a^r \|\nabla_{\partial_t} Y\|^2 + R^M(\dot\gamma^M, Y, \dot\gamma^M, Y) \\ &\leq \int_a^r \|\nabla_{\partial_t} Y\|^2 + R^N(\dot\gamma^N, J^N, \dot\gamma^N, J^N) \text{ (by (iv))} \\ &= I_r^N(J^N, J^N) \end{aligned}$$

Recalling the definition of Y we obtain

$$I_r^M(J^M, J^M)/\left\|J^M(r)\right\|^2 \leq I_r^N(J^N, J^N)/\left\|J^N(r)\right\|^2$$

and so $h_M(r) - h_N(r) \leq 0$ but r was arbitrary and so we are done. ∎

23.0.2 Bishop's Volume Comparison Theorem

`under construction`

23.0.3 Comparison Theorems in semi-Riemannian manifolds

`under construction`

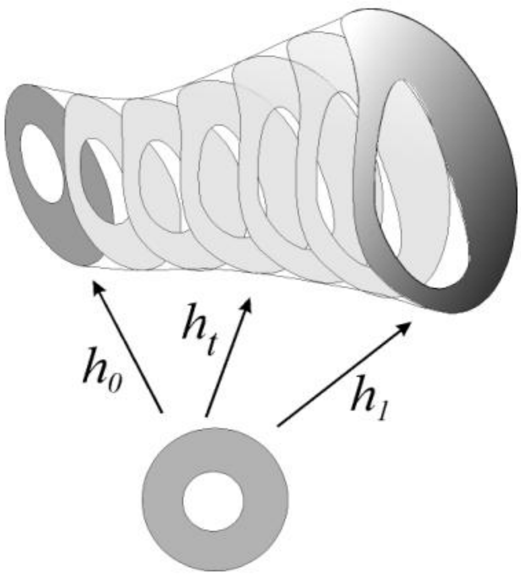
Chapter 24

Algebraic Topology

`UNDER CONSTRUCTION`

24.1 Topological Sum

24.2 Homotopy



Homotopy as a family of maps.

Definition 24.1 *Let $f_0, f_1 : X \to Y$ be maps. A homotopy from f_0 to f_1 is a one parameter family of maps $\{h_t : X \to Y : 0 \leq t \leq 1\}$ such that $h_0 = f_0$, $h_1 = f_1$ and such that $(x, t) \mapsto h_t(x)$ defines a (jointly continuous) map*

$X \times [0,1] \to Y$. If there exists such a homotopy we write $f_0 \simeq f_1$ and say that f_0 is homotopic to f_1. If there is a subspace $A \subset X$ such that $h_t|\, A = f_0|\, A$ for all $t \in [0,1]$ then we say that f_0 is homotopic to f_1 relative to A and we write $f_0 \simeq f_1(\mathrm{rel}A)$.

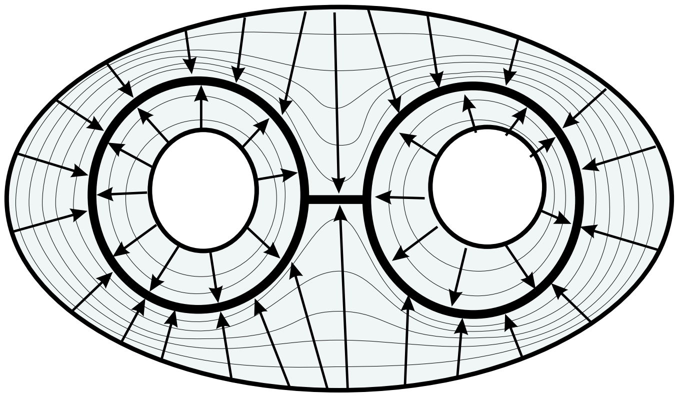
It is easy to see that homotopy equivalence is in fact an equivalence relation. The set of homotopy equivalence classes of maps $X \to Y$ is denoted $[X, Y]$ or $\pi(X, Y)$.

Definition 24.2 *Let $f_0, f_1 : (X, A) \to (Y\ , B)$ be maps of topological pairs. A homotopy from f_0 to f_1 is a homotopy h of the underlying maps $f_0, f_1 : X \to Y$ such that $h_t(A) \subset B$ for all $t \in [0,1]$. If $S \subset X$ then we say that f_0 is homotopic to f_1 relative to S if $h_t|\, S = f_0|\, S$ for all $t \in [0,1]$.*

The set of homotopy equivalence classes of maps $(X, A) \to (Y\ , B)$ is denoted $[(X, A), (Y, B)]$ or $\pi((X, A), (Y, B))$. As a special case we have the notion of a homotopy of **pointed maps** $f_0, f_1 : (X, x_0) \to (Y, y_0)$. The points x_0 and y_0 are called the **base points** and are commonly denoted by the generic symbol $*$. The set of all homotopy classes of pointed maps between pointed topological spaced is denoted $[(X, x_0), (Y, y_0)]$ or $\pi((X, x_0), (Y, y_0))$ but if the base points are fixed and understood then we denote the space of pointed homotopy classes as $[X, Y]_0$ or $\pi(X, Y)_0$. We may also wish to consider morphisms of **pointed pairs** such as $f : (X, A, a_0) \to (Y, B, b_0)$ which is given by a map $f : (X, A) \to (Y\ , B)$ such that $f(a_0) = b_0$. Here usually have $a_0 \in A$ and $b_0 \in B$. A homotopy between two such morphisms, say f_0 and $f_1 : (X, A, a_0) \to (Y, B, b_0)$ is a homotopy h of the underlying maps $(X, A) \to (Y\ , B)$ such that $h_t(a_0) = b_0$ for all $t \in [0,1]$. Clearly there are many variations on this theme of restricted homotopy.

Remark 24.1 *Notice that if $f_0, f_1 : (X, A) \to (Y\ , y_0)$ are homotopic as maps of topological pairs then we automatically have $f_0 \simeq f_1$(rel A). However, this is not necessarily the case if $\{y_0\}$ is replaced by a set $B \subset Y$ with more than one element.*

Definition 24.3 *A (strong)* ***deformation retraction*** *of X onto subspace $A \subset X$ is a homotopy f_t from $f_0 = \mathrm{id}_X$ to f_1 such that $f_1(X) \subset A$ and $f_t|\, A = \mathrm{id}_A$ for all $t \in [0,1]$. If such a retraction exists then we say that A is a (strong)* ***deformation retract*** *of X.*

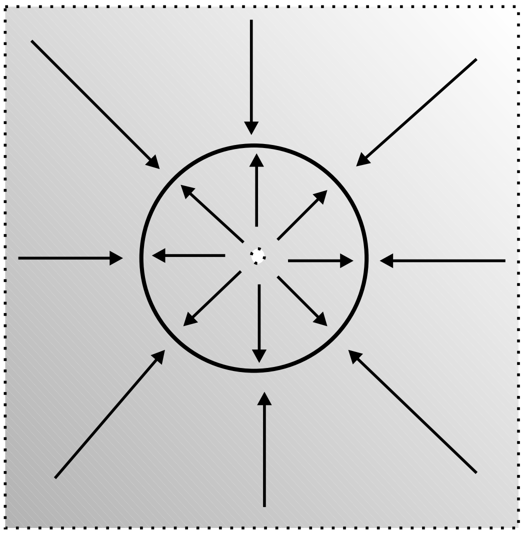


Retraction onto "eyeglasses"

Example 24.1 *Let $f_t : \mathbb{R}^n \backslash \{0\} \to \mathbb{R}^n \backslash \{0\}$ be defined by*

$$f_t(x) := t\frac{x}{|x|} + (1-t)x$$

for $0 \leq t \leq 1$. Then f_t gives a deformation retraction of $\mathbb{R}^n \backslash \{0\}$ onto $S^{n-1} \subset \mathbb{R}^n$.



Retraction of Punctured Plane

Definition 24.4 *A map $f : X \to Y$ is called a* ***homotopy equivalence*** *if there is a map $g : Y \to X$ such that $f \circ g \simeq \mathrm{id}_Y$ and $g \circ f \simeq \mathrm{id}_X$. The maps are then said to be homotopy inverses of each other. In this case we say that X and Y are* ***homotopy equivalent*** *and are said to be of the same* ***homotopy type****. We denote this relationship by $X \simeq Y$*

Definition 24.5 *A space X is called* ***contractible*** *if it is homotopy equivalent to a one point space.*

Definition 24.6 *A map $f : X \to Y$ is called* ***null-homotopic*** *if it is homotopic to a constant map.*

Equivalently, one can show that X is contractible if and only if every map $f : X \to Y$ is null-homotopic.

24.3 Cell Complexes

Let I denote the closed unit interval and let $I^n := I \times \cdots \times I$ be the n-fold Cartesian product of I with itself. The boundary of I is $\partial I = \{0, 1\}$ and the boundary of I^2 is $\partial I^2 = (I \times \{0, 1\}) \cup (\{0, 1\} \times I)$. More generally, the boundary of I^n is the union of the sets of the form $I \times \cdots \times \partial I \cdots \times I$. Also, recall that the closed unit n-disk D^n is the subset of $\mathbb{R}^n$ given by $\{|x|^2 \le 1\}$ and has as boundary the sphere S^{n-1}. From the topological point of view the pair $(I^n, \partial I^n)$ is indistinguishable from the pair (D^n, S^{n-1}). In other words , $(I^n, \partial I^n)$ is homeomorphic to (D^n, S^{n-1}).

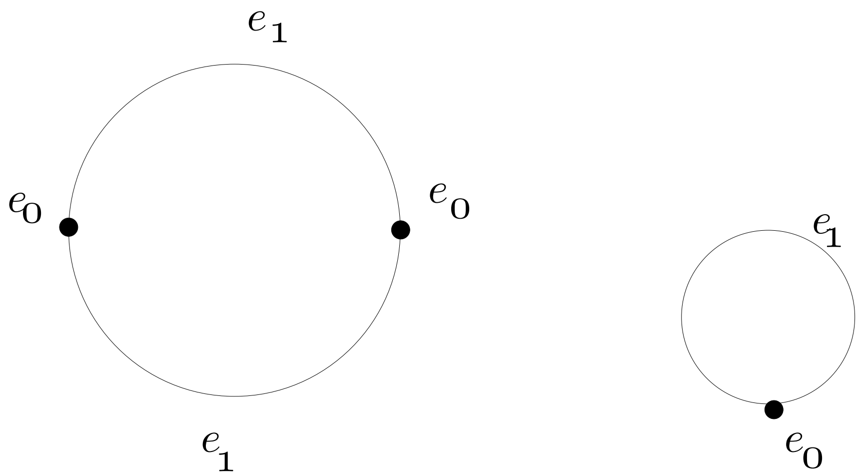
There is a generic notation for any homeomorphic copy of $I^n \cong D^n$ which is simply $\overline{e}^n$. Any such homeomorph of D^n is referred to as a closed n-cell. If we wish to distinguish several copies of such a space we might add an index to the notation as in $\overline{e}^n_1, \overline{e}^n_2$...etc. The interior of $\overline{e}^n$ is called an open n-cell and is generically denoted by e^n. The boundary is denoted by $\partial\overline{e}^n$ (or just ∂e^n). Thus we always have $(\overline{e}^n, \partial\overline{e}^n) \cong (D^n, S^{n-1})$.

An important use of the attaching idea is the construction of so called cell complexes . The open unit ball in $\mathbb{R}^n$ or any space homeomorphic to it is referred to as an open n-cell and is denoted by e^n. The closed ball is called a closed n-cell and has as boundary the $n-1$ sphere. A 0-cell is just a point and a 1-cell is a (homeomorph of) the unit interval the boundary of which is a pair of points. We now describe a process by which one can construct a large and interesting class of topological spaces called cell complexes. The steps are as follows:

1. Start with any discrete set of points and regard these as 0-cells.
2. Assume that one has completed the $n-1$ step in the construction with a resulting space X^{n-1}, construct X^n by attaching some number of copies of n-cells $\{e^n_\alpha\}_{\alpha \in A}$ (indexed by some set A) by attaching maps $f_\alpha : \partial e^n_\alpha = S^{n-1} \to X^{n-1}$.
3. Stop the process with a resulting space X^n called a **finite cell complex** or continue indefinitely according to some recipe and let $X = \bigcup_{n \ge 0} X^n$ and define a topology on X as follows: A set $U \subset X$ is defined to be open if and only if $U \cap X^n$ is open in X^n (with the relative topology). The space X is called a **CW complex** or just a **cell complex** .

Definition 24.7 *Given a cell complex constructed as above the set X^n constructed at the n-th step is called the n-***skeleton***. If the cell complex is finite* *then the highest step n reached in the construction is the whole space and the cell complex is said to have dimension n. In other words, a finite cell complex has* ***dimension*** *n if it is equal to its own nskeleton.*

It is important to realize that the stratification of the resulting topological space by the via the skeletons and also the open cells that are homeomorphically embedded are part of the definition of a cell complex and so two different cell complexes may in fact be homeomorphic without being the same cell complex. For example, one may realize the circle S^1 by attaching a 1-cell to a 0-cell or by attaching two 1-cells to two different 0-cells as in figure **??**.



Two cell structures for circle

Another important example is the projective space $P^n(\mathbb{R})$ which can be thought of as a the hemisphere $\overline{S^n_+} = \{x \in \mathbb{R}^{n+1} : x^{n+1} \geq 0\}$ modulo the identification of antipodal points of the boundary $\partial\overline{S^n_+} = S^{n-1}$. But S^{n-1} with antipodal point identified becomes $P^{n-1}(\mathbb{R})$ and so we can obtain $P^n(\mathbb{R})$ by attaching an n-cell e^n to $P^{n-1}(\mathbb{R})$ with the attaching map $\partial e^n = S^{n-1} \to P^{n-1}(\mathbb{R})$ given as the quotient map of S^{n-1} onto $P^{n-1}(\mathbb{R})$. By repeating this analysis inductively we conclude that $P^n(\mathbb{R})$ can be obtained from a point by attaching one cell from each dimension up to n :

$$P^n(\mathbb{R}) = e^0 \cup e^1 \cup \cdots \cup e^n$$

and so $P^n(\mathbb{R})$ is a finite cell complex of dimension n.

24.4 Axioms for a Homology Theory

Consider the category $\mathcal{TP}$ of all topological pairs (X, A) where X is a topological space, A is a subspace of X and where a morphism $f : (X, A) \to (X', A')$ is given by a map $f : X \to X'$ such that $f(A) \subset A'$. We may consider the category of topological spaces and maps as a subcategory of $\mathcal{TP}$ by identifying $(X, \emptyset)$ with X. We will be interested in functors from some subcategory $\mathcal{NTP}$ to the category $\mathbb{Z}-\mathcal{GAG}$ of $\mathbb{Z}$-graded abelian groups. The subcategory $\mathcal{NTP}$ (tentatively called "nice topological pairs") will vary depending of the situation but one example for which things work out nicely is the category of finite cell complex pairs. Let $\sum A_k$ and $\sum B_k$ be graded abelian groups. A morphism of $\mathbb{Z}$-graded abelian groups is a sequence $\{h_k\}$ of group homomorphisms $h_k : A_k \to B_k$. Such a morphism may also be thought of as combined to give a degree preserving map on the graded group; $h : \sum A_k \to \sum B_k$.

In the following we write $H_p(X)$ for $H_p(X, \emptyset)$. A homology theory H with coefficient group G is a covariant functor h_G from a category of nice topological pairs $\mathcal{NTP}$ to the category $\mathbb{Z}-\mathcal{GAG}$ of $\mathbb{Z}$-graded abelian groups;

$$h_G : \begin{cases} (X, A) \mapsto H(X, A, G) = \sum_{p \in \mathbb{Z}} H_p(X, A, G) \\ \qquad\qquad f \mapsto f_* \end{cases}$$

and that satisfies the following axioms:

1. $H_p(X, A) = 0$ for $p < 0$.

2. (**Dimension axiom**) $H_p(pt) = 0$ for all $p \geq 1$ and $H_0(pt) = G$.

3. If $f : (X, A) \to (X', A')$ is homotopic to $g : (X, A) \to (X', A')$ then $f_* = g_*$

4. (**Boundary map axiom**) To each pair (X, A) and each $p \in \mathbb{Z}$ there is a boundary homomorphism $\partial_p : H_p(X, A; G) \to H_{p-1}(A; G)$ such that for all maps $f : (X, A) \to (X', A')$ the following diagram commutes:

$$\begin{array}{ccc} H_p(X, A; G) & \xrightarrow{f_*} & H_p(X', A'; G) \\ \partial_p \downarrow & & \partial_p \downarrow \\ H_{p-1}(A; G) & \underset{(f|A)_*}{\rightarrow} & H_{p-1}(A'; G) \end{array}$$

5. (**Excision axiom**) For each inclusion $\iota : (B, B \cap A) \to (A \cup B, A)$ the induced map $\iota_* : H(B, B \cap A; G) \to H(A \cup B, A; G)$ is an isomorphism.

6. For each pair (X, A) and inclusions $i : A \hookrightarrow X$ and $j : (X, \emptyset) \hookrightarrow (X, A)$ there is a long exact sequence

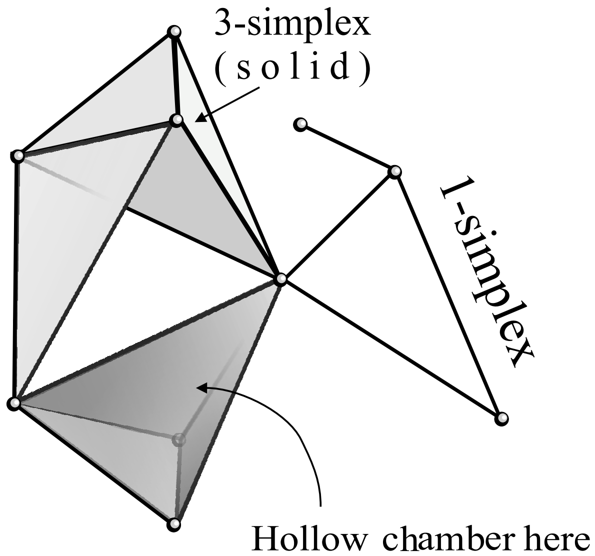
$$\begin{array}{cccc} \cdots \to H_{p+1}(A) \xrightarrow{i_*} H_{p+1}(X) & \xrightarrow{j_*} & H_{p+1}(X, A) & \\ & & \partial_{p+1} \swarrow & \\ & & H_{p+1}(A) \xrightarrow{i_*} & H_{p+1}(X) \xrightarrow{j_*} \cdots \end{array}$$

where we have suppressed the reference to G for brevity.

24.5 Simplicial Homology

Simplicial homology is a perhaps the easiest to understand in principle.

And we have



Simplicial Complex

24.6 Singular Homology

These days an algebraic topology text is like to emphasize singular homology.

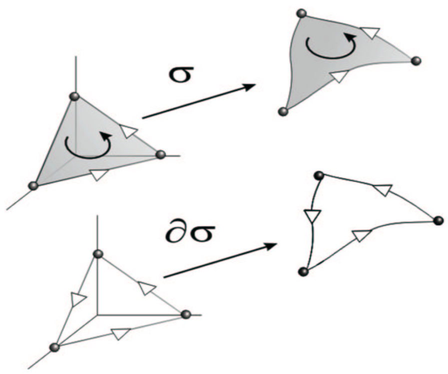


Figure 24.1: Singular 2-simplex and boundary

24.7 Cellular Homology

UNDER CONSTRUCTION

24.8 Universal Coefficient theorem

UNDER CONSTRUCTION

24.9 Axioms for a Cohomology Theory

UNDER CONSTRUCTION

24.10 Topology of Vector Bundles

In this section we study vector bundles with finite rank. Thus, the typical fiber may be taken to be $\mathbb{R}^n$ (or $\mathbb{C}^n$ for a complex vector bundle) for some positive integer n. We would also like to study vectors bundles over spaces that are not necessarily differentiable manifolds; although this will be our main interest. All the spaces in this section will be assumed to be paracompact Hausdorff spaces. We shall refer to continuous maps simply as maps. In many cases the theorems will makes sense in the differentiable category and in this case one reads map as "smooth map".

Recall that a (rank n) real vector bundle is a triple (π_E, E, M) where E and M are paracompact spaces and $\pi_E : E \to M$ is a surjective map such that there is a cover of M by open sets U_α together with corresponding trivializing maps (VB-charts) $\phi_\alpha : \pi_E^{-1}(U_\alpha) \to U_\alpha \times \mathbb{R}^n$ of the form $\phi_\alpha = (\pi_E, \Phi_\alpha)$. Here $\Phi_\alpha : \pi_E^{-1}(U_\alpha) \to \mathbb{R}^n$ has the property that $\Phi_\alpha|_{E_x} : E_x \to \mathbb{R}^n$ is a diffeomorphism for each fiber $E_x := \pi_E^{-1}(x)$. Furthermore, in order that we may consistently transfer the linear structure of $\mathbb{R}^n$ over to E_x we must require that when $U_\alpha \cap U_\beta \neq \emptyset$ and $x \in U_\alpha \cap U_\beta$ then function

$$\Phi_{\beta\alpha;x} = \Phi_\beta|_{E_x} \circ \Phi_\alpha|_{E_x}^{-1} : \mathbb{R}^n \to \mathbb{R}^n$$

is a linear isomorphism. Thus the fibers are vectors spaces isomorphic to $\mathbb{R}^n$. For each nonempty overlap $U_\alpha \cap U_\beta$ we have a map $U_\alpha \cap U_\beta \to \mathrm{GL}(n)$

$$x \mapsto \Phi_{\beta\alpha;x}.$$

We have already seen several examples of vector bundles but let us add one more to the list:

Example 24.2 *The normal bundle to $S^n \subset \mathbb{R}^{n+1}$ is the subset $N(S^n)$ of $S^n \times \mathbb{R}^{n+1}$ given by*

$$N(S^n) := \{(x, v) : x \cdot v = 0\}.$$

The bundle projection $\pi_{N(S^n)}$ *is given by* $(x, v) \mapsto x$. *We may define bundle charts by taking opens sets* $U_\alpha \subset S^n$ *that cover* S^n *and then since any* $(x, v) \in \pi_{N(S^n)}^{-1}(U_\alpha)$ *is of the form* (x, tx) *for some* $t \in \mathbb{R}$ *we may define*

$$\phi_\alpha : (x, v) = (x, tx) \mapsto (x, t).$$

Now there is a very important point to be made from the last example. Namely, it seems we could have taken *any* cover $\{U_\alpha\}$ with which to build the VB-charts. But can we just take the cover consisting of the single open set $U_1 := S^n$ and thus get a VB-chart $N(S^n) \to S^n \times \mathbb{R}$? The answer is that in this case we can. This is because $N(S^n)$ is itself a trivial bundle; that is, it is isomorphic to the product bundle $S^n \times \mathbb{R}$. This is not the case for vector bundles in general. In particular, we will later be able to show that the tangent bundle of an even dimensional sphere is always nontrivial. Of course, we have already seen that the Möbius line bundle is nontrivial.

24.11 De Rham's Theorem

UNDER CONSTRUCTION

24.12 Sheaf Cohomology

UNDER CONSTRUCTION

24.13 Characteristic Classes

UNDER CONSTRUCTION

Chapter 25

Ehressman Connections and Cartan Connections

Halmos, Paul R.

Don't just read it; fight it! Ask your own questions, look for your own examples, discover your own proofs. Is the hypothesis necessary? Is the converse true? What happens in the classical special case? What about the degenerate cases? Where does the proof use the hypothesis?

Halmos, Paul R.

I Want to be a Mathematician, Washington: MAA Spectrum, 1985.

25.1 Principal and Associated $G-$Bundles

In this chapter we take a (C^r) **fiber bundle** over M with typical fiber F to be a quadruple (π, E, M, F) where $\pi : E \to M$ is a smooth C^r-submersion such that for every $p \in M$ there is an open set U containing p with a C^r-isomorphism $\phi = (\pi, \Phi) : \pi^{-1}(U) \to U \times F$. To review; we denote the fiber at p by $E_p = \pi^{-1}(p)$ and for each $p \in U$ the map $\Phi|_{E_p} : E_p \to F$ is a C^r-diffeomorphism. Given two such trivializations $(\pi, \Phi_\alpha) : \pi^{-1}(U_\alpha) \to U_\alpha \times F$ and $(\pi, \Phi_\beta) : \pi^{-1}(U_\beta) \to U_\beta \times F$ then for each $p \in U_\alpha \cap U_\beta$ there is a diffeomorphism $\Phi_{\alpha\beta}|_p : E_p \to E_p$ $U_\alpha \cap U_\beta \to \mathrm{Diff}(F)$ defined by $p \mapsto \Phi_{\alpha\beta}(p) = \Phi_{\alpha\beta}|_p$. These are called **transition maps** or **transition functions**.

Remark 25.1 *Recall that a group action $\rho : G \times F \to F$ is equivalently thought of as a representation $\overline{\rho} : G \to \mathrm{Diff}(F)$ given by $\overline{\rho}(g)(f) = \rho(g, f)$. We will forgo the separate notation $\overline{\rho}$ and simple write ρ for the action and the corresponding representation.*

Returning to our discussion of fiber bundles, suppose that there is a Lie group action $\rho : G \times F \to F$ and cover of E by trivializations $\{U_\alpha, \phi_\alpha\}$ such that for each α, β we have

$$\Phi_{\alpha\beta}(p)(f) = \rho(g_{\alpha\beta}(p), f) := g_{\alpha\beta}(p) \cdot f$$

for some smooth map $g_{\alpha\beta} : U_\alpha \cap U_\beta \to G$ then we say that we have presented (π, E, M, F) as G **bundle under the representation** ρ. We also say that the transition functions **live in** G (via ρ). In many but not all cases the representation ρ will be faithful, i.e. the action will be **effective** and so G can be considered as a subgroup of $\mathrm{Diff}(F)$. A notable exception is the case of spin bundles. We call (π, E, M, F, G) a (G, ρ) **bundle** or just a $G-$bundle if the representation is understood or standard in some way. It is common to call G the **structure group** but since the action in question may not be effective we should really refer to the **structure group representation (or action)** ρ.

A fiber bundle is determined if we are given an open cover $\mathcal{U} = \{U_\alpha\}$ and maps $\Phi_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{diff}(F)$ such that for all α, β, γ

$$\begin{aligned} \Phi_{\alpha\alpha}(p) &= \mathrm{id}\,\text{for } p \in U_\alpha \\ \Phi_{\alpha\beta}(p) &= \Phi_{\beta\alpha}^{-1}(p) \text{ for } p \in U_\alpha \cap U_\beta \\ \Phi_{\alpha\beta}(p) \circ \Phi_{\beta\gamma}(p) \circ \Phi_{\gamma\alpha}(p) &= \text{ id for } p \in U_\alpha \cap U_\beta \cap U_\gamma \end{aligned}$$

If we want a G bundle under a representation ρ then we further require that $\Phi_{\alpha\beta}(p)(f) = \rho(g_{\alpha\beta}(p))(f)$ as above and that the maps $g_{\alpha\beta}$ themselves satisfy the cocycle condition:

$$\begin{aligned} g_{\alpha\alpha}(p) &= \mathrm{id} \text{ for } p \in U_\alpha \\ g_{\alpha\beta}(p) &= g_{\beta\alpha}^{-1}(p) \text{ for } p \in U_\alpha \cap U_\beta \\ g_{\alpha\beta}(p) \circ g_{\beta\gamma}(p) \circ g_{\gamma\alpha}(p) &= \mathrm{id} \text{ for } p \in U_\alpha \cap U_\beta \cap U_\gamma \end{aligned} \tag{25.1}$$

We shall also call the maps $g_{\alpha\beta}$ **transition functions** or **transition maps**. Notice that if ρ is effective the last condition follows from the first. The family $\{U_\alpha\}$ together with the maps $\Phi_{\alpha\beta}$ form a **cocycle** and we can construct a bundle by taking the disjoint union $\bigsqcup(U_\alpha \times F) = \bigcup U_\alpha \times F \times \{\alpha\}$ and then taking the equivalence classes under the relation $(p, f, \beta) \backsim (p, \Phi_{\alpha\beta}(p)(f), \alpha)$ so that

$$E = \left(\bigcup U_\alpha \times F \times \{\alpha\}\right) / \backsim$$

and $\pi([p, f, \beta]) = p$.

Let $H \subset G$ be a closed subgroup. Suppose that we can, by throwing out some of the elements of $\{ U_\alpha, \Phi_\alpha\}$ arrange that all of the transition functions live in H. That is, suppose we have that $\Phi_{\alpha\beta}(p)(f) = \rho(g_{\alpha\beta}(p), f)$ where $g_{\alpha\beta} : U_\alpha \cap U_\beta \to H$. Then we have a **reduction the structure group (**or **reduction of the structure representation** in case the action needs to be specified).

Next, suppose that we have an surjective Lie group homomorphism $h : \overline{G} \to G$ **.** We then have the lifted representation $\overline{\rho} : \overline{G} \times F \to F$ given by $\overline{\rho}(\overline{g}, f) = \rho(h(\overline{g}), f)$. Under suitable topological conditions we may be able to lift the maps $g_{\alpha\beta}$ to maps $\overline{g}_{\alpha\beta} : U_\alpha \cap U_\beta \to \overline{G}$ and by choosing a subfamily we can even arrange that the $\overline{g}_{\alpha\beta}$ satisfy the cocycle condition. Note that $\Phi_{\alpha\beta}(p)(f) = \rho(g_{\alpha\beta}, f) = \rho(h(\overline{g}_{\alpha\beta}), f) = \overline{\rho}(\overline{g}_{\alpha\beta}(p), f)$. In this case we say that we have **lifted** the structure representation to $\overline{\rho}$.

Example 25.1 *The simplest class of examples of fiber bundles over a manifold M are the product bundles. These are just Cartesian products $M \times F$ together with the projection map $pr_1 : M \times F \to M$. Here, the structure group can be reduced to the trivial group $\{e\}$ acting as the identity map on F. On the other hand, this bundle can also be prolonged to any Lie group acting on F.*

Example 25.2 *A covering manifold $\pi : \widetilde{M} \to M$ is a G-bundle where G is the group of deck transformations. In this example the group G is a discrete (0-dimensional) Lie group.*

Example 25.3 (The Hopf Bundle) *Identify S^1 as the group of complex numbers of unit modulus. Also, we consider the sphere S^3 as it sits in $\mathbb{C}^2$:*

$$S^3 = \{(z_1, z_2) \in \mathbb{C}^2 : |z_1|^2 + |z_2|^2 = 1\}.$$

The group S^1 acts on S^2 by $u \cdot (z_1, z_2) = (uz_1, uz_2)$. Next we get S^2 into the act. We want to realize S^2 as the sphere of radius 1/2 in $\mathbb{R}^3$ and having two coordinate maps coming from stereographic projection from the north and south poles onto copies of $\mathbb{C}$ embedded as planes tangent to the sphere at the two poles. The chart transitions then have the form $w = 1/z$. Thus we may view S^2 as two copies of $\mathbb{C}$, say the z plane $\mathbb{C}_1$ and the w plane $\mathbb{C}_2$ glued together under the identification $\phi : z \mapsto 1/z = w$

$$S^2 = \mathbb{C}_1 \cup_\phi \mathbb{C}_2.$$

With this in mind define a map $\pi : S^3 \subset \mathbb{C}^2 \to S^2$ by

$$\pi(z_1, z_2) = \begin{cases} z_2/z_1 \in \mathbb{C}_2 \subset \mathbb{C}_1 \cup_\phi \mathbb{C}_2 & \text{if} \quad z_1 \neq 0 \\ z_1/z_2 \in \mathbb{C}_1 \subset \mathbb{C}_1 \cup_\phi \mathbb{C}_2 & \text{if} \quad z_2 \neq 0 \end{cases}.$$

Note that this gives a well defined map onto S^2.

Claim 25.1 $u \cdot (z_1, z_2) = u \cdot (w_1, w_2)$ *if and only if* $\pi(z_1, z_2) = \pi(w_1, w_2)$.

Proof. If $u \cdot (z_1, z_2) = u \cdot (w_1, w_2)$ and $z_1 \neq 0$ then $w_1 \neq 0$ and $\pi(w_1, w_2) = w_2/w_1 = uw_2/uw_1 = \pi(w_1, w_2) = \pi(z_1, z_2) = uz_2/uz_1 = z_2/z_1 = \pi(z_1, z_2)$. A similar calculation show applies when $z_2 \neq 0$. On the other hand, if $\pi(w_1, w_2) = \pi(z_1, z_2)$ then by a similar chain of equalities we also easily get that $u \cdot (w_1, w_2) = ... = \pi(w_1, w_2) = \pi(z_1, z_2) = ... = u \cdot (z_1, z_2)$. ∎

Using these facts we see that there is a fiber bundle atlas on $\pi_{Hopf} = \pi : S^3 \to S^2$ given by the following trivializations:

$$\varphi_1 : \pi^{-1}(C_1) \to C_1 \times S^1$$
$$\varphi_1 : (z_1, z_2) = (z_2/z_1, z_1/|z_1|)$$

and

$$\varphi_2 : \pi^{-1}(C_2) \to C_2 \times S^1$$
$$\varphi_2 : (z_1, z_2) = (z_1/z_2, z_2/|z_2|).$$

The transition map is

$$(z, u) \mapsto (1/z, \frac{z}{|z|}u)$$

which is of the correct form since $u \mapsto \frac{z}{|z|} \cdot u$ is a circle action. Thus the Hopf bundle is an S^1-bundle with typical fiber S^1 itself. It can be shown that the inverse image of a circle on S^2 by the Hopf projection π_{Hopf} is a torus. Since the sphere S^2 is foliated by circles degenerating at the poles we have a foliation of S^3-{two circles} by tori degenerating to circles at the fiber over the two poles. Since $S^3\backslash\{\text{pole}\}$ is diffeomorphic to $\mathbb{R}^3$ we expect to be able to get a picture of this foliation by tori. In fact, the following picture depicts this foliation.

25.2 Principal and Associated Bundles

An important case for a bundle with structure group G is where the typical fiber is the group itself. In fact we may obtain such a bundle by taking the transition functions $g_{\alpha\beta}$ from any effective G bundle $E \to M$ or just any smooth maps $g_{\alpha\beta} : U_\alpha \cap U_\beta \to G$ that form a cocycle with respect to some cover $\{U_\alpha\}$ of M. We let G act on itself by left multiplication and then use the bundle construction method above. Thus if $\{U_\alpha\}$ is the cover of M corresponding to the cocycle $\{g_{\alpha\beta}\}_{\alpha,\beta}$ then we let

$$P = \left(\bigcup U_\alpha \times G \times \{\alpha\}\right) / \backsim$$

where $(p, g, \alpha) \backsim (p, g_{\alpha\beta}(p)g, \beta)$ gives the equivalence relation. In this way we construct what is called the a principal bundle. Notice that for $g \in G$ we have $(p, g_1, \beta) \backsim (p, g_2, \alpha)$ if and only if $(p, g_1 g, \beta) \backsim (p, g_2 g, \alpha)$ and so there is a well defined right action on any bundle principal bundle. On the other hand there is a more direct way chart free way to define the notion of principal bundle. The advantage of defining a principal bundle without explicit reference to transitions functions is that we may then use the principal bundle to give another definition of a G-bundle that doesn't appeal directly to the notion of transition functions. We will see that every G bundle is given by a choice of a principal G-bundle and an action of G on some manifold F (the typical fiber).

First we define the **trivial principal G bundle over U** to be the trivial bundle $pr_1 : U \times G \to M$ together with the right G action $(U \times G) \times G$ given by

$$(x, g_1)g := (x, g_1 g).$$

An automorphism of the G-space $U \times G$ is a bundle map $\delta : U \times G \to U \times G$ such that $\delta(x, g_1 g) = \delta(x, g_1)g$ for all $g_1, g \in G$ and all $x \in U$. Now δ must have the form given by $\delta(x, g) = (x, \Delta(x, g))$ and so

$$\Delta(x, g) = \Delta(x, e)g.$$

If we then let the function $x \mapsto \Delta(x,e)$ be denoted by $g_\delta()$ then we have $\delta(x,g) = (x, g_\delta(x)g)$. Thus we obtain the following

Lemma 25.1 *Every automorphism of a trivial principal G bundle over and open set U has the form $\delta : (x,g) \mapsto (x, g_\delta(x)g)$ for some smooth map $g_\delta : U \to G$.*

Definition 25.1 *A **principal G-bundle** is a fiber bundle $\pi_P : P \to M$ together with a right G action $P \times G \to P$ that is locally equivalent as a right G space to the trivial principal G bundle over M. This means that for each point $x \in M$ there is an open neighborhood U_x and a trivialization ϕ*

$$\begin{array}{ccc} \pi_P^{-1}(U_x) & \xrightarrow{\phi} & U_x \times G \\ & \searrow \quad \swarrow & pr_1 \\ & U_x & \end{array}$$

*that is G equivariant. Thus we require that $\phi(pg) = \phi(p)g$. We shall call such a trivialization an **equivariant trivialization**.*

Note that $\phi(pg) = (\pi_P(pg), \Phi(pg))$ while on the other hand $\phi(p)g = (\pi_P(pg), \Phi(p)g)$ so it is necessary and sufficient that $\Phi(p)g = \Phi(pg)$. Now we want to show that this means that the structure representation of $\pi_P : P \to M$ is left multiplication by elements of G. Let ϕ_1, U_1 and ϕ_2, U_2 be two equivariant trivializations such that $U_1 \cap U_2 \neq \emptyset$. On the overlap we have the diagram

$$\begin{array}{ccccc} U_1 \cap U_2 \times G & \xleftarrow{\phi_2} & \pi_P^{-1}(U_1 \cap U_2) & \xrightarrow{\phi_1} & U_1 \cap U_2 \times G \\ & \searrow & \downarrow & \swarrow & \\ & & U_1 \cap U_2 & & \end{array}$$

The map $\phi_2 \circ \phi_1^{-1}\big|_{U_1 \cap U_2}$ clearly must an G-bundle automorphism of $U_1 \cap U_2 \times G$ and so by 25.1 must have the form $\phi_2 \circ \phi_1^{-1}\big|_{U_1 \cap U_2}(x,g) = (x, \Phi_{12}(x)g)$. We conclude that a principal G-bundle is a G-bundle with typical fiber G as defined in section 25.1. The maps on overlaps such as Φ_{12} are the transition maps. Notice that $\Phi_{12}(x)$ acts on G by left multiplication and so the structure representation is left multiplication.

Proposition 25.1 *If $\pi_P : P \to M$ is a principal G-bundle then the right action $P \times G \to P$ is free and the action restricted to any fiber is transitive.*

Proof. Suppose $p \in P$ and $pg = p$ for some $g \in G$. Let $\pi_P^{-1}(U_x) \xrightarrow{\phi} U_x \times G$ be an (equivariant) trivialization over U_x where U_x contains $\pi_P(p) = x$. Then we have

$$\begin{aligned} \phi(pg) = \phi(p) \quad & \Rightarrow \\ (x, g_0 g) = (x, g_0) \quad & \Rightarrow \\ g_0 g = g_0 & \end{aligned}$$

and so $g = e$.

Now let $P_x = \pi_P^{-1}(x)$ and let $p_1, p_2 \in P_x$. Again choosing an (equivariant) trivialization over U_x as above we have that $\phi(p_i) = (x, g_i)$ and so letting $g := g_1^{-1} g_2$ we have $\phi(p_1 g) = (x, g_1 g) = (x, g_2) = \phi(p_2)$ and since ϕ is injective $p_1 g = p_2$. ■

The reader should realize that this result is in some sense "obvious" since the upshot is just that the result is true for the trivial principal bundle and then it follows for the general case since a general principal bundle is locally G-bundle isomorphic to a trivial principal bundle.

Remark 25.2 *Some authors define a principal bundle to be fiber bundle with typical fiber G and with a free right action that is transitive on each fiber. This approach turns out to be equivalent to the present one.*

Our first and possibly most important example of a principal bundle is the frame bundle associated to a rank k vector bundle $\pi : E \to M$. To achieve an appropriate generality let us assume that the structure group of the bundle can be reduced to some matrix group G (e.g. $SO(k)$). In this case we may single out a special class of frames in each fiber which are related to each other by elements of the group G. Let us call these frames "$G-$frames". A moving frame $\sigma_\alpha = (F_1, ..., F_k)$ which is such that $\sigma_\alpha(x)$ is always a $G-$frame will be called a moving G-frame. Let $P_{G,x}(\pi)$ be the set of all $G-$frames at x and define

$$P_G(\pi) = \bigcup_{x \in Mx} P_{G,x}(\pi).$$

Also, let $\wp : GF(\pi) \to M$ be the natural projection map that takes any frame $\mathcal{F}_x \in GF(\pi)$ with $\mathcal{F}_x \in GF_x(\pi)$ to its base x. A moving $G-$frame is clearly the same thing as a local section of $GF(\pi)$. $GF(\pi)$ is a smooth manifold and in fact a principal bundle. We describe a smooth atlas for $GF(\pi)$. Let us adopt the convention that $(\mathcal{F}_x)_i := f_i$ is the i-th vector in the frame $\mathcal{F}_x = (f_1, ..., f_n)$. Let $\{U_\alpha, \mathtt{x}_\alpha\}_{\alpha \in A}$ be an atlas for M. We can assume without loss that each chart domains U_α is also the domain of a moving $G-$frame $\sigma_\alpha : U_\alpha \to GF(\pi)$ and use the notation $\sigma_\alpha = (F_1, ..., F_k)$. For each chart α on we define a chart $\widetilde{U}_\alpha, \widetilde{\mathtt{x}}_\alpha$ by letting

$$\widetilde{U}_\alpha := \wp^{-1}(U_\alpha) = \bigcup_{x \in U_\alpha} GF_x(\pi)$$

and

$$\widetilde{\mathtt{x}}_\alpha(\mathcal{F}_x) := (x^i(x), f_i^j) \in \mathbb{R}^n \times \mathbb{R}^{k \times k}$$

where (f_j^i) is matrix such that $(\mathcal{F}_x)_i = \sum f_i^j F_j(x)$. We leave it to the reader to find the change of coordinate maps and see that they are smooth. The right G action on $GF(\pi)$ is given by matrix multiplication $(\mathcal{F}_x, g) \mapsto \mathcal{F}_x g$ where we think of $\mathcal{F}_x$ as row of basis vectors.

Example 25.4 (Full frame bundle of a manifold) *The frame bundle of the tangent bundle of a manifold M is the set of all frames without restriction and* *so the structure group is* $\mathrm{GL}(n)$ *where* $n = \dim(M)$*. This frame bundle (usually called* the *frame bundle of* M*) is also traditionally denoted* $L(M)$*.*

Example 25.5 (Orthonormal Frame Bundle) *If* $E \to M$ *is a real vector bundle with a bundle metric then the structure group is reduced to* $\mathrm{O}(k)$ *and the corresponding frame bundle is denoted* $P_{\mathrm{O}(k)}(\pi)$ *or* $P_{\mathrm{O}(k)}(E)$ *and is called the orthonormal frame bundle of* E*. In the case of the tangent bundle of a Riemannian manifold* M, g *we denote the orthonormal frame of* TM *by* $P_{\mathrm{O}(n)}(TM)$ *although the notation* $F(M)$ *is also popular in the literature.*

25.3 Ehressman Connections

UNDER CONSTRUCTION

As we have pointed out, a moving $G-$frame on a vector bundle may clearly be thought of as a local section of the frame bundle $\sigma : U \to F(E)$. If

25.4 Gauge Fields

UNDER CONSTRUCTION

25.5 Cartan Connections

UNDER CONSTRUCTION

Part II

Part II

Chapter 26

Analysis on Manifolds

The best way to escape from a problem is to solve it.
-Alan Saporta

26.1 Basics

Now E is a Hermitian or a Riemannian vector bundle. First, if E is a Riemannian vector bundle then so is $E \otimes T^*M^{\otimes k}$ since T^*M also has a metric (locally given by g^{ij}). If $\xi = \xi^r_{i_1 \dots i_k} \epsilon_r \otimes \theta^{i_1} \otimes \cdots \otimes \theta^{i_k}$ and $\mu = \mu^r_{i_1 \dots i_k} \epsilon_r \otimes \theta^{i_1} \otimes \cdots \otimes \theta^{i_k}$ then

$$\langle \xi, \mu \rangle = g^{i_1 j_1} \cdots g^{i_1 j_1} h_{sr} \xi^r_{i_1 \dots i_k} \mu^s_{j_1 \dots j_k}$$

We could denote the metric on $E \otimes T^*M^{\otimes k}$ by $h \otimes (g^{-1})^{\otimes k}$ but we shall not often have occasion to use this notation will stick with the notation $\langle ., . \rangle$ whenever it is clear which metric is meant.

Let M, g be an oriented Riemannian manifold. The volume form vol_g allows us to integrate smooth functions of compact support. This gives a functional $C^\infty_c(M) \longrightarrow \mathbb{C}$ which extends to a positive linear functional on $C^0(M)$. A standard argument from real analysis applies here and gives a measure μ_g on the Borel sigma algebra $\mathcal{B}(M)$ generated by open set on M which is characterized by the fact that for every open set $\mu_g(O) = \sup\{\int f \, vol_g : f \prec O\}$ where $f \prec O$ means that $\operatorname{supp} f$ is a compact subset of O and $0 \leq f \leq 1$. We will denote integration with respect to this measure by $f \mapsto \int_M f(x) \mu_g(dx)$ or by a slight abuse of notation $\int f \, vol_g$. In fact, if f is smooth then $\int_M f(x) \mu_g(dx) = \int_M f \, vol_g$. We would like to show how many of the spaces and result from analysis on $\mathbb{R}^n$ still make sense in a more global geometric setting.

26.1.1 L^2, L^p, L^∞

Let $\pi : E \to M$ be a Riemannian vector bundle or a Hermitian vector bundle. Thus there is a real or complex inner product $h_p() = \langle ., . \rangle_p$ on every fiber E_p which varies smoothly with $p \in M$. Now if $v_x \in E_x$ we let $|v_x| := \langle v_x, v_x \rangle^{1/2} \in \mathbb{R}$

and for each section $\sigma \in \Gamma(M)$ let $|\sigma| = \langle \sigma, \sigma \rangle^{1/2} \in C^0(M)$. We need a measure on the base space M and so for convenience we assume that M is oriented and has a Riemannian metric. Then the associated volume element vol_g induces a Radon measure which is equivalent to Lebesgue measure in every coordinate chart. We now define the norms

$$\|\sigma\|_p := \left(\int_M |\sigma|^p \, vol_g \right)^{1/p}$$
$$\|\sigma\|_\infty := \sup_{x \in M} |\sigma(x)|$$

First of all we must allow sections of the vector bundle which are not necessarily C^∞. It is quite easy to see what it means for a section to be continuous and a little more difficult but still rather easy to see what it means for a section to be measurable.

Definition 26.1 $L^p_g(M, E)$ *is the space of measurable sections of* $\pi : E \to M$ *such that* $\left(\int_M |\sigma|^p \, vol_g \right)^{1/p} < \infty$.

With the norm $\|\sigma\|_p$ this space is a Banach space. For the case $p = 2$ we have the obvious Hilbert space inner product on $L^2_g(M, E)$ defined by

$$(\sigma, \eta) := \int_M \langle \sigma, \eta \rangle vol_g$$

Many, in fact, most of the standard facts about L^p spaces of functions on a measure space still hold in this context. For example, if $\sigma \in L^p(M, E), \eta \in L^q(M, E), \frac{1}{p} + \frac{1}{q} = 1, p, q \geq 1$ then $|\sigma| \, |\eta| \in L^1(M)$ and Hölder's inequality holds:

$$\int_M |\sigma| \, |\eta| \, vol_g \leq \left(\int_M |\sigma|^p \, vol_g \right)^{1/p} \left(\int_M |\eta|^q \, vol_g \right)^{1/q}$$

To what extent do the spaces $L^p_g(M, E)$ depend on the metric? If M is compact, it is easy to see that for any two metrics g_1 and g_2 there is a constant $C > 0$ such that

$$\frac{1}{C} \left(\int_M |\sigma|^p \, vol_{g_2} \right)^{1/p} \leq \left(\int_M |\sigma|^p \, vol_{g_1} \right)^{1/p} \leq C \left(\int_M |\sigma|^p \, vol_{g_2} \right)^{1/p}$$

uniformly for all $\sigma \in L^p_{g_1}(M, E)$. Thus $L^p_{g_1}(M, E) = L^p_{g_2}(M, E)$ and the norms are equivalent. For this reason we shall forego the subscript which references the metric.

Now we add one more piece of structure into the mix.

26.1.2 Distributions

For every integer $m \geq 0$ and compact subset $K \subset M$ we define (semi) norm $p_{K,m}$ on the space $\mathfrak{X}^k_l(M)$ by

$$p_{K,m}(\tau) := \sum_{j \leq m} \sup_{x \in K} \{ \left| \nabla^{(j)} \tau \right| (x) \}$$

Let $\mathfrak{X}_l^k(K)$ denote the set of restrictions of elements of $\mathfrak{X}_l^k(M)$ to K. The set $\mathfrak{X}_l^k(K)$ is clearly a vector space and $\{p_{K,m}\}_{1\leq m<\infty}$ is a family of norms that turns $\mathfrak{X}_l^k(K)$ into a Frechet space. Now let $\mathcal{D}(M,\mathfrak{X}_l^k)$ denote the space of (k,l)-tensor fields with compact support $\mathfrak{X}_l^k(M)_c \subset \mathfrak{X}_l^k(M)$ but equipped with the inductive limit topology of the family of Frechet spaces $\{\mathfrak{X}_l^k(K) : K \subset M \text{ compact}\}$. What we need to know is what this means in practical terms. Namely, we need a criterion for the convergence of a sequence (or net) of elements from $\mathfrak{X}_l^k(M)_c$.

Criterion 26.1 *Let $\{\tau_\alpha\} \subset \mathcal{D}(M,\mathfrak{X}_l^k) = \mathfrak{X}_l^k(M)_c$. Then $\tau_\alpha \to \tau$ if and only if given any $\epsilon > 0$ there is a compact set K_ϵ and $N > 0$ such that $\operatorname{supp}\tau_\alpha \subset K$ and such that $p_{K,m}(\tau_\alpha) < \epsilon$ whenever $\alpha > N$.*

Now the space of generalized tensors or tensor distributions $\mathcal{D}'(M,\mathfrak{X}_l^k)$ of type (k,l) is the set of all linear functionals on the space $\mathcal{D}(M,\mathfrak{X}_l^k)$ which are continuous in the following sense:

Criterion 26.2 *A linear functional $F : \mathcal{D}(M,\mathfrak{X}_l^k) \to \mathbb{C}$ is continuous if and only if for every compact set $K \subset M$ there are constants $C(K)$ and $m(K)$ such that*

$$|\langle F,\tau\rangle| \leq C(K)p_{K,m}(\tau) \text{ for all } \tau \in \mathfrak{X}_l^k(K)$$

Definition 26.2 *For each integer $p \geq 1$ let $\|\tau\|_p := \left(\int_M |\tau|^p \, dV\right)^{1/p}$. Let $L^p\mathfrak{X}_l^k(M)$ denote the completion of $\mathfrak{X}_l^k(M)$ with respect to this norm. Similarly, for every compact subset $K \subset M$ define $\|\tau\|_{p,K} := \left(\int_K |\tau|^p \, dV\right)^{1/p}$.*

The family of norms $\{\|.\|_{p,K}\}$, where K runs over all compact K, provides $\mathfrak{X}_l^k(M)$ with a Frechet space structure which we denote by $\mathfrak{X}_l^k(M)$

Definition 26.3 *For each integer $p \geq 1$ and integer $r \geq 0$ let $\|T\|_{m,p} := \sum_{|r|\leq m} \|\nabla^r T\|_p$. This defines a norm on $\mathfrak{X}_l^k(M)$ called the Sobolev norm. Let $W_{r,p}\mathfrak{X}_l^k(M)$ denote the completion of $\mathfrak{X}_l^k(M)$ with respect to this norm. For $k,l = 0,0$ so that we are dealing with functions we write $W_{r,p}(M)$ instead of $W_{r,p}\mathfrak{X}_0^0(M)$.*

26.1.3 Elliptic Regularity

26.1.4 Star Operator II

The definitions and basic algebraic results concerning the star operator on a scalar product space globalize to the tangent bundle of a Riemannian manifold in a straightforward way.

Definition 26.4 *Let M, g be a semi-Riemannian manifold. Each tangent space is a scalar product space and so on each tangent space T_pM we have a metric volume element vol_p and then the map $p \mapsto \mathrm{vol}_p$ gives a section of $\bigwedge^n T^*M$*

called the metric volume element of M, g. *Also on each fiber* $\bigwedge T_p^*M$ *of* $\bigwedge T^*M$ *we have a star operator* $*_p : \bigwedge^k T_p^*M \to \bigwedge^{n-k} T_p^*M$. *These induce a bundle map* $* : \bigwedge^k T^*M \to \bigwedge^{n-k} T^*M$ *and thus a map on sections (i.e. smooth forms)* $* : \Omega^k(M) \to \Omega^{n-k}(M)$.

Definition 26.5 *The star operator is sometimes referred to as the* ***Hodge star operator****.*

Definition 26.6 *Globalizing the scalar product on the Grassmann algebra we get a* ***scalar product bundle*** $\Omega(M), \langle .,. \rangle$ *where for every* $\eta, \omega \in \Omega^k(M)$ *we have a smooth function* $\langle \eta, \omega \rangle$ *defined by*

$$p \mapsto \langle \eta(p), \omega(p) \rangle$$

and thus a $C^\infty(M)$*-bilinear map* $\langle .,. \rangle : \Omega^k(M) \times \Omega^k(M) \to C^\infty(M)$. *Declaring forms of differing degree to be orthogonal as before we extend to a* C^∞ *bilinear map* $\langle .,. \rangle : \Omega(M) \times \Omega(M) \to C^\infty(M)$.

Theorem 26.1 *For any forms* $\eta, \omega \in \Omega(M)$ *we have* $\langle \eta, \omega \rangle \operatorname{vol} = \eta \wedge *\omega$

Now let M, g be a Riemannian manifold so that $\mathrm{g} = \langle .,. \rangle$ is positive definite. We can then define a Hilbert space of square integrable differential forms:

Definition 26.7 *Let an inner product be defined on* $\Omega_c(M)$, *the elements of* $\Omega(M)$ *with compact support, by*

$$(\eta, \omega) := \int_M \eta \wedge *\omega = \int_M \langle \eta, \omega \rangle \operatorname{vol}$$

and let $L^2(\Omega(M))$ *denote the* L^2 *completion of* $\Omega_c(M)$ *with respect to this inner product.*

26.2 The Laplace Operator

The exterior derivative operator $d : \Omega^k(M) \to \Omega^{k+1}(M)$ has a formal adjoint $\delta : \Omega^{k+1}(M) \to \Omega^k(M)$ defined by the requirement that for all $\alpha, \beta \in \Omega_c^k(M)$ with compact support we have

$$(d\alpha, \beta) = (\alpha, \delta\beta).$$

On a Riemannian manifold M the Laplacian of a function $f \in C(M)$ is given in coordinates by

$$\Delta f = -\frac{1}{\sqrt{g}} \sum_{j,k} \partial_j (g^{jk} \sqrt{g} \partial_k f)$$

where g^{ij} is the inverse of g_{ij} the metric tensor and g is the determinant of the matrix $G = (g_{ij})$. We can obtain a coordinate free definition as follows. First we recall that the divergence of a vector field $X \in \mathfrak{X}(M)$ is given at $p \in M$ by the trace of the map $\nabla X|_{T_pM}$. Here $\nabla X|_{T_pM}$ is the map

$$v \mapsto \nabla_v X.$$

Thus

$$\operatorname{div}(X)(p) := tr(\nabla X|_{T_pM}).$$

Then we have

$$\Delta f := \operatorname{div}(\operatorname{grad}(f))$$

Eigenvalue problem: For a given compact Riemannian manifold M one is interested in finding all $\lambda \in R$ such that there exists a function $f \neq 0$ in specified subspace $S \subset L^2(M)$ satisfying $\Delta f = \lambda f$ together with certain boundary conditions in case $\partial M \neq 0$.

The reader may be a little annoyed that we have not specified S more clearly. The reason for this is twofold. First, the theory will work even for relatively compact open submanifolds with rather unruly topological boundary and so regularity at the boundary becomes and issue. In general, our choice of S will be influenced by boundary conditions. Second, even though it may appear that S must consist of C^2 functions, we may also seek "weak solutions" by extending Δ in some way. In fact, Δ is essentially self adjoint in the sense that it has a unique extension to a self adjoint unbounded operator in $L^2(M)$ and so eigenvalue problems could be formulated in this functional analytic setting. It turns out that under very general conditions on the form of the boundary conditions, the solutions in this more general setting turn out to be smooth functions. This is the result of the general theory of elliptic regularity.

Definition 26.8 *A boundary operator is a linear map $b : S \to C^0(\partial M)$.*

Using this notion of a boundary operator we can specify boundary conditions as the requirement that the solutions lie in the kernel of the boundary map. In fact, the whole eigenvalue problem can be formulated as the search for λ such that the linear map

$$(\triangle - \lambda) \oplus b : S \to L^2(M) \oplus C^0(\partial M)$$

has a nontrivial kernel. If we find such a λ then this kernel is denoted $E_\lambda \subset L^2(M)$ and by definition $\Delta f = \lambda f$ and $bf = 0$ for all $f \in E_\lambda$. Such a function is called an **eigenfunction** corresponding to the eigenvalue λ. We shall see below that in each case of interest (for compact M) the eigenspaces E_λ will be finite dimensional and the eigenvalues form a sequence of nonnegative numbers increasing without bound. The dimension $\dim(E_\lambda)$ is called the **multiplicity** of λ. We shall present the sequence of eigenvalues in two ways:

1. If we write the sequence so as to include repetitions according to multiplicity then the eigenvalues are written as $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \uparrow \infty$. Thus it is possible, for example, that we might have $\lambda_2 = \lambda_3 = \lambda_4$ if $\dim(E_{\lambda_2}) = 3$.

2. If we wish to list the eigenvalues without repetition then we use an overbar:

$$0 \leq \bar{\lambda}_1 < \bar{\lambda}_2 < \ldots \uparrow \infty$$

The sequence of eigenvalues is sometimes called the **spectrum** of M.

To make thing more precise we divide things up into four cases:

The closed eigenvalue problem: In this case M is a compact Riemannian manifold without boundary the specified subspace of $L^2(M)$ can be taken to be $C^2(M)$. The kernel of the map $\Delta - \lambda : C^2(M) \to C^0(M)$ is the λ eigenspace and denoted by E_λ It consists of eigenfunctions for the eigenvalue λ.

The Dirichlet eigenvalue problem: In this case M is a compact Riemannian manifold without nonempty boundary ∂M. Let $\overset{\circ}{M}$ denote the interior of M. The specified subspace of $L^2(M)$ can be taken to be $C^2(\overset{\circ}{M}) \cap C^0(M)$ and the boundary conditions are $f|\,\partial M \equiv 0$ (**Dirichlet boundary conditions)** so the appropriate boundary operator is the restriction map $b_D : f \longmapsto f|\,\partial M$. The solutions are called **Dirichlet eigenfunctions** and the corresponding sequence of numbers λ for which a nontrivial solution exists is called the **Dirichlet spectrum** of M**.**

The Neumann eigenvalue problem: In this case M is a compact Riemannian manifold without nonempty boundary ∂M but . The specified subspace of $L^2(M)$ can be taken to be $C^2(\overset{\circ}{M}) \cap C^1(M)$. The problem is to find nontrivial solutions of $\Delta f = \lambda f$ with $f \in C^2(\overset{\circ}{M}) \cap C^0(\partial M)$ that satisfy $\nu f|\,\partial M \equiv 0$ (**Neumann boundary conditions)**.Thus the boundary map here is $b_N : C^1(M) \to C^0(\partial M)$ given by $f \mapsto \nu f|\,\partial M$ where ν is a smooth unit normal vector field defined on ∂M and so the νf is the normal derivative of f . The solutions are called **Neumann eigenfunctions** and the corresponding sequence of numbers λ for which a nontrivial solution exists is called the **Neumann spectrum** of M**.**

Recall that the completion of $C^k(M)$ (for any $k \geq 0$) with respect to the inner product

$$(f, g) = \int_M f g dV$$

is the Hilbert space $L^2(M)$. The Laplace operator has a natural extension to a self adjoint operator on $L^2(M)$ and a careful reformulation of the above eigenvalue problems in this Hilbert space setting together with the theory of elliptic regularity lead to the following

Theorem 26.2 *1) For each of the above eigenvalue problems the set of eigenvalues (the spectrum) is a sequence of nonnegative numbers which increases without bound:* $0 \leq \bar{\lambda}_1 < \bar{\lambda}_2 < \cdots \uparrow \infty$.

2) Each eigenfunction is a C^∞ *function on* $M = \overset{\circ}{M} \cup \partial M$.

3) Each eigenspace $E_{\bar{\lambda}_i}$ *(or* $E^D_{\bar{\lambda}_i}$ *or* $E^N_{\bar{\lambda}_i}$*) is finite dimensional, that is, each eigenvalue has finite multiplicity.*

4) If $\varphi_{\bar{\lambda}_i,1},...,\varphi_{\bar{\lambda}_i,m_i}$ is an orthonormal basis for the eigenspace $E_{\bar{\lambda}_i}$ (or $E^D_{\bar{\lambda}_i}$ or $E^N_{\bar{\lambda}_i}$) then the set $B = \cup_i\{\varphi_{\bar{\lambda}_i,1},...,\varphi_{\bar{\lambda}_i,m_i}\}$ is a complete orthonormal set for $L^2(M)$. In particular, if we write the spectrum with repetitions by multiplicity, $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \uparrow \infty$, then we can reindex this set of functions B as $\{\varphi_1, \varphi_2, \varphi_3, ...\}$ to obtain an ordered orthonormal basis for $L^2(M)$ such that φ_i is an eigenfunction for the eigenvalue λ_i.

The above can be given the following physical interpretation. If we think of M as a vibrating homogeneous membrane then the transverse motion of the membrane is described by a function $f : M \times (0, \infty) \to R$ satisfying

$$\Delta f + \frac{\partial^2 f}{\partial t^2} = 0$$

and if $\partial M \neq \emptyset$ then we could require $f|\partial M \times (0, \infty) = 0$ which means that we are holding the boundary fixed. A similar discussion for the Neumann boundary conditions is also possible and in this case the membrane is free at the boundary. If we look for the solutions of the form $f(x,t) = \phi(x)T(t)$ then we are led to conclude that ϕ must satisfy $\Delta\phi = \lambda\phi$ for some real number λ with $\phi = 0$ on ∂M. This is the Dirichlet eigenvalue problem discussed above.

Theorem 26.3 *For each of the eigenvalue problems defined above*

Now explicit solutions of the above eigenvalue problems are very difficult to obtain except in the simplest of cases. It is interesting therefore, to see if one can tell something about the eigenvalues from the geometry of the manifold. For instance we may be interested in finding upper and/or lower bounds on the eigenvalues of the manifold in terms of various geometric attributes of the manifold. A famous example of this is the Faber–Krahn inequality which states that if Ω is a regular domain in say $\mathbb{R}^n$ and D is a ball or disk of the same volume then

$$\lambda(\Omega) \geq \lambda(D)$$

where $\lambda(\Omega)$ and $\lambda(D)$ are the lowest nonzero Dirichlet eigenvalues of Ω and D respectively. Now it is of interest to ask whether one can obtain geometric information about the manifold given a degree of knowledge about the eigenvalues. There is a 1966 paper by M. Kac entitled "Can One Hear the Shape of a Drum?" which addresses this question. Kac points out that Weyl's asymptotic formula shows that the sequence of eigenvalues does in fact determine the volume of the manifold. Weyl's formula is

$$(\lambda_k)^{n/2} \sim \left(\frac{(2\pi)^n}{\omega_n}\right)\frac{k}{vol(M)} \quad \text{as} \quad k \longrightarrow \infty$$

where ω_n is the volume of the unit ball in $\mathbb{R}^n$ and M is the given compact manifold. In particular,

$$\left(\frac{(2\pi)^n}{\omega_n}\right)\lim_{k\to\infty}\frac{k}{(\lambda_k)^{n/2}} = vol(M).$$

So the volume is indeed determined by the spectrum[1].

26.3 Spectral Geometry

Legend has it that Pythagoras was near a black-smith's shop one day and hearing the various tones of a hammer hitting an anvil was lead to ponder the connection between the geometry (and material composition) of vibrating objects and the pitches of the emitted tones. This lead him to experiment with vibrating strings and a good deal of mathematics ensued. Now given a string of uniform composition it is essentially the length of the string that determines the possible pitches. Of course, there isn't much Riemannian geometry in a string because the dimension is 1. Now we have seen that a natural mathematical setting for vibration in higher dimensions is the Riemannian manifold and the wave equation associated with the Laplacian. The spectrum of the Laplacian corresponds the possible frequencies of vibration and it is clearly only the metric together with the total topological structure of the manifold that determines the spectrum. If the manifold is a Lie group or is a homogeneous space acted on by a Lie group, then the topic becomes highly algebraic but simultaneously involves fairly heavy analysis. This is the topic of harmonic analysis and is closely connected with the study of group representations. One the other hand, the Laplacian and its eigenvalue spectrum are defined for arbitrary (compact) Riemannian manifolds and, generically, a Riemannian manifold is far from being a Lie group or homogeneous space. The isometry group may well be trivial. Still the geometry must determine the spectrum. But what is the exact relationship between the geometry of the manifold and the spectrum? Does the spectrum determine the geometry. Is it possible that two manifolds can have the same spectrum without being isometric to each other? That the answer is yes has been known for quite a while now it wasn't until (??whenn) that the question was answered for planar domains? This was Mark Kac's original question: "Can one hear the shape of a drum?" It was shown by Carolyn Gordon and (??WhoW?) that the following two domains have the same Dirichlet spectrum but are not isometric:

finsh>

[1]Notice however, one may ask still how far out into the spectrum must one "listen" in order to gain an estimate of $vol(M)$ to a given accuracy.

26.4 Hodge Theory

26.5 Dirac Operator

It is often convenient to consider the differential operator $D = i\frac{\partial}{\partial x}$ instead of $\frac{\partial}{\partial x}$ even when one is interested mainly in real valued functions. For one thing $D^2 = -\frac{\partial^2}{\partial x^2}$ and so D provides a sort of square root of the positive Euclidean Laplacian $\triangle = -\frac{\partial^2}{\partial x^2}$ in dimension 1. Dirac wanted a similar square root for the wave operator $\square = \partial_0^2 - \sum_{i=1}^3 \partial_i^2$ (the Laplacian in $\mathbb{R}^4$ for the Minkowski inner metric) and found that an operator of the form $D = \partial_0 - \sum_{i=1}^3 \gamma_i \partial_i$ would do the job if it could be arranged that $\gamma_i\gamma_j + \gamma_i\gamma_j = 2\eta_{ij}$ where

$$(\eta_{ij}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

One way to do this is to allow the γ_i to be matrices.

Now lets consider finding a square root for $\triangle = -\sum_{i=1}^n \partial_i^2$. We accomplish this by an $\mathbb{R}$-linear embedding of $\mathbb{R}^n$ into an $N \times N$ real or complex matrix algebra A by using n linearly independent matrices $\{\gamma_i : i = 1, 2, ..., n\}$ (so called "gamma matrices") and mapping

$$(x^1, ..., x^n) \mapsto x^i\gamma_i \text{ (sum)}.$$

and where $\gamma_1, ..., \gamma_n$ are matrices satisfying the basic condition

$$\gamma_i\gamma_j + \gamma_i\gamma_j = -2\delta_{ij}.$$

We will be able to arrange[2] that $\{1, \gamma_1, ..., \gamma_n\}$ generates an algebra of dimension 2^n spanned as vector space by the identity matrix 1 and all products of the form $\gamma_{i_1} \cdots \gamma_{i_k}$ with $i_1 < i_2 < \cdots < i_k$. Thus we aim to identify $\mathbb{R}^n$ with the linear span of these gamma matrices. Now if we can find matrices with the property that $\gamma_i\gamma_j + \gamma_i\gamma_j = -2\delta_{ij}$ then our "Dirac operator" will be

$$D = \sum_{i=1}^n \gamma_i \partial_i$$

which is now acting on N-tuples of smooth functions.

Now the question arises: What are the differential operators $\partial_i = \frac{\partial}{\partial x^i}$ acting on exactly. The answer is that they act on whatever we take the algebra spanned by the gamma matrices to be acting on. In other words we should have some vector space S that is a module over the algebra spanned by the gamma matrices.

[2]It is possible that gamma matrices might span a space of half the dimension we are interested in. This fact has gone unnoticed in some of the literature. The dimension condition is to assure that we get a universal Clifford algebra.

Then we take as our "fields" smooth maps $f : \mathbb{R}^n \to S$. Of course since the $\gamma_i \in \mathbb{M}_{N\times N}$ we may always take $S = \mathbb{R}^N$ with the usual action of $\mathbb{M}_{N\times N}$ on $\mathbb{R}^N$. The next example shows that there are other possibilities.

Example 26.1 *Notice that with* $\frac{\partial}{\partial z} := \frac{\partial}{\partial x} - i\frac{\partial}{\partial y}$ *and* $\frac{\partial}{\partial \bar{z}} := \frac{\partial}{\partial x} + i\frac{\partial}{\partial y}$ *we have*

$$\begin{aligned} &\begin{bmatrix} 0 & -\frac{\partial}{\partial z} \\ \frac{\partial}{\partial \bar{z}} & 0 \end{bmatrix}\begin{bmatrix} 0 & -\frac{\partial}{\partial z} \\ \frac{\partial}{\partial \bar{z}} & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & -\frac{\partial}{\partial z}\frac{\partial}{\partial \bar{z}} \\ -\frac{\partial}{\partial \bar{z}}\frac{\partial}{\partial z} & 0 \end{bmatrix} = \begin{bmatrix} \triangle & 0 \\ 0 & \triangle \end{bmatrix} \\ &= \triangle 1 \end{aligned}$$

where $\triangle = -\sum \partial_i^2$. *On the other hand*

$$\begin{bmatrix} 0 & -\frac{\partial}{\partial z} \\ \frac{\partial}{\partial \bar{z}} & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\frac{\partial}{\partial x} + \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}\frac{\partial}{\partial y}.$$

¿From this we can see that appropriate gamma matrices for this case are $\gamma_1 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$ *and* $\gamma_2 = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$.

Now let E^0 *be the span of* $1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ *and* $\gamma_2\gamma_1 = \begin{pmatrix} -i & 0 \\ 0 & -i \end{pmatrix}$. *Let* E^1 *be the span of* γ_1 *and* γ_2. *Refer to* E^0 *and* E^1 *the even and odd parts of* $Span\{1, \gamma_1, \gamma_2, \gamma_2\gamma_1\}$. *Then we have that* $D = \begin{bmatrix} 0 & -\frac{\partial}{\partial z} \\ \frac{\partial}{\partial \bar{z}} & 0 \end{bmatrix}$ *maps* E^0 *to* E^1 *and writing a typical element of* E^0 *as* $f(x, y) = u(x, y) + \gamma_2\gamma_1 v(x, y)$ *is easy to show that* $Df = 0$ *is equivalent to the Cauchy-Riemann equations.*

The reader should keep this last example in mind as this kind of decomposition into even and odd part will be a general phenomenon below.

26.5.1 Clifford Algebras

A Clifford algebra is the type of algebraic object that allows us to find differential operators that square to give Laplace type operators. The matrix approach described above is in fact quite general but there are other approaches that are more abstract and encourage one to think about a Clifford algebra as something that contains the scalar and the vector space we start with. The idea is similar to that of the complex numbers. We seldom think about complex numbers as "pairs of real numbers" while we are calculating unless push comes to shove. After all, there are other good ways to represent complex numbers; as matrices for example. And yet there is one underlying abstract object called the complex numbers which is quite concrete once one get used to using them. Similarly we encourage the reader to learn to think about abstract Clifford algebras in the same way. Just compute!

Clifford algebras are usually introduced in connection with a quadratic form q on some vector space but in fact we are just as interested in the associated symmetric bilinear form and so in this section we will generically use the same symbol for a quadratic form and the bilinear form obtained by polarization and write both $q(v)$ and $q(v, w)$.

Definition 26.9 *Let* V *be an* n *dimensional vector space over a field* $\mathbb{K}$ *with characteristic not equal to* 2*. Suppose that* q *is a quadratic form on* V *and let* q *be the associated symmetric bilinear form obtained by polarization. A* ***Clifford algebra based on*** V,q *is an algebra with unity* 1 $Cl(\mathrm{V},q,\mathbb{K})$ *containing* V *(or an isomorphic image of* V*) such that the following relations hold:*

$$vw + wv = -2q(v, w)1$$

and such that $Cl(\mathrm{V},q,\mathbb{K})$ *is universal in the following sense: Given any linear map* $L : \mathrm{V} \to A$ *into an associative* $\mathbb{K}-$*algebra with unity* $\mathbf{1}$ *such that*

$$L(v)L(w) + L(w)L(v) = -2q(v, w)\mathbf{1}$$

then there is a unique extension of L *to an algebra homomorphism* $\bar{L} : Cl(\mathrm{V},q,\mathbb{K}) \to A$*.*

If $e_1, ..., e_n$ is an orthonormal basis for V, q then we must have

$$\begin{aligned} e_i e_j + e_j e_i &= 0 \text{ for } i \neq j \\ e_i^2 &= -q(e_i) = \pm 1 \text{ or } 0 \end{aligned}$$

A common choice is the case when q is a nondegenerate inner product on a real vector space. In this case we have a particular realization of the Clifford algebra obtained by introducing a new product into the Grassmann vector space $\wedge\mathrm{V}$. The said product is the unique linear extension of the following rule for $v \in \wedge^1\mathrm{V}$ and $w \in \wedge^k\mathrm{V}$:

$$\begin{aligned} v \cdot w &:= v \wedge w - v^\flat \lrcorner w \\ w \cdot v &:= (-1)^k (v \wedge w + v^\flat \lrcorner w) \end{aligned}$$

We will refer to this as a **geometric algebra** on $\wedge\mathrm{V}$ and this version of the Clifford algebra will be called the **form presentation of** $Cl(\mathrm{V},q)$. Now once we have a definite inner product on V we have an inner product on V^* and $\mathrm{V} \cong \mathrm{V}^*$. The Clifford algebra on V^* is generated by the following more natural looking formulas

$$\begin{aligned} \alpha \cdot \beta &:= \alpha \wedge \beta - (\sharp\alpha) \lrcorner \beta \\ \beta \cdot \alpha &:= (-1)^k (\alpha \wedge \beta + (\sharp\alpha) \lrcorner \beta) \end{aligned}$$

for $\alpha \in \wedge^1\mathrm{V}$ and $\beta \in \wedge\mathrm{V}$.

Now we have seen that one can turn $\wedge\mathrm{V}$ (or $\wedge\mathrm{V}^*$) into a Clifford algebra and we have also seen that one can obtain a Clifford algebra whenever appropriate gamma matrices can be found. A slightly more abstract construction is also common: Denote by $I(q)$ the ideal of the full tensor algebra $T(\mathrm{V})$ generated by elements of the form $x \otimes x - q(x) \cdot 1$. The **Clifford algebra** is (up to isomorphism) given by

$$Cl(\mathrm{V}, q, \mathbb{K}) = T(\mathrm{V})/I(q).$$

We can use the canonical injection

$$i : \mathrm{V} \longrightarrow C_K$$

to identify V with its image in $Cl(\mathrm{V}, q, \mathbb{K})$. (The map turns out that i is 1–1 onto $i(\mathrm{V})$ and we will just accept this without proof.)

Exercise 26.1 *Use the universal property of $Cl(\mathrm{V}, q, \mathbb{K})$ to show that it is unique up to isomorphism.*

Remark 26.1 *Because of the form realization of a Clifford algebra we see that $\wedge\mathrm{V}$ is a $Cl(\mathrm{V}, q, \mathbb{R})-$module. But even if we just have some abstract $Cl(\mathrm{V}, q, \mathbb{R})$ we can use the universal property to extend the action of* V *on $\wedge\mathrm{V}$ given by*

$$v \mapsto v \cdot w := v \wedge w - v^\flat \lrcorner w$$

to an action of $Cl(\mathrm{V}, q, \mathbb{K})$ on $\wedge\mathrm{V}$ thus making $\wedge\mathrm{V}$ a $Cl(\mathrm{V}, q, \mathbb{R})-$module.

Definition 26.10 *Let $\mathbb{R}^n_{(r,s)}$ be the real vector space $\mathbb{R}^n$ with the inner product of signature (r, s) given by*

$$\langle x, y \rangle := \sum_{i=1}^{r} x_i y_i - \sum_{i=r+1}^{r+s=n} x_i y_i.$$

The Clifford algebra formed from this inner product space is denoted $Cl_{r,s}$. In the special case of $(p, q) = (n, 0)$ we write Cl_n.

Definition 26.11 *Let $\mathbb{C}^n$ be the complex vector space of n-tuples of complex numbers together with the standard symmetric $\mathbb{C}-$bilinear form*

$$b(z, w) := \sum_{i=1}^{n} z_i w_i.$$

The (complex) Clifford algebra obtained is denoted $\mathbb{C}l_n$.

Remark 26.2 *The complex Clifford algebra $\mathbb{C}l_n$ is based on a complex symmetric form and* not *on a Hermitian form.*

Exercise 26.2 *Show that for any nonnegative integers p, q with $p + q = n$ we have $Cl_{p,q} \otimes \mathbb{C} \cong \mathbb{C}l_n$.*

Example 26.2 *The Clifford algebra based on $\mathbb{R}^1$ itself with the relation $x^2 = -1$ is just the complex number system.*

The Clifford algebra construction can be globalized in the obvious way. In particular, we have the option of using the form presentation so that the above formulas $\alpha \cdot \beta := \alpha \wedge \beta - (\sharp\alpha) \lrcorner \beta$ and $\beta \cdot \alpha := (-1)^k(\alpha \wedge \beta + (\sharp\alpha) \lrcorner \beta)$ are interpreted as equations for differential forms $\alpha \in \wedge^1 T^*M$ and $\beta \in \wedge^k T^*M$ on a semi-Riemannian manifold M, g. In any case we have the following

Definition 26.12 *Given a Riemannian manifold M, g, the Clifford algebra bundle is $Cl(T^*M, \mathrm{g}) = Cl(T^*M) := \cup_x Cl(T_x^*M)$.*

Since we take each tangent space to be embedded $T_x^*M \subset Cl(T_x^*M)$, the elements θ^i of a local orthonormal frame $\theta^1, ..., \theta^n \in \Omega^1$ are also local sections of $Cl(T^*M, \mathrm{g})$ and satisfy

$$\theta^i\theta^j + \theta^j\theta^i = -\langle\theta^i, \theta^j\rangle = -\epsilon^i\delta^{ij}$$

Recall that $\varepsilon^1, ..., \varepsilon^n$ is a list of numbers equal to ± 1 (or even 0 if we allow degeneracy) and giving the index of the metric $\mathrm{g}(.,.) = \langle .,. \rangle$.

Obviously, we could also work with the bundle $Cl(TM) := \cup_x Cl(T_xM)$ which is naturally isomorphic to $Cl(T^*M)$ in which case we would have

$$e^ie^j + e^je^i = -\langle e^i, e^j\rangle = -\varepsilon^i\delta^{ij}$$

for orthonormal frames. Of course it shouldn't make any difference to our development since one can just identify TM with T^*M by using the metric. On the other hand, we could define $Cl(T^*M, \mathrm{b})$ even if b is a *degenerate* bilinear tensor and then we recover the Grassmann algebra bundle $\wedge T^*M$ in case $\mathrm{b} \equiv 0$. These comments should make it clear that $Cl(T^*M, \mathrm{g})$ is in general a sort of deformation of the Grassmann algebra bundle.

There are a couple of things to notice about $Cl(T^*M)$ when we realize it as $\wedge T^*M$ with a new product. First of all if $\alpha, \beta \in \wedge T^*M$ and $\langle\alpha, \beta\rangle = 0$ then $\alpha \cdot \beta = \alpha \wedge \beta$ where as if $\langle\alpha, \beta\rangle \neq 0$ then in general $\alpha\beta$ is not a homogeneous element. Second, $Cl(T^*M)$ is locally generated by $\{1\} \cup \{\theta^i\} \cup \{\theta^i\theta^j : i < j\} \cup \cdots \cup \{\theta^1\theta^2 \cdots \theta^n\}$ where $\theta^1, \theta^2, ..., \theta^n$ is a local orthonormal frame. Now we can immediately define our current objects of interest:

Definition 26.13 *A* ***bundle of modules over*** $Cl(T^*M)$ *is a vector bundle $\Sigma = (E, \pi, M)$ such that each fiber E_x is a module over the algebra $Cl(T_x^*M)$ and such that for each $\theta \in \Gamma(Cl(T^*M))$ and each $\sigma \in \Gamma(\Sigma)$ the map $x \mapsto \theta(x)\sigma(x)$ is smooth. Thus we have an induced map on smooth sections: $\Gamma(Cl(T^*M)) \times \Gamma(\Sigma) \to \Gamma(\Sigma)$.*

Proposition 26.1 *The bundle $Cl(T^*M)$ is a Clifford module over itself and the Levi-Civita connection ∇ on M induces a connection on $Cl(T^*M)$ (this connection is also denoted ∇) such that*

$$\nabla(\sigma_1\sigma_2) = (\nabla\sigma_1)\sigma_2 + \sigma_1\nabla\sigma_2$$

for all $\sigma_1, \sigma_2 \in \Gamma(Cl(T^*M))$. In particular, if $X, Y \in \mathfrak{X}(M)$ and $\sigma \in \Gamma(Cl(T^*M))$ then

$$\nabla_X(Y\sigma) = (\nabla_X Y)\sigma + Y\nabla_X\sigma.$$

Proof. Realize $Cl(T^*M)$ as $\wedge T^*M$ with Clifford multiplication and let ∇ be usual induced connection on $\wedge T^*M \subset \otimes T^*M$. We have for an local orthonormal frame $e_1, ..., e_n$ with dual frame $\theta^1, \theta^2, ..., \theta^n$. Then $\nabla_\xi \theta^i = -\Gamma^i_j(\xi)\theta^j$

$$\begin{aligned}
\nabla_\xi(\theta^i\theta^j) &= \nabla_\xi(\theta^i \wedge \theta^j) \\
&= \nabla_\xi\theta^i \wedge \theta^j + \theta^i \wedge \nabla_\xi\theta^j \\
&= -\Gamma^i_k(\xi)\theta^k \wedge \theta^j - \Gamma^j_k(\xi)\theta^k \wedge \theta^i \\
&= -\Gamma^i_k(\xi)\theta^k\theta^j - \Gamma^j_k(\xi)\theta^k\theta^i = (\nabla_\xi\theta^i)\theta^j + \theta^i\nabla_\xi\theta^j
\end{aligned}$$

The result follows by linearity and a simple induction since a general section σ can be written locally as $\sigma = \sum a_{i_1 i_2 \ldots i_k}\theta^{i_1}\theta^{i_2}\cdots\theta^{i_k}$. ■

Definition 26.14 *Let M, g be a (semi-) Riemannian manifold. A compatible connection for a **bundle of modules** Σ **over** $Cl(T^*M)$ is a connection ∇^Σ on Σ such that*

$$\nabla^\Sigma(\sigma \cdot s) = (\nabla\sigma) \cdot s + \sigma \cdot \nabla^\Sigma s$$

for all $s \in \Gamma(\Sigma)$ and all $\sigma \in \Gamma(Cl(T^*M))$

Definition 26.15 *Let $\Sigma = (E, \pi, M)$ be a bundle of modules over $Cl(T^*M)$ with a compatible connection $\nabla = \nabla^\Sigma$. The associated **Dirac operator** is defined as a differential operator Σ on by*

$$Ds := \sum \theta^i \cdot \nabla^\Sigma_{e_i} s$$

for $s \in \Gamma(\Sigma)$.

Notice that Clifford multiplication of $Cl(T^*M)$ on $\Sigma = (E, \pi, M)$ is a zeroth order operator and so is well defined as a fiberwise operation $Cl(T^*_xM) \times E_x \to E_x$.

There are still a couple of convenient properties that we would like to have. These are captured in the next definition.

Definition 26.16 *Let $\Sigma = (E, \pi, M)$ be a bundle of modules over $Cl(T^*M)$ such that Σ carries a Riemannian metric and compatible connection $\nabla = \nabla^\Sigma$. We call $\Sigma = (E, \pi, M)$ a **Dirac bundle** if the following equivalent conditions hold:*

1) $\langle es_1, es_2\rangle = \langle s_1, s_2\rangle$ for all $s_1, s_2 \in E_x$ and all $e \in T^*_xM \subset Cl(T^*_xM))$ with $|e| = 1$. In other words, Clifford multiplication by a unit (co)vector is required to be an isometry of the Riemannian metric on each fiber of Σ. Since , $e^2 = -1$ it follows that this is equivalent to requiring .

2) $\langle es_1, s_2\rangle = -\langle s_1, es_2\rangle$ for all $s_1, s_2 \in E_x$ and all $e \in T^*_xM \subset Cl(T^*_xM))$ with $|e| = 1$.

Assume in the sequel that q is nondegenerate. Let denote the subalgebra generated by all elements of the form $x_1 \cdots x_k$ with k even. And similarly, $Cl_1(\mathrm{V},q)$, with k odd. Thus $Cl(\mathrm{V},q)$ has the structure of a Z_2-graded algebra :

$$Cl(\mathrm{V},q) = Cl_0(\mathrm{V},q) \oplus Cl_1(\mathrm{V},q)$$

$$Cl_0(\mathrm{V},q) \cdot Cl_0(\mathrm{V},q) \subset Cl_0(\mathrm{V},q)$$
$$Cl_0(\mathrm{V},q) \cdot Cl_1(\mathrm{V},q) \subset Cl_1(\mathrm{V},q)$$
$$Cl_1(\mathrm{V},q) \cdot Cl_1(\mathrm{V},q) \subset Cl_0(\mathrm{V},q)$$

$Cl_0(\mathrm{V},q)$ and $Cl_1(\mathrm{V},q)$ are referred to as the **even** and **odd** part respectively. A Z_2-graded algebra is also called a **superalgebra**. There exists a fundamental automorphism α of $Cl(\mathrm{V},q)$ such that $\alpha(x) = -x$ for all $x \in \mathrm{V}$. Note that $\alpha^2 = \mathrm{id}$. It is easy to see that $Cl_0(\mathrm{V},q)$ and $Cl_1(\mathrm{V},q)$ are the $+1$ and -1 eigenspaces of $\alpha : Cl(\mathrm{V},q) \to Cl(\mathrm{V},q)$.

26.5.2 The Clifford group and spinor group

Let G be the group of all invertible elements $s \in C_K$ such that $s\mathrm{V}s^{-1} = \mathrm{V}$. This is called the **Clifford group** associated to q. The special Clifford group is $G^+ = G \cap C_0$. Now for every $s \in G$ we have a map $\phi_s : v \longrightarrow svs^{-1}$ for $v \in \mathrm{V}$. It can be shown that ϕ is a map from G into $\mathrm{O}(q)$, the orthogonal group of q. The kernel is the invertible elements in the center of C_K.

It is a useful and important fact that if $x \in G \cap \mathrm{V}$ then $q(x) \neq 0$ and $-\phi_x$ is reflection through the hyperplane orthogonal to x. Also, if s is in G^+ then ϕ_s is in $S\mathrm{O}(q)$. In fact, $\phi(G^+) = S\mathrm{O}(q)$.

Besides the fundamental automorphism α mentioned above, there is also a fundament anti-automorphism or **reversion** $\beta : Cl(\mathrm{V},q) \to Cl(\mathrm{V},q)$ which is determined by the requirement that $\beta(v_1 v_2 \cdots v_k) = v_k v_{k-1} \cdots v_1$ for $v_1, v_2, ..., v_k \in \mathrm{V} \subset Cl(\mathrm{V},q)$. We can use this anti-automorphism β to put a kind of "norm" on G^+;

$$N; G^+ \longrightarrow \mathbb{K}^*$$

where $\mathbb{K}^*$ is the multiplicative group of nonzero elements of $\mathbb{K}$ and $N(s) = \beta(s)s$. This is a homomorphism and if we "mod out" the kernel of N we get the so called reduced Clifford group G_0^+.

We now specialize to the real case $\mathbb{K} = \mathbb{R}$. The identity component of G_0^+ is called the **spin group** and is denoted by $\mathrm{Spin}(\mathrm{V}, q)$.

26.6 The Structure of Clifford Algebras

Now if $\mathbb{K} = \mathbb{R}$ and
$$q(x) = \sum_{i=1}^{r} (x_i)^2 - \sum_{i=r+1}^{r+s} (x_i)^2$$
we write $C(\mathbb{R}^{r+s}, q, \mathbb{R}) = Cl(r, s)$. Then one can prove the following isomorphisms.
$$Cl(r+1, s+1) \cong Cl(1,1) \otimes C(r,s)$$
$$Cl(s+2, r) \cong Cl(2,0) \otimes Cl(r,s)$$
$$Cl(s, r+2) \cong Cl(0,2) \otimes Cl(r,s)$$

and

$$Cl(p,p) \cong \bigotimes^{p} Cl(1,1)$$

$$Cl(p+k,p) \cong \bigotimes^{p} Cl(1,1) \bigotimes Cl(k,0)$$

$$Cl(k,0) \cong Cl(2,0) \otimes Cl(0,2) \otimes Cl(k-4,0) \qquad k>4$$

Using the above type of periodicity relations together with

$$Cl(2,0) \cong Cl(1,1) \cong M_2(\mathbb{R})$$

$$Cl(1,0) \cong \mathbb{R} \oplus \mathbb{R}$$

and

$$Cl(0,1) \cong \mathbb{C}$$

we can piece together the structure of $Cl(r,s)$ in terms of familiar matrix algebras. We leave out the resulting table since for one thing we are more interested in the simpler complex case. Also, we will explore a different softer approach below.

The complex case . In the complex case we have a much simpler set of relations;

$$Cl(2r) \cong Cl(r,r) \otimes \mathbb{C} \cong M_{2^r}(\mathbb{C})$$

$$Cl(2r+1) \cong Cl(1) \otimes Cl(2r)$$

$$\cong Cl(2r) \oplus Cl(2r) \cong M_{2^r}(\mathbb{C}) \oplus M_{2^r}(\mathbb{C})$$

These relations remind us that we may use matrices to represent our Clifford algebras. Lets return to this approach and explore a bit.

26.6.1 Gamma Matrices

Definition 26.17 *A set of real or complex matrices $\gamma_1, \gamma_2, ..., \gamma_n$ are called gamma matrices for $Cl(r,s)$ if*

$$\gamma_i \gamma_j + \gamma_j \gamma_i = -2g_{ij}$$

where $(g_{ij}) = diag(1, ..., 1, -1, ...,)$ is the diagonalized matrix of signature (r,s).

Example 26.3 *Recall the Pauli matrices:*

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad , \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\sigma_2 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} \quad , \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

It is easy to see that σ_1, σ_2 serve as gamma matrices for $Cl(0,2)$ while $i\sigma_1, i\sigma_2$ serve as gamma matrices for $Cl(2,0)$.

$Cl(2,0)$ is spanned as a vector space of matrices by $\sigma_0, i\sigma_1, i\sigma_2, i\sigma_3$ and is (algebra) isomorphic to the quaternion algebra $\mathbb{H}$ under the identification

$$\begin{aligned}\sigma_0 &\mapsto 1\\ i\sigma_1 &\mapsto I\\ i\sigma_2 &\mapsto J\\ i\sigma_3 &\mapsto K\end{aligned}$$

Example 26.4 *The matrices $\sigma_0, \sigma_1, \sigma_2, \sigma_3$ are gamma matrices for $Cl(1,3)$.*

26.7 Clifford Algebra Structure and Representation

26.7.1 Bilinear Forms

We will need some basic facts about bilinear forms. We review this here.

(1) Let E be a module over a commutative ring R. Typically E is a vector space over a field $\mathbb{K}$. A bilinear map $g : \mathrm{E} \times \mathrm{E} \longrightarrow R$ is called symmetric if $g(x,y) = g(y,x)$ and antisymmetric if $g(x,y) = -g(y,x)$ for $(x,y) \in \mathrm{E} \times \mathrm{E}$. If R has an automorphism of order two, $a \mapsto \bar{a}$ we say that g is Hermitian if $g(ax,y) = ag(x,y)$ and $g(x,ay) = \bar{a}g(x,y)$ for all $a \in R$ and $(x,y) \in \mathrm{E} \times \mathrm{E}$. If g is any of symmetric,antisymmetric,or Hermitian then the " left kernel" of g is equal to the "right kernel". That is

$$\begin{aligned}\ker g &= \{x \in \mathrm{E} : g(x,y) = 0 \quad \forall y \in \mathrm{E}\}\\ &= \{y \in \mathrm{E} : g(x,y) = 0 \quad \forall x \in \mathrm{E}\}\end{aligned}$$

If $\ker g = 0$ we say that g is nondegenerate. In case E is a vector space of finite dimension g is nondegenerate if and only if $x \mapsto g(x,\cdot) \in \mathrm{E}^*$ is an isomorphism. An orthogonal basis for g is a basis $\{v_i\}$ for E such that $g(v_i, v_i) = 0$ for $i \neq j$.

Definition 26.18 *Let E be a vector space over a three types above. If $\mathrm{E} = \mathrm{E} \oplus \mathrm{E}_2$ for subspaces $\mathrm{E}_i \subset \mathrm{E}$ and $g(x_1, x_2) = 0 \qquad \forall x_1 \in \mathrm{E}, x_2 \in \mathrm{E}_2$ then we write*

$$\mathrm{E} = \mathrm{E}_1 \perp \mathrm{E}_2$$

and say that E is the orthogonal direct sum of E_1 and E_2.

Proposition 26.2 *Suppose E, g is as above with*

$$\mathrm{E} = \mathrm{E}_1 \perp \mathrm{E}_2 \perp \cdots \perp \mathrm{E}_k$$

Then g is non-degenerate if and only if its restrictions $g|_{\mathrm{E}_i}$ are and

$$\ker \mathrm{E} = \mathrm{E}_1^o \perp \mathrm{E}_2^o \perp \cdots \perp \mathrm{E}_k^o$$

Proof. Nearly obvious. ■

Terminology: If g is one of symmetric, antisymmetric or Hermitian we say that g is geometric.

Proposition 26.3 *Let g be a geometric bilinear form on a vector space* E *(over* $\mathbb{K}$*). Suppose g is nondegenerate. Then g is nondegenerate on a subspace F if and only if* $\mathrm{E} = F \perp F^{\perp}$ *where*

$$F^{\perp} = \{x \in \mathrm{E} : g(x, f) = 0 \qquad \forall f \in F\}$$

Definition 26.19 *A map q is called quadratic if there is a symmetric g such that $q(x) = g(x, x)$. Note that g can be recovered from q:*

$$2g(x, y) = q(x + y) - q(x) - q(y)$$

26.7.2 Hyperbolic Spaces And Witt Decomposition

E, g is a vector space with symmetric form g. If E has dimension 2 we call E a hyperbolic plane. If $\dim \mathrm{E} \geq 2$ and $\mathrm{E} = \mathrm{E}_1 \perp \mathrm{E}_2 \perp \cdots \perp \mathrm{E}_k$ where each E_i is a hyperbolic plane for $g|_{\mathrm{E}_i}$ then we call E a hyperbolic space. For a hyperbolic plane one can easily construct a basis f_1, f_2 such that $g(f_1, f) = g(f_2, f_2) = 0$ and $g(f_1, f_2) = 1$. So that with respect to this basis g is given by the matrix

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

This pair $\{f_1, f_2\}$ is called a hyperbolic pair for E, g. Now we return to $\dim\mathrm{E} \geq 2$. Let $\operatorname{rad} F \equiv F^{\perp} \cap F = \ker g|_F$

Lemma 26.1 *There exists a subspace* $U \subset \mathrm{E}$ *such that* $\mathrm{E} = \operatorname{rad} \mathrm{E} \perp U$ *and U is nondegenerate.*

Proof. It is not to hard to see that $\operatorname{rad} U = \operatorname{rad} U^{\perp}$. If $\operatorname{rad} U = 0$ then $\operatorname{rad} U^{\perp} = 0$ and visa versa. Now $U + U^{\perp}$ is clearly direct since $0 = \operatorname{rad} U = U \cap U^{\perp}$. Thus $\mathrm{E} = U \perp U^{\perp}$. ■

Lemma 26.2 *Let g be nondegenerate and* $U \subset \mathrm{E}$ *some subspace. Suppose that* $U = \operatorname{rad} U \perp W$ *where* $\operatorname{rad} W = 0$*. Then given a basis $\{u_1, \cdots, u_s\}$ for* $\operatorname{rad} U$ *there exists $v_1, \cdots, v_s \in W^{\perp}$ such that each $\{u_i, v_i\}$ is a hyperbolic pair. Let $P_i = span\{u_i, v_i\}$. Then*

$$\mathrm{E} = W \perp P_1 \perp \cdots \perp P_s.$$

Proof. Let $W_1 = \operatorname{span}\{u_2, u_3, \cdots, u_s\} \oplus W$. Then $W_1 \subsetneq \operatorname{rad} U \oplus W$ so $(\operatorname{rad} U \oplus W)^{\perp} \subsetneq W_1^{\perp}$. Let $w_1 \in W_1^{\perp}$ but assume $w_1 \notin (\operatorname{rad} U \oplus W)^{\perp}$. Then we have $g(u_1, w_1) \neq 0$ so that $P_1 = span\{u_1, w_1\}$ is a hyperbolic plane. Thus we can find v_1 such that u_1, v_1 is a hyperbolic pair for P_1. We also have

$$U_1 = (u_2, u_3 \cdots u_s) \perp P_1 \perp W$$

so we can proceed inductively since $u_2, U_3, \ldots u_s \in \operatorname{rad} U_1$. ■

Definition 26.20 *A subspace* $U \subset \mathrm{E}$ *is called totally isotropic if* $g|_U \equiv 0$.

Proposition 26.4 *(Witt decomposition) Suppose that* $U \subset \mathrm{E}$ *is a maximal totally isotropic subspace and* $e_1, e_2, \ldots e_r$ *a basis for* U. *Then there exist (null) vectors* $f_1, f_2, \ldots, f_r$ *such that each* $\{e_i, f_i\}$ *is a hyperbolic pair and* $U' = span\{f_i\}$ *is totally isotropic. Further*

$$\mathrm{E} = U \oplus U' \perp G$$

where $G = (U \oplus U')^{\perp}$.

Proof. Using the proof of the previous theorem we have $\operatorname{rad} U = U$ and $W = 0$. The present theorem now follows. ∎

Proposition 26.5 *If* g *is symmetric then* $g|_G$ *is definite.*

Example 26.5 *Let* $\mathrm{E}, g = \mathbb{C}^{2k}, g_0$ *where*

$$g_0(z, w) = \sum_{i=1}^{2k} z_i w_i$$

Let $\{e_1, ..., e_k, e_{k+1}, ..., e_{2k}\}$ *be the standard basis of* $\mathbb{C}^{2k}$. *Define*

$$\epsilon_j = \frac{1}{\sqrt{2}}(e_j + ie_{k+j}) \qquad j = 1, ..., k$$

and

$$\eta_j = \frac{1}{\sqrt{2}}(e_i - ie_{k+j}).$$

Then letting $F = span\{\epsilon_i\}$, $F' = span\{\eta_j\}$ *we have* $\mathbb{C}^{2k} = F \oplus F'$ *and* F *is a maximally isotropic subspace. Also, each* $\{\epsilon_j, \eta_j\}$ *is a hyperbolic pair.*

This is the most important example of a neutral space:

Proposition 26.6 *A vector space* E *with quadratic form is called neutral if the rank, that is, the dimension of a totally isotropic subspace, is* $r = dim\mathrm{E}/2$. *The resulting decomposition* $F \oplus F'$ *is called a (weak) polarization.*

26.7.3 Witt's Decomposition and Clifford Algebras

Even Dimension Suppose that V, Q is quadratic space over $\mathbb{K}$. Let $\dim \mathrm{V} = r$ and suppose that V, Q is neutral. Then we have that $C_{\mathbb{K}}$ is isomorphic to $\mathrm{End}(S)$ for an r dimensional space S (spinor space). In particular, $C_{\mathbb{K}}$ is a simple algebra.

Proof. Let $F \oplus F'$ be a polarization of V. Here, F and F' are maximal totally isotropic subspaces of V. Now let $\{x_1, ..., x_r, y_1, ..., y_r\}$ be a basis for V such that $\{x_i\}$ is a basis for F and $\{y_i\}$ a basis for F'. Set $f = y_1 y_2 \cdots y_h$. Now let S be the span of elements of the form $x_{i_1}x_{i_2}\cdots x_{i_r}f$ where $1 \le i_1 < ... < i_h \le r$. S is an ideal of $C_{\mathbb{K}}$ of dimension 2^r. We define a representation ρ of $C_{\mathbb{K}}$ in S by

$$\rho(u)s = us$$

This can be shown to be irreducible so that we have the desired result. ■

Now since we are interested in Spin which sits inside and in fact generates C_0 we need the following

Proposition 26.7 *C_0 is isomorphic to* $\mathrm{End}(S^+) \times \mathrm{End}(S^-)$ *where* $S^+ = C_0 \cap S$ *and* $S^- = C_1 \cap S$.

This follows from the obvious fact that each of $C_0 f$ and $C_1 f$ are invariant under multiplication by C_0.

Now consider a real quadratic space V, Q where Q is positive definite. We have $\mathrm{Spin}(n) \subset Cl^0(0) \subset C_0$ and $\mathrm{Spin}(n)$ generates C_0. Thus the complex spin group representation of is just given by restriction and is semisimple factoring as $S^+ \oplus S^-$.

Odd Dimension In the odd dimensional case we can not expect to find a polarization but this cloud turns out to have a silver lining. Let x_0 be a non-isotropic vector from V and set $\mathrm{V}_1 = (x_0)^{\perp}$. On V_1 we define a quadratic form Q_1 by

$$Q_1(y) = -Q(x_0)Q(y)$$

for $y \in \mathrm{V}_1$. It can be shown that Q_1 is non-degenerate. Now notice that for $y \in \mathrm{V}_1$ then $x_0 y = -y x_0$ and further

$$(x_0 y)^2 = -x_0^2 y^2 = -Q(x_0)Q(y) = Q_1(y)$$

so that by the universal mapping property the map

$$y \longrightarrow x_0 y$$

can be extended to an algebra morphism h from $Cl(Q_1, \mathrm{V}_1)$ to $C_{\mathbb{K}}$. Now these two algebras have the same dimension and since C_o is simple it must be an isomorphism. Now if Q has rank r then Q_1, V_1 is neutral and we obtain the following

Theorem 26.4 *If the dimension of* V *is odd and Q has rank r then C_0 is represented irreducibly in a space S^+ of dimension 2^r. In particular $C_0 \cong End(S^+)$.*

26.7.4 The Chirality operator

Let V be a Euclidean vector space with associated positive definite quadratic form Q. Let $\{e_1, ..., e_n\}$ be an oriented orthonormal. frame for V. We define the Chirality operator τ to be multiplication in the associated (complexified) Clifford algebra by the element

$$\tau = (\sqrt{-1})^{n/2} e_1 \cdots e_n$$

if n is even and by

$$\tau = (\sqrt{-1})^{(n+1)/2} e_1 \cdots e_n$$

if n is odd. Here $\tau \in Cl(n)$ and does not depend on the choice of orthonormal. oriented frame. We also have that $\tau v = -v\tau$ for $v \in$ V and $\tau^2 = 1$.

Let us consider the case of n even. Now we have seen that we can write V $\otimes C = F \oplus \bar{F}$ where F is totally isotropic and of dimension n. In fact we may assume that F has a basis $\{e_{2j-1} - ie_{2j} : 1 \leq j \leq n/2\}$, where the e_i come from an oriented orthonormal basis. Lets use this polarization to once again construct the spin representation.

First note that Q (or its associated bilinear form) places F and $\bar{F}$ in duality so that we can identify $\bar{F}$ with the dual space F'. Now set $S = \wedge F$. First we show how V act on S. Given $v \in$ Vconsider $v \in$ V$\otimes C$ and decompose $v = w + \bar{w}$ according to our decomposition above. Define $\phi_w s = \sqrt{2} w \wedge s$ and

$$\phi_{\bar{w}} s = -\iota(\bar{w}) s.$$

where ι is interior multiplication. Now extend ϕ linearly to V. Exercise Show that ϕ extends to a representation of $C \otimes Cl(n)$. Show that $S^+ = \wedge^+ F$ is invariant under C_0. It turns out that ϕ_τ is $(-1)^k$ on $\wedge^k F$

26.7.5 Spin Bundles and Spin-c Bundles

26.7.6 Harmonic Spinors

Chapter 27

Classical Mechanics

Every body continues in its state of rest or uniform motion in a straight line, except insofar as it doesn't.

Arthur Eddington, Sir

27.1 Particle motion and Lagrangian Systems

If we consider a *single* particle of mass m then Newton's law is

$$m\frac{d^2\mathbf{x}}{dt^2} = \mathbf{F}(\mathbf{x}(t), t)$$

The force $\mathbf{F}$ is **conservative** if it doesn't depend on time and there is a potential function $V : \mathbb{R}^3 \to \mathbb{R}$ such that $\mathbf{F}(\mathbf{x}) = -\operatorname{grad} V(\mathbf{x})$. Assume this is the case. Then Newton's law becomes

$$m\frac{d^2}{dt^2}\mathbf{x}(t) + \operatorname{grad} V(\mathbf{x}(t)) = 0.$$

Consider the Affine space $C(I, \mathbf{x}_1, \mathbf{x}_2)$ of all C^2 paths from $\mathbf{x}_1$ to $\mathbf{x}_2$ in $\mathbb{R}^3$ defined on the interval $I = [t_1, t_2]$. This is an Affine space modelled on the Banach space $C_0^r(I)$ of all C^r functions $\varepsilon : I \to \mathbb{R}^3$ with $\varepsilon(t_1) = \varepsilon(t_1) = 0$ and with the norm

$$\|\varepsilon\| = \sup_{t\in I}\{|\varepsilon(t)| + |\varepsilon'(t)| + |\varepsilon''(t)|\}.$$

If we define the fixed affine linear path $\mathbf{a} : I \to \mathbb{R}^3$ by $\mathbf{a}(t) = \mathbf{x}_1 + \frac{t-t_1}{t_2-t_1}(\mathbf{x}_2 - \mathbf{x}_1)$ then all we have a coordinatization of $C(I, \mathbf{x}_1, \mathbf{x}_2)$ by $C_0^r(I)$ given by the single chart $\psi : \mathbf{c} \mapsto \mathbf{c} - \mathbf{a} \in C_0^2(I)$. Then the tangent space to $C(I, \mathbf{x}_1, \mathbf{x}_2)$ at a fixed path c_0 is just $C_0^2(I)$. Now we have the function S defined on $C(I, \mathbf{x}_1, \mathbf{x}_2)$ by

$$S(\mathbf{c}) = \int_{t_1}^{t_2} \left(\frac{1}{2}m\,\|\mathbf{c}'(t)\|^2 - V(\mathbf{c}(t))\right) dt$$

The variation of S is just the 1-form $\delta S : C_0^2(I) \to \mathbb{R}$ defined by

$$\delta S \cdot \varepsilon = \left.\frac{d}{d\tau}\right|_{\tau=0} S(\mathbf{c}_0 + \tau\varepsilon)$$

Let us suppose that $\delta S = 0$ at c_0. Then we have

$$\begin{aligned}
0 &= \left.\frac{d}{d\tau}\right|_{\tau=0} S(\mathbf{c}_0'(t) + \tau\varepsilon) \\
&= \left.\frac{d}{d\tau}\right|_{\tau=0} \int_{t_1}^{t_2} \left(\frac{1}{2}m \left\|\mathbf{c}_0'(t) + \tau\varepsilon'(t)\right\|^2 - V(\mathbf{c}_0(t) + \tau\varepsilon(t))\right) dt \\
&= \int_{t_1}^{t_2} \left(m \left\langle \mathbf{c}_0'(t), \varepsilon'(t)\right\rangle - \frac{\partial V}{\partial x^i}(\mathbf{c}_0)\frac{d\varepsilon^i}{dt}(0)\right) dt \\
&= \int_{t_1}^{t_2} \left(m\mathbf{c}_0'(t) \cdot \frac{d}{dt}\varepsilon(t) - \operatorname{grad} V(\mathbf{c}_0(t)) \cdot \varepsilon(t)\right) dt \\
&\int_{t_1}^{t_2} (m\mathbf{c}_0''(t) - \operatorname{grad} V(\mathbf{c}_0(t))) \cdot \varepsilon(t) dt
\end{aligned}$$

Now since this is true for every choice of $\varepsilon \in C_0^2(I)$ we see that

$$m\mathbf{c}_0''(t) - \operatorname{grad} V(\mathbf{c}_0(t)) = 0$$

thus we see that $\mathbf{c}_0(t) = \mathbf{x}(t)$ is a critical point in $C(I, \mathbf{x}_1, \mathbf{x}_2)$, that is, a stationary path, if and only if **??** is satisfied.

27.1.1 Basic Variational Formalism for a Lagrangian

In general we consider a differentiable manifold Q as our state space and then a Lagrangian density function L is given on TQ. For example we can take a potential function $V : Q \to \mathbb{R}$, a Riemannian metric g on Q and define the action functional S on the space of smooth paths $I \to Q$ beginning and ending at a fixed points p_1 and p_2 given by

$$\begin{aligned}
S(c) = &\int_{t_1}^{t_2} \mathrm{L}(c'(t))dt = \\
&\int_{t_1}^{t_2} \frac{1}{2}m \left\langle c'(t), c'(t)\right\rangle - V(c(t))dt
\end{aligned}$$

The tangent space at a fixed c_0 is the Banach space $\Gamma_0^2(c_0^*TQ)$ of C^2 vector fields ε: $I \to TQ$ along c_0 that vanish at t_1 and t_2. A curve with tangent ε at c_0 is just a variation $v : (-\epsilon, \epsilon) \times I \to Q$ such that $\varepsilon(t) = \left.\frac{\partial}{\partial s}\right|_{s=0} v(s,t)$ is the variation vector field. Then we have

$$\begin{aligned}\delta S\cdot\varepsilon &= \left.\frac{\partial}{\partial s}\right|_{s=0}\int_{t_1}^{t_2}\mathrm{L}\left(\frac{\partial v}{\partial t}(s,t)\right)dt\\ &= \int_{t_1}^{t_2}\left.\frac{\partial}{\partial s}\right|_{s=0}\mathrm{L}\left(\frac{\partial v}{\partial t}(s,t)\right)dt\\ &= etc.\end{aligned}$$

Let us examine this in the case of $Q = U \subset \mathbb{R}^n$. With $\mathbf{q} = (q^1, ...q^n)$ being (general curvilinear) coordinates on U we have natural (tangent bundle chart) coordinates $\mathbf{q}, \dot{\mathbf{q}}$ on $TU = U \times \mathbb{R}^n$. Assume that the variation has the form $\mathbf{q}(s,t) = \mathbf{q}(t) + s\varepsilon(t)$. Then we have

$$\begin{aligned}\delta S\cdot\varepsilon &= \left.\frac{\partial}{\partial s}\right|_{s=0}\int_{t_1}^{t_2}\mathrm{L}(\mathbf{q}(s,t),\dot{\mathbf{q}}(s,t))dt\\ &\int_{t_1}^{t_2}\left.\frac{\partial}{\partial s}\right|_{s=0}\mathrm{L}(\mathbf{q}+s\varepsilon,\dot{\mathbf{q}}+s\varepsilon)dt\\ &= \int_{t_1}^{t_2}(\frac{\partial \mathrm{L}}{\partial \mathbf{q}}(\mathbf{q},\dot{\mathbf{q}})\cdot\varepsilon + \frac{\partial \mathrm{L}}{\partial \dot{\mathbf{q}}}(\mathbf{q},\dot{\mathbf{q}})\cdot\varepsilon)dt\\ &= \int_{t_1}^{t_2}(\frac{\partial \mathrm{L}}{\partial \mathbf{q}}(\mathbf{q},\dot{\mathbf{q}})\cdot\varepsilon - \frac{d}{dt}\frac{\partial \mathrm{L}}{\partial \dot{\mathbf{q}}}(\mathbf{q},\dot{\mathbf{q}})\cdot\varepsilon)dt\\ &= \int_{t_1}^{t_2}(\frac{\partial \mathrm{L}}{\partial \mathbf{q}}(\mathbf{q},\dot{\mathbf{q}}) - \frac{d}{dt}\frac{\partial \mathrm{L}}{\partial \dot{\mathbf{q}}}(\mathbf{q},\dot{\mathbf{q}}))\cdot\varepsilon dt\end{aligned}$$

and since ε was arbitrary we get the Euler-Lagrange equations for the motion

$$\frac{\partial \mathrm{L}}{\partial \mathbf{q}}(\mathbf{q},\dot{\mathbf{q}}) - \frac{d}{dt}\frac{\partial \mathrm{L}}{\partial \dot{\mathbf{q}}}(\mathbf{q},\dot{\mathbf{q}}) = 0$$

In general, a time-independent Lagrangian on a manifold Q is a smooth function on the tangent bundle:

$$\mathrm{L} : TQ \to Q$$

and the associated action functional is a map from the space of smooth curves $C^\infty([a,b], Q)$ defined for $c : [a,b] \to Q$ by

$$\mathcal{S}_L(c) = \int_a^b \mathrm{L}(\dot{c}(t))dt$$

where $\dot{c} : [a,b] \to TQ$ is the canonical lift (velocity). A time dependent Lagrangian is a smooth map

$$L : \mathbb{R} \times TQ \to Q$$

where the first factor $\mathbb{R}$ is the time t, and once again we have the associated action functional $\mathcal{S}_L(c) = \int_a^b L(t, \dot{c}(t))dt$.

Let us limit ourselves initially to the time independent case.

Definition 27.1 *A smooth variation of a curve $c : [a, b] \to Q$ is a smooth map $\nu : [a, b] \times (-\epsilon, \epsilon) \to Q$ for small ϵ such that $\nu(t, 0) = c(t)$. We call the variation a* ***variation with fixed endpoints*** *if $\nu(a, s) = c(a)$ and $\nu(b, s) = c(b)$ for all $s \in (-\epsilon, \epsilon)$. Now we have a family of curves $\nu_s = \nu(., s)$. The infinitesimal variation at ν_0 is the vector field along c defined by $V(t) = \frac{d\nu}{ds}(t, 0)$. This V is called the* ***variation vector field*** *for the variation. The differential of the functional δS_L (classically called the first variation) is defined as*

$$\begin{aligned}\delta S_{\mathrm{L}}(c) \cdot V &= \left.\frac{d}{ds}\right|_{s=0} S_{\mathrm{L}}(\nu_s) \\ &= \left.\frac{d}{ds}\right|_{s=0} \int_a^b \mathrm{L}(\nu_s(t))dt\end{aligned}$$

Remark 27.1 *Every smooth vector field along c is the variational vector field coming from some variation of c and for any other variation ν' with $V(t) = \frac{d\nu'}{ds}(t, 0)$ the about computed quantity $\delta S_L(c) \cdot V$ will be the same.*

At any rate, if $\delta S_L(c) \cdot V = 0$ for all variations vector fields V along c and vanishing at the endpoints then we write $\delta S_L(c) = 0$ and call c **critical** (or **stationary**) for L.

Now consider the case where the image of c lies in some coordinate chart $U, \psi = q^1, q^2, ...q^n$ and denote by $TU, T\psi = (q^1, q^2, ..., q^n, \dot{q}^1, \dot{q}^2, ..., \dot{q}^n)$ the natural chart on $TU \subset TQ$. In other words, $T\psi(\xi) = (q^1 \circ \tau(\xi), q^2 \circ \tau(\xi), ..., q^n \circ \tau(\xi), dq^1(\xi), dq^2(\xi), ..., dq^n(\xi))$. Thus the curve has coordinates

$$(c, \dot{c}) = (q^1(t), q^2(t), ..., q^n(t), \dot{q}^1(t), \dot{q}^2(t), ..., \dot{q}^n(t))$$

where now the $\dot{q}^i(t)$ *really are* time derivatives. In this local situation we can choose our variation to have the form $q^i(t) + s\delta q^i(t)$ for some functions $\delta q^i(t)$ vanishing at a and b and some parameter s with respect to which we will differentiate. The lifted variation is $(\mathbf{q}(t) + s\delta(t), \dot{\mathbf{q}}(t) + s\delta\dot{\mathbf{q}}(t))$ which is the obvious abbreviation for a path in $T\psi(TU) \subset \mathbb{R}^n \times \mathbb{R}^n$. Now we have seen above that the path c will be critical if

$$\left.\frac{d}{ds}\right|_{s=0} \int \mathrm{L}(\mathbf{q}(t) + s\delta(t), \dot{\mathbf{q}}(t) + s\delta\dot{\mathbf{q}}(t)))dt = 0$$

for all such variations and the above calculations lead to the result that

$$\frac{\partial}{\partial \mathbf{q}} \mathrm{L}(\mathbf{q}(t), \dot{\mathbf{q}}(t)) - \frac{d}{dt}\frac{\partial}{\partial \dot{\mathbf{q}}} \mathrm{L}(\mathbf{q}(t), \dot{\mathbf{q}}(\mathbf{t})) = \mathbf{0} \quad \text{Euler-Lagrange}$$

for any L-critical path (with image in this chart). Here $n = \dim(Q)$.

It can be show that even in the case that the image of c does not lie in the domain of a chart that c is L-critical path if it can be subdivided into sub-paths lying in charts and L-critical in each such chart.

27.1.2 Two examples of a Lagrangian

Example 27.1 *Suppose that we have a 1-form $\theta \in \mathfrak{X}^*(Q)$. A 1-form is just a map $\theta : TQ \to \mathbb{R}$ that happens to be linear on each fiber T_pQ. Thus we may examine the special case of $\mathrm{L} = \theta$. In canonical coordinates $(q, \dot{q})$ again,*

$$\mathrm{L} = \theta = \sum a_i(q) dq^i$$

for some functions $a_i(q)$. An easy calculation shows that the Euler-Lagrange equations become

$$\left(\frac{\partial a_i}{\partial q^k} - \frac{\partial a_k}{\partial q^i}\right) \dot{q}^i = 0$$

but on the other hand

$$d\theta = \frac{1}{2} \sum \left(\frac{\partial a_j}{\partial q^i} - \frac{\partial a_i}{\partial q^j}\right) \partial q^i \wedge \partial q^j$$

and one can conclude that if $c = (q^i(t))$ is critical for $L = \theta$ then for any vector field X defined on the image of c we have

$$\frac{1}{2} \sum \left(\frac{\partial a_j}{\partial q^i}(q^i(t)) - \frac{\partial a_i}{\partial q^j}(q^i(t))\right) \dot{q}^i(t) X^j$$

or $d\theta(\dot{c}(t), X) = 0$. This can be written succinctly as

$$\iota_{\dot{c}(t)} d\theta = 0.$$

Example 27.2 *Now let us take the case of a Riemannian manifold M, g and let $\mathrm{L}(v) = \frac{1}{2}\mathrm{g}(v, v)$. Thus the action functional is the "energy"*

$$S_g(c) = \int g(\dot{c}(t), \dot{c}(t)) dt$$

In this case the critical paths are just geodesics.

27.2 Symmetry, Conservation and Noether's Theorem

Let G be a Lie group acting on a smooth manifold M.

$$\lambda : G \times M \to M$$

As usual we write $g \cdot x$ for $\lambda(g, x)$. We have a fundamental vector field $\xi^\natural$ associated to every $\xi \in \mathfrak{g}$ defined by the rule

$$\xi^\natural(p) = T_{(e,p)}\lambda \cdot (., 0)$$

or equivalently by the rule

$$\xi^{\natural}(p) = \left.\frac{d}{dt}\right|_0 \exp(t\xi)\cdot p$$

The map $\natural : \xi \mapsto \xi^{\natural}$ is a Lie algebra anti-homomorphism. Of course, here we are using the flow associated to ξ

$$\mathrm{Fl}^{\xi}(t,p) := \mathrm{Fl}^{\xi^{\natural}}(t,p) = \exp(t\xi)\cdot p$$

and it should be noted that $t \mapsto \exp(t\xi)$ is the one parameter subgroup associated to ξ and to get the corresponding left invariant vector field $X^{\xi} \in \mathfrak{X}^L(G)$ we act on the right:

$$X^{\xi}(g) = \left.\frac{d}{dt}\right|_0 g\cdot \exp(t\xi)$$

Now a diffeomorphism acts on a covariant k-tensor field contravariantly according to

$$(\phi^*K)(p)(v_1, ...v_k) = K(\phi(p))(T\phi v_1, ...T\phi v_k)$$

Suppose that we are given a covariant tensor field $\Upsilon \in \mathfrak{T}(M)$ on M. We think of Υ as defining some kind of extra structure on M. The two main examples for our purposes are

1. $\Upsilon = \langle .,. \rangle$ a nondegenerate covariant symmetric 2-tensor. Then $M, \langle .,. \rangle$ is a (semi-) Riemannian manifold.
2. $\Upsilon = \omega \in \Omega^2(M)$ a non-degenerate 2-form. Then M, ω is a **symplectic manifold**.

Then G acts on Υ since G acts on M as diffeomorphisms. We denote this natural (left) action by $g \cdot \Upsilon$. If $g\cdot\Upsilon = \Upsilon$ for all $g \in G$ we say that G acts by symmetries of the pair M, Υ.

Definition 27.2 *In general, a vector field X on M, Υ is called an* ***infinitesimal symmetry*** *of the pair M, Υ if $\mathcal{L}_X\Upsilon = 0$. Other terminology is that X is a $\Upsilon-$**Killing field** . The usual notion of a Killing field in (pseudo-) Riemannian geometry is the case when $\Upsilon = \langle , \rangle$ is the metric tensor.*

Example 27.3 *A group G is called a symmetry group of a symplectic manifold M, ω if G acts by symplectomorphisms so that $g\cdot\omega = \omega$ for all $g \in G$. In this case, each $\xi \in \mathfrak{g}$ is an infinitesimal symmetry of M, ω meaning that*

$$\mathcal{L}_{\xi}\omega = 0$$

where $\mathcal{L}_{\xi}$ is by definition the same as $\mathcal{L}_{\xi^{\natural}}$. This follows because if we let $g_t = \exp(t\xi)$ then each g_t is a symmetry so $g_t^*\omega = 0$ and

$$\mathcal{L}_{\xi}\omega = \left.\frac{d}{dt}\right|_0 g_t^*\omega = 0$$

27.2.1 Lagrangians with symmetries.

We need two definitions

Definition 27.3 *If $\phi : M \to M$ is a diffeomorphism then the induced tangent map $T\phi : TM \to TM$ is called the* ***canonical lift***.

Definition 27.4 *Given a vector field $X \in \mathfrak{X}(M)$ there is a lifting of X to $\widetilde{X} \in \mathfrak{X}(TM) = \Gamma(TM, TTM)$*

$$\begin{array}{rcc} \widetilde{X}: & TM \to & TTM \\ & \downarrow & \downarrow \\ X: & M \to & TM \end{array}$$

such that the flow $\mathrm{Fl}^{\widetilde{X}}$ is the canonical lift of Fl^{X}

$$\begin{array}{rcc} \mathrm{Fl}_t^{\widetilde{X}}: & TM \to & TM \\ & \downarrow & \downarrow \\ \mathrm{Fl}_t^{X}: & M \to & M \end{array}.$$

In other words, $\mathrm{Fl}^{\widetilde{X}} = T\mathrm{Fl}_t^X$. We simply define $\widetilde{X}(v) = \frac{d}{dt}(T\mathrm{Fl}_t^X \cdot v)$.

Definition 27.5 *Let ω_{L} denote the unique 1-form on Q that in canonical coordinates is $\omega_{\mathrm{L}} = \sum_{i=1}^{n} \frac{\partial \mathrm{L}}{\partial \dot{q}^i} dq^i$.*

Theorem 27.1 (E. Noether) *If X is an infinitesimal symmetry of the Lagrangian then the function $\omega_{\mathrm{L}}(\widetilde{X})$ is constant along any path $c : I \subset \mathbb{R}$ that is stationary for the action associated to* L.

Let's prove this using local coordinates $(q^i, \dot{q}^i)$ for $TU_\alpha \subset TQ$. It turn out that locally,

$$\widetilde{X} = \sum_i \left(a^i \frac{\partial}{\partial q^i} + \sum_j \frac{\partial a^i}{\partial q^j} \frac{\partial}{\partial \dot{q}^i} \right)$$

where a^i is defined by $X = \sum a^i(q) \frac{\partial}{\partial q^i}$. Also, $\omega_L(\widetilde{X}) = \sum a^i \frac{\partial L}{\partial \dot{q}^i}$. Now suppose that $q^i(t), \dot{q}^i(t) = \frac{d}{dt} q^i(t)$ satisfies the Euler-Lagrange equations. Then

$$\begin{aligned}
&\frac{d}{dt}\omega_{\mathrm{L}}(\widetilde{X})(q^i(t), \dot{q}^i(t)) \\
&= \frac{d}{dt} \sum a^i(q^i(t)) \frac{\partial \mathrm{L}}{\partial \dot{q}^i}(q^i(t), \dot{q}^i(t)) \\
&= \sum \frac{da^i}{dt}(q^i(t)) \frac{\partial \mathrm{L}}{\partial \dot{q}^i}(q^i(t), \dot{q}^i(t)) + a^i(q^i(t)) \frac{d}{dt} \frac{\partial \mathrm{L}}{\partial \dot{q}^i}(q^i(t), \dot{q}^i(t)) \\
&= \sum_i \left[\sum_j \frac{da^i}{dq^j} \dot{q}^j(t) \frac{\partial \mathrm{L}}{\partial \dot{q}^i}(q^i(t), \dot{q}^i(t)) + a^i(q^i(t)) \frac{\partial \mathrm{L}}{\partial q^i}(q^i(t), \dot{q}^i(t)) \right] \\
&= d\mathrm{L}(X) = \mathcal{L}_X \mathrm{L} = 0
\end{aligned}$$

This theorem tells us one case when we get a conservation law. A conservation law is a function C on TQ (or T^*Q for the Hamiltonian flow) such that C is constant along solution paths. (i.e. stationary for the action or satisfying the Euler-Lagrange equations.)

$$\mathrm{L} : TQ \to Q$$

let $X \in T(TQ)$.

27.2.2 Lie Groups and Left Invariants Lagrangians

Recall that G act on itself by left translation $l_g : G \to G$. The action lifts to the tangent bundle $Tl_g : TG \to TG$. Suppose that $\mathrm{L} : TG \to \mathbb{R}$ is invariant under this left action so that $\mathrm{L}(Tl_g X_h) = \mathrm{L}(X_p)$ for all $g, h \in G$. In particular, $\mathrm{L}(Tl_g X_e) = L(X_e)$ so L is completely determined by its restriction to $T_eG = \mathfrak{g}$. Define the restricted Lagrangian function by $\Lambda = \mathrm{L}|_{T_eG}$. We view the differential $d\Lambda$ as a map $d\Lambda : \mathfrak{g} \to \mathbb{R}$ and so in fact $d\lambda \in \mathfrak{g}^*$. Next, recall that for any $\xi \in \mathfrak{g}$ the map $\mathrm{ad}_\xi : \mathfrak{g} \to \mathfrak{g}$ is defined by $\mathrm{ad}_\xi v = [\xi, v]$ and we have the adjoint map $\mathrm{ad}^*_\xi : \mathfrak{g}^* \to \mathfrak{g}^*$. Now let $t \mapsto g(t)$ be a motion of the system and define the "body velocity" by $\nu_c(t) = Tl_{c(t)^{-1}} \cdot c'(t) = \omega_G(c'(t))$. Then we have

Theorem 27.2 *Assume L is invariant as above. The curve $c(.)$ satisfies the Euler-Lagrange equations for* L *if and only if*

$$\frac{d}{dt} d\Lambda(\nu_c(t)) = \mathrm{ad}^*_{\nu_c(t)} d\Lambda$$

27.3 The Hamiltonian Formalism

Let us now examine the change to a description in cotangent chart $\mathbf{q}, \mathbf{p}$ so that for a covector at $\mathbf{q}$ given by $\mathbf{a}(\mathbf{q}) \cdot d\mathbf{q}$ has coordinates $\mathbf{q}, \mathbf{a}$. Our method of transfer to the cotangent side is via the Legendre transformation induced by L. In fact, this is just the fiber derivative defined above. We must assume that the map $F : (\mathbf{q}, \dot{\mathbf{q}}) \mapsto (\mathbf{q}, \mathbf{p}) = (\mathbf{q}, \frac{\partial L}{\partial \dot{\mathbf{q}}}(\mathbf{q}, \dot{\mathbf{q}}))$ is a diffeomorphism (this is written with respect to the two natural charts on TU and T^*U). Again this just means that the Lagrangian is nondegenerate. Now if $v(t) = (\mathbf{q}(t), \dot{\mathbf{q}}(t))$ is (a lift of) a solution curve then defining the **Hamiltonian** function

$$\widetilde{H}(\mathbf{q}, \dot{\mathbf{q}}) = \frac{\partial \mathrm{L}}{\partial \dot{\mathbf{q}}}(\mathbf{q}, \dot{\mathbf{q}}) \cdot \dot{\mathbf{q}} - \mathrm{L}(\mathbf{q}, \dot{\mathbf{q}})$$

we compute with $\dot{\mathbf{q}} = \frac{d}{dt}\mathbf{q}$

$$\begin{aligned} \frac{d}{dt}\widetilde{H}(\mathbf{q}, \dot{\mathbf{q}}) &= \frac{d}{dt}(\frac{\partial \mathrm{L}}{\partial \dot{\mathbf{q}}}(\mathbf{q}, \dot{\mathbf{q}}) \cdot \dot{\mathbf{q}}) - \frac{d}{dt}\mathrm{L}(\mathbf{q}, \dot{\mathbf{q}}) \\ &= \frac{\partial \mathrm{L}}{\partial \dot{\mathbf{q}}}(\mathbf{q}, \dot{\mathbf{q}}) \frac{d}{dt}\dot{\mathbf{q}} + \frac{d}{dt}\frac{\partial \mathrm{L}}{\partial \dot{\mathbf{q}}}(\mathbf{q}, \dot{\mathbf{q}}) \cdot \dot{\mathbf{q}} - \frac{d}{dt}\mathrm{L}(\mathbf{q}, \dot{\mathbf{q}}) \\ &= 0 \end{aligned}$$

we have used that the Euler-Lagrange equations $\frac{\partial \mathrm{L}}{\partial \mathbf{q}}(\mathbf{q}, \dot{\mathbf{q}}) - \frac{d}{dt}\frac{\partial \mathrm{L}}{\partial \dot{\mathbf{q}}}(\mathbf{q}, \dot{\mathbf{q}}) = 0$. Thus differential form $d\widetilde{H} = \frac{\partial \widetilde{H}}{\partial \mathbf{q}} d\mathbf{q} + \frac{\partial \widetilde{H}}{\partial \dot{\mathbf{q}}} d\dot{\mathbf{q}}$ is zero on the velocity $v'(t) = \frac{d}{dt}(\mathbf{q}, \dot{\mathbf{q}})$

$$\begin{aligned} d\widetilde{H} \cdot v'(t) &= d\widetilde{H} \cdot \frac{d}{dt}(\mathbf{q}, \dot{\mathbf{q}}) \\ &= \frac{\partial \widetilde{H}}{\partial \mathbf{q}} \frac{d\mathbf{q}}{dt} + \frac{\partial \widetilde{H}}{\partial \dot{\mathbf{q}}} \frac{d\dot{\mathbf{q}}}{dt} = 0 \end{aligned}$$

We then use the inverse of this diffeomorphism to transfer the Hamiltonian function to a function $H(\mathbf{q}, \mathbf{p}) = F^{-1*}\widetilde{H}(\mathbf{q}, \dot{\mathbf{q}}) = \mathbf{p} \cdot \dot{\mathbf{q}}(\mathbf{q}, \mathbf{p}) - \mathrm{L}(\mathbf{q}, \dot{\mathbf{q}}(\mathbf{q}, \mathbf{p}))$.. Now if $\mathbf{q}(t), \dot{\mathbf{q}}(t)$ is a solution curve then its image $b(t) = F \circ v(t) = (\mathbf{q}(t), \mathbf{p}(t))$ satisfies

$$\begin{aligned} dH(b'(t)) &= (dH \cdot TF.v'(t)) \\ &= (F^* dH) \cdot v'(t) \\ &= d(F^* H) \cdot v'(t) \\ &= d\widetilde{H} \cdot v'(t) = 0 \end{aligned}$$

so we have that

$$0 = dH(b'(t)) = \frac{\partial H}{\partial \mathbf{q}} \cdot \frac{d\mathbf{q}}{dt} + \frac{\partial H}{\partial \mathbf{p}} \cdot \frac{d\mathbf{p}}{dt}$$

but also

$$\frac{\partial}{\partial \mathbf{p}} H(\mathbf{q}, \mathbf{p}) = \dot{\mathbf{q}} + \mathbf{p} \cdot \frac{\partial \dot{\mathbf{q}}}{\partial \mathbf{p}} - \frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \frac{\partial \dot{\mathbf{q}}}{\partial \mathbf{p}} = \dot{\mathbf{q}} = \frac{d\mathbf{q}}{dt}$$

solving these last two equations simultaneously we arrive at Hamilton's equations of motion:

$$\begin{aligned} \frac{d}{dt}\mathbf{q}(t) &= \frac{\partial H}{\partial \mathbf{p}}(\mathbf{q}(t), \mathbf{p}(t)) \\ \frac{d}{dt}\mathbf{p}(t) &= -\frac{\partial H}{\partial \mathbf{q}}(\mathbf{q}(t), \mathbf{p}(t)) \end{aligned}$$

or

$$\frac{d}{dt}\begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{pmatrix} \frac{\partial H}{\partial \mathbf{q}} \\ \frac{\partial H}{\partial \mathbf{p}} \end{pmatrix}$$

Remark 27.2 *One can calculate directly that* $\frac{dH}{dt}(\mathbf{q}(t), \mathbf{p}(t)) = 0$ *for solutions these equations. If the Lagrangian was originally given by* $\mathrm{L} = \frac{1}{2}K - V$ *for some kinetic energy function and a potential energy function then this amounts to conservation of energy. We will see that this follows from a general principle below.*

Chapter 28

Symplectic Geometry

Equations are more important to me, because politics is for the present, but an equation is something for eternity

-Einstein

28.1 Symplectic Linear Algebra

A (real) **symplectic vector space** is a pair V, α where V is a (real) vector space and α is a nondegenerate alternating (skew-symmetric) bilinear form $\alpha : \mathrm{V} \times \mathrm{V} \to \mathbb{R}$. The basic example is $\mathbb{R}^{2n}$ with

$$\alpha_0(x, y) = x^t J_n y$$

where

$$J_n = \begin{pmatrix} 0 & I_{n \times n} \\ -I_{n \times n} & 0 \end{pmatrix}.$$

The standard symplectic form on α_0 is typical. It is a standard fact from linear algebra that for any N dimensional symplectic vector space V, α there is a basis $e_1, ..., e_n, f^1, ..., f^n$ called a **symplectic basis** such that the matrix that represents α with respect to this basis is the matrix J_n. Thus we may write

$$\alpha = e^1 \wedge f_1 + ... + e^n \wedge f_n$$

where $e^1, ..., e^n, f_1, ..., f_n$ is the dual basis to $e_1, ..., e_n, f^1, ..., f^n$. If V, η is a vector space with a not necessarily nondegenerate alternating form η then we can define the null space

$$N_\eta = \{v \in \mathrm{V} : \eta(v, w) = 0 \text{ for all } w \in \mathrm{V}\}.$$

On the quotient space $\overline{\mathrm{V}} = \mathrm{V}/N_\eta$ we may define $\overline{\eta}(\overline{v}, \overline{w}) = \eta(v, w)$ where v and w represent the elements $\overline{v}, \overline{w} \in \overline{\mathrm{V}}$. Then $\overline{\mathrm{V}}, \overline{\eta}$ is a symplectic vector space called the symplectic reduction of V, η.

Proposition 28.1 *For any $\eta \in \bigwedge V^*$ (regarded as a bilinear form) there is linearly independent set of elements $e^1, ..., e^k, f_1, ..., f_k$ from V^* such that*

$$\eta = e^1 \wedge f_1 + ... + e^k \wedge f_k$$

where $\dim(V) - 2k \geq 0$ *is the dimension of* N_η.

Definition 28.1 *Note: The number k is called the* ***rank*** *of η. The matrix that represents η actually has rank $2k$ and so some might call k the* ***half rank*** *of η.*

Proof. Consider the symplectic reduction $\overline{V}, \overline{\eta}$ of V, η and choose set of elements $e^1, ..., e^k, f_1, ..., f_k$ such that $\bar{e}^1, ..., \bar{e}^k, \bar{f}_1, ..., \bar{f}_k$ form a symplectic basis of $\overline{V}, \overline{\eta}$. Add to this set a basis $b_1, ..., b_l$ a basis for N_η and verify that $e^1, ..., e^k, f_1, ..., f_k, b_1, ..., b_l$ must be a basis for V. Taking the dual basis one can check that

$$\eta = e^1 \wedge f_1 + ... + e^k \wedge f_k$$

by testing on the basis $e^1, ..., e^k, f_1, ..., f_k, b_1, ..., b_l$. ■

Now if W is a subspace of a symplectic vector space then we may define

$$W^\perp = \{v \in V : \eta(v, w) = 0 \text{ for all } w \in W\}$$

and it is true that $\dim(W) + \dim(W^\perp) = \dim(V)$ but it is **not** necessarily the case that $W \cap W^\perp = 0$. In fact,we classify subspaces W by two numbers: $d = \dim(W)$ and $\nu = \dim(W \cap W^\perp)$. If $\nu = 0$ then $\eta|_W, W$ is a symplectic space and so we call W a **symplectic subspace** . At the opposite extreme, if $\nu = d$ then W is called a **Lagrangian subspace** . If $W \subset W^\perp$ we say that W is an **isotropic subspace**.

A linear transformation between symplectic vector spaces $\ell : V_1, \eta_1 \to V_2, \eta_2$ is called a **symplectic linear map** if $\eta_2(\ell(v), \ell(w)) = \eta_1(v, w)$ for all $v, w \in V_1$; In other words, if $\ell^*\eta_2 = \eta_1$. The set of all symplectic linear isomorphisms from V, η to itself is called the **symplectic group** and denoted $Sp(V,\eta)$. With respect to a symplectic basis $\mathcal{B}$ a symplectic linear isomorphism ℓ is represented by a matrix $A = [\ell]_\mathcal{B}$ that satisfies

$$A^t J A = J$$

where $J = J_n$ is the matrix defined above and where $2n = \dim(V)$. Such a matrix is called a symplectic matrix and the group of all such is called the **symplectic matrix group** and denoted $Sp(n, \mathbb{R})$. Of course if $\dim(V) = 2n$ then $Sp(V,\eta) \cong Sp(n, \mathbb{R})$ the isomorphism depending a choice of basis. If η is a symplectic from on V with $\dim(V) = 2n$ then $\eta^n \in \wedge^{2n}V$ is nonzero and so orients the vector space V.

Lemma 28.1 *If $A \in Sp(n, \mathbb{R})$ then* $\det(A) = 1$.

Proof. If we use A as a linear transformation $\mathbb{R}^{2n} \to \mathbb{R}^{2n}$ then $A^*\alpha_0 = \alpha_0$ and $A^*\alpha_0^n = \alpha_0^n$ where α_0 is the standard symplectic form on $\mathbb{R}^{2n}$ and $\alpha_0^n \in \wedge^{2n}\mathbb{R}^{2n}$ is top form. Thus $\det A = 1$. ■

Theorem 28.1 (Symplectic eigenvalue theorem) *If λ is a (complex) eigenvalue of a symplectic matrix A then so is $1/\lambda$, $\bar{\lambda}$ and $1/\bar{\lambda}$.*

Proof. Let $p(\lambda) = \det(A - \lambda I)$ be the characteristic polynomial. It is easy to see that $J^t = -J$ and $JAJ^{-1} = (A^{-1})^t$. Using these facts we have

$$\begin{aligned} p(\lambda) &= \det(J(A - \lambda I)J^{-1}) = \det(A^{-1} - \lambda I) \\ &= \det(A^{-1}(I - \lambda A)) = \det(I - \lambda A) \\ &= \lambda^{2n} \det(\frac{1}{\lambda} I - A)) = \lambda^{2n} p(1/\lambda). \end{aligned}$$

So we have $p(\lambda) = \lambda^{2n} p(1/\lambda)$. Using this and remembering that 0 is not an eigenvalue one concludes that $1/\lambda$ and $\bar{\lambda}$ are eigenvalues of A. ■

Exercise 28.1 *With respect to the last theorem, show that λ and $1/\lambda$ have the same multiplicity.*

28.2 Canonical Form (Linear case)

Suppose one has a vector space W with dual W*. We denote the pairing between W and W* by $\langle .,. \rangle$. There is a simple way to produce a symplectic form on the space Z = W × W* which we will call the **canonical symplectic form**. This is defined by

$$\Omega((v_1, \alpha_1), (v_2, \alpha_2)) := \langle \alpha_2, v_1 \rangle - \langle \alpha_1, v_2 \rangle.$$

If W is an inner product space with inner product $\langle .,. \rangle$ then we may form the canonical symplectic from on Z = W × W by the same formula. As a special case we get the standard symplectic form on $\mathbb{R}^{2n} = \mathbb{R}^n \times \mathbb{R}^n$ given by

$$\Omega((x, y), (\widetilde{x}, \widetilde{y})) = \widetilde{y} \cdot x - y \cdot \widetilde{x}.$$

28.3 Symplectic manifolds

Definition 28.2 *A **symplectic form** on a manifold M is a* nondegenerate closed *2-form $\omega \in \Omega^2(M) = \Gamma(M, T^*M)$. A **symplectic manifold** is a pair (M, ω) where ω is a symplectic form on M. If there exists a symplectic form on M we say that M has a symplectic structure or admits a symplectic structure.*

A map of symplectic manifolds, say $f : (M, \omega) \to (N, \varpi)$ is called a **symplectic map** if and only if $f^*\varpi = \omega$. We will reserve the term **symplectomorphism** to refer to diffeomorphisms that are symplectic maps. Notice that since a symplectic form such as ω is nondegenerate, the $2n$ form $\omega^n = \omega \wedge \cdots \wedge \omega$ is nonzero and global. Hence a symplectic manifold is orientable (more precisely, it is oriented).

Definition 28.3 *The form $\Omega_\omega = \frac{(-1)^n}{(2n)!}\omega^n$ is called the **canonical volume form** or **Liouville volume**.*

We immediately have that if $f : (M, \omega) \to (M, \omega)$ is a symplectic diffeomorphism then $f^*\Omega_\omega = \Omega_\omega$.

Not every manifold admits a symplectic structure. Of course if M does admit a symplectic structure then it must have even dimension but there are other more subtle obstructions. For example, the fact that $H^2(S^4) = 0$ can be used to show that S^4 does not admit ant symplectic structure. To see this, suppose to the contrary that ω is a closed nondegenerate 2-form on S^4. The since $H^2(S^4) = 0$ there would be a 1-form θ with $d\theta = \omega$. But then since $d(\omega \wedge \theta) = \omega \wedge \omega$ the 4-form $\omega \wedge \omega$ would be exact also and Stokes' theorem would give $\int_{S^4} \omega \wedge \omega = \int_{S^4} d(\omega \wedge \theta) = \int_{\partial S^4 = \emptyset} \omega \wedge \theta = 0$. But as we have seen $\omega^2 = \omega \wedge \omega$ is a nonzero top form so we must really have $\int_{S^4} \omega \wedge \omega \neq 0$. So in fact, S^4 does not admit a symplectic structure. We will give a more careful examination to the question of obstructions to symplectic structures but let us now list some positive examples.

Example 28.1 (surfaces) *Any orientable surface with volume form (area form) qualifies since in this case the volume ω itself is a closed nondegenerate two form.*

Example 28.2 (standard) *The form $\omega_{can} = \sum_{i=1}^n dx^i \wedge dx^{i+n}$ on $\mathbb{R}^{2n}$ is the prototypical symplectic form for the theory and makes $\mathbb{R}^n$ a symplectic manifold. (See Darboux's theorem 28.2 below)*

Example 28.3 (cotangent bundle) *We will see in detail below that the cotangent bundle of any smooth manifold has a natural symplectic structure. The symplectic form in a natural bundle chart (q, p) has the form $\omega = \sum_{i=1}^n dq^i \wedge dp_i$. (warning: some authors use $-\sum_{i=1}^n dq^i \wedge dp_i = \sum_{i=1}^n dp_i \wedge dq^i$ instead).*

Example 28.4 (complex submanifolds) *The symplectic $\mathbb{R}^{2n}$ may be considered the realification of $\mathbb{C}^n$ and then multiplication by i is thought of as a map $J : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$. We have that $\omega_{can}(v, Jv) = -|v|^2$ so that ω_{can} is nondegenerate on any complex submanifold M of $\mathbb{R}^{2n}$ and so $M, \omega_{can}|_M$ is a symplectic manifold.*

Example 28.5 (coadjoint orbit) *Let G be a Lie group. Define the coadjoint map $\mathrm{Ad}^\dagger : G \to GL(\mathfrak{g}^*)$, which takes g to $\mathrm{Ad}_g^\dagger$, by*

$$\mathrm{Ad}_g^\dagger(\xi)(x) = \xi(\mathrm{Ad}_{g^{-1}}(x)).$$

The action defined by $\mathrm{Ad}^\dagger$,

$$g \to g \cdot \xi = \mathrm{Ad}_g^\dagger(\xi),$$

is called the ***coadjoint action****. Then we have an induced map $\mathrm{ad}^\dagger : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g}^*)$ at the Lie algebra level;*

$$\mathrm{ad}^\dagger(x)(\xi)(y) = -\xi([x, y]).$$

The orbits of the action given by Ad^* are called coadjoint orbits and we will show in theorem below that each orbit is a symplectic manifold in a natural way.

28.4 Complex Structure and Kähler Manifolds

Recall that a complex manifold is a manifold modeled on $\mathbb{C}^n$ and such that the chart overlap functions are all biholomorphic. Every (real) tangent space T_pM of a complex manifold M has a complex structure $J_p : T_pM \to T_pM$ given in biholomorphic coordinates $z = x + iy$ by

$$J_p(\left.\frac{\partial}{\partial x^i}\right|_p) = \left.\frac{\partial}{\partial y^i}\right|_p$$
$$J_p(\left.\frac{\partial}{\partial y^i}\right|_p) = -\left.\frac{\partial}{\partial x^i}\right|_p$$

and for any (biholomorphic) overlap function $\Delta = \varphi \circ \psi^{-1}$ we have $T\Delta \circ J = J \circ T\Delta$.

Definition 28.4 *An **almost complex structure** on a smooth manifold M is a bundle map $J : TM \to TM$ covering the identity map such that $J^2 = -\mathrm{id}$. If one can choose an atlas for M such that all the coordinate change functions (overlap functions) Δ satisfy $T\Delta \circ J = J \circ T\Delta$ then J is called a **complex structure** on M.*

Definition 28.5 *An **almost symplectic structure** on a manifold M is a nondegenerate smooth 2-form ω that is not necessarily closed.*

Theorem 28.2 *A smooth manifold M admits an almost complex structure if and only if it admits an almost symplectic structure.*

Proof. First suppose that M has an almost complex structure J and let g be any Riemannian metric on M. Define a quadratic form q_p on each tangent space by

$$q_p(v) = \mathrm{g}_p(v, v) + \mathrm{g}_p(Jv, Jv).$$

Then we have $q_p(Jv) = q_p(v)$. Now let h be the metric obtained from the quadratic form q by polarization. It follows that $\mathrm{h}(v, w) = \mathrm{h}(Jv, Jw)$ for all $v, w \in TM$. Now define a two form ω by

$$\omega(v, w) = \mathrm{h}(v, Jw).$$

This really is skew-symmetric since $\omega(v, w) = \mathrm{h}(v, Jw) = \mathrm{h}(Jv, J^2w) = -\mathrm{h}(Jv, w) = \omega(w, v)$. Also, ω is nondegenerate since if $v \neq 0$ then $\omega(v, Jv) = \mathrm{h}(v, v) > 0$.

Conversely, let ω be a nondegenerate two form on a manifold M. Once again choose a Riemannian metric g for M. There must be a vector bundle map $\Omega : TM \to TM$ such that

$$\omega(v, w) = \mathrm{g}(\Omega v, w) \text{ for all } v, w \in TM.$$

Since ω is nondegenerate the map Ω must be invertible. Furthermore, since Ω is clearly anti-symmetric with respect to g the map $-\Omega \circ \Omega = -\Omega^2$ must be symmetric and positive definite. From linear algebra applied fiberwise we know that there must be a positive symmetric square root for $-\Omega^2$. Denote this by $P = \sqrt{-\Omega^2}$. Finite dimensional spectral theory also tell us that $P\Omega = \Omega P$. Now let $J = \Omega P^{-1}$ and notice that

$$J^2 = (\Omega P^{-1})(\Omega P^{-1}) = \Omega^2 P^{-2} = -\Omega^2 \Omega^{-2} = -\operatorname{id}.$$

■

One consequence of this result is that there must be characteristic class obstructions to the existence of a symplectic structure on a manifolds. In fact, if M, ω is a symplectic manifold then it is certainly almost symplectic and so there is an almost complex structure J on M. The tangent bundle is then a complex vector bundle with J giving the action of multiplication by $\sqrt{-1}$ on each fiber T_pM. Denote the resulting complex vector bundle by TM^J and then consider the total Chern class

$$c(TM^J) = c_n(TM^J) + ... + c_1(TM^J) + 1.$$

Here $c_i(TM^J) \in H^{2i}(M, \mathbb{Z})$. Recall that with the orientation given by ω^n the top class $c_n(TM^J)$ is the Euler class $e(TM)$ of TM. Now for the real bundle TM we have the total Pontrijagin class

$$p(TM) = p_n(TM) + ... + p_1(TM) + 1$$

which are related to the Chern classes by the Whitney sum

$$\begin{aligned} p(TM) &= c(TM^J) \oplus c(TM^{-J}) \\ &= (c_n(TM^J) + ... + c_1(TM^J) + 1)((-1)^n c_n(TM^J) - +... + c_1(TM^J) + 1) \end{aligned}$$

where TM^{-J} is the complex bundle with $-J$ giving the multiplication by $\sqrt{-1}$. We have used the fact that

$$c_i(TM^{-J}) = (-1)^i c_i(TM^J).$$

Now the classes $p_k(TM)$ are invariants of the diffeomorphism class of M an so can be considered constant over all possible choices of J. In fact, from the above relations one can deduce a quadratic relation that must be satisfied:

$$p_k(TM) = c_k(TM^J)^2 - 2c_{k-1}(TM^J)c_{k+1}(TM^J) + \cdots + (-1)^k 2c_{2k}(TM^J).$$

Now this places a restriction on what manifolds might have almost complex structures and hence a restriction on having an almost symplectic structure. Of course some manifolds might have an almost symplectic structure but still have no symplectic structure.

Definition 28.6 *A positive definite real bilinear form h on an almost complex manifold M, J is will be called Hermitian metric or J-metric if h is J invariant. In this case h is the real part of a Hermitian form on the complex vector bundle TM, J given by*

$$\langle v, w \rangle = h(v, w) + ih(Jv, w)$$

Definition 28.7 *A diffeomorphism $\phi : M, J, h \to M, J, h$ is called a Hermitian isometry if and only if $T\phi \circ J = J \circ T\phi$ and*

$$h(T\phi v, T\phi w) = h(v, w).$$

A group action $\rho : G \times M \to M$ is called a Hermitian action if $\rho(g, .)$ is a Hermitian isometry for all g. In this case, we have for every $p \in M$ a the representation $d\rho_p : H_p \to \mathrm{Aut}(T_pM, J_p)$ of the isotropy subgroup H_p given by

$$d\rho_p(g)v = T_p\rho_g \cdot v.$$

Definition 28.8 *Let M, J be a complex manifold and ω a symplectic structure on M. The manifold is called a* ***Kähler manifold*** *if $h(v, w) := \omega(v, Jw)$ is positive definite.*

Equivalently we can define a ***Kähler manifold as*** *a complex manifold M, J with Hermitian metric h with the property that the nondegenerate 2-form $\omega(v, w) := h(v, Jw)$ is closed.*

Thus we have the following for a Kähler manifold:

1. A complex structure J,
2. A J-invariant positive definite bilinear form b,
3. A Hermitian form $\langle v, w\rangle = h(v, w) + ih(Jv, w)$.
4. A symplectic form ω with the property that $\omega(v, w) = h(v, Jw)$.

Of course if M, J is a complex manifold with Hermitian metric h then $\omega(v, w) := h(v, Jw)$ automatically gives a nondegenerate 2-form; the question is whether it is closed or not. Mumford's criterion is useful for this purpose:

Theorem 28.3 (Mumford) *Let $\rho : G\times M \to M$ be a smooth Lie group action by Hermitian isometries. For $p \in M$ let H_p be the isometry subgroup of the point p. If $J_p \in d\rho_p(H_p)$ for every p then we have that ω defined by $\omega(v, w) := h(v, Jw)$ is closed.*

Proof. It is easy to see that since ρ preserves both h and J it also preserves ω and $d\omega$. Thus for any given $p \in M$, we have

$$d\omega(d\rho_p(g)u, d\rho_p(g)v, d\rho_p(g)w) = d\omega(u, v, w)$$

for all $g \in H_p$ and all $u, v, w \in T_pM$. By assumption there is a $g_p \in H_p$ with $J_p = d\rho_p(g_p)$. Thus with this choice the previous equation applied twice gives

$$\begin{aligned} d\omega(u, v, w) &= d\omega(J_pu, J_pv, J_pw) \\ &= d\omega(J_p^2u, J_p^2v, J_p^2w) \\ &= d\omega(-u, -v, -w) = -d\omega(u, v, w) \end{aligned}$$

so $d\omega = 0$ at p which was an arbitrary point so $d\omega = 0$. ∎

Since a Kähler manifold is a posteriori a Riemannian manifold it has associated with it the Levi-Civita connection ∇. In the following we view J as an element of $\mathfrak{X}(M)$.

Theorem 28.4 *For a Kähler manifold M, J, h with associated symplectic form ω we have that*

$$d\omega = 0 \quad \textit{if and only if} \quad \nabla J = 0.$$

28.5 Symplectic musical isomorphisms

Since a symplectic form ω on a manifold M is nondegenerate we have a map

$$\omega_\flat : TM \to T^*M$$

given by $\omega_\flat(X_p)(v_p) = \omega(X_p, v_p)$ and the inverse $\omega^\sharp$ is such that

$$\iota_{\omega^\sharp(\alpha)}\omega = \alpha$$

or

$$\omega(\omega^\sharp(\alpha_p), v_p) = \alpha_p(v_p)$$

Let check that $\omega^\sharp$ really is the inverse. (one could easily be off by a sign in this business.) We have

$$\omega_\flat(\omega^\sharp(\alpha_p))(v_p) = \omega(\omega^\sharp(\alpha_p), v_p) = \alpha_p(v_p) \text{ for all } v_p$$
$$\implies \omega_\flat(\omega^\sharp(\alpha_p)) = \alpha_p.$$

Notice that $\omega^\sharp$ induces a map on sections also denoted by $\omega^\sharp$ with inverse $\omega_\flat : \mathfrak{X}(M) \to \mathfrak{X}^*(M)$.

Notation 28.1 *Let us abbreviate $\omega^\sharp(\alpha)$ to $\sharp\alpha$ and $\omega_\flat(v)$ to $\flat v$.*

28.6 Darboux's Theorem

Lemma 28.2 (Darboux's theorem) *On a $2n$- manifold (M, ω) with a closed 2-form ω with $\omega^n \neq 0$ (for instance if (M, ω) is symplectic) there exists a sub-atlas consisting of charts called symplectic charts (canonical coordinates) characterized by the property that the expression for ω in such a chart is*

$$\omega_U = \sum_{i=1}^{n} dx^i \wedge dx^{i+n}$$

and so in particular M must have even dimension $2n$.

Remark 28.1 *Let us agree that the canonical coordinates can be written (x^i, y_i) instead of (x^i, x^{i+n}) when convenient.*

Remark 28.2 *It should be noticed that if x^i, y_i is a symplectic chart then $\sharp dx^i$ must be such that*

$$\sum_{r=1}^{n} dx^r \wedge dy^r(\sharp dx^i, \frac{\partial}{\partial x^j}) = \delta^i_j$$

but also

$$\sum_{r=1}^{n} dx^r \wedge dy^r(\sharp dx^i, \frac{\partial}{\partial x^j}) = \sum_{r=1}^{n} \left(dx^r(\sharp dx) dy^r(\frac{\partial}{\partial x^j}) - dy^r(\sharp dx^i) dx^r(\frac{\partial}{\partial x^j}) \right)$$
$$= -dy^j(\sharp dx^i)$$

and so we conclude that $\sharp dx^i = -\frac{\partial}{\partial y^i}$ *and similarly* $\sharp dy^i = \frac{\partial}{\partial x^i}$.

Proof. We will use induction and follow closely the presentation in [**?**]. Assume the theorem is true for symplectic manifolds of dimension $2(n-1)$. Let $p \in M$. Choose a function y^1 on some open neighborhood of p such that $dy_1(p) \neq 0$. Let $X = \sharp dy_1$ and then X will not vanish at p. We can then choose another function x^1 such that $Xx^1 = 1$ and we let $Y = -\sharp dx^1$. Now since $d\omega = 0$ we can use Cartan's formula to get

$$\mathcal{L}_X \omega = \mathcal{L}_Y \omega = 0.$$

In the following we use the notation $\langle X, \omega \rangle = \iota_X \omega$ (see notation 8.1). Contract ω with the bracket of X and Y :

$$\langle [X,Y], \omega \rangle = \langle \mathcal{L}_X Y, \omega \rangle = \mathcal{L}_X \langle Y, \omega \rangle - \langle Y, \mathcal{L}_X \omega \rangle$$
$$= \mathcal{L}_X(-dx^1) = -d(X(x^1)) = -d1 = 0.$$

Now since ω is nondegenerate this implies that $[X,Y] = 0$ and so there must be a local coordinate system $(x^1, y_1, w^1, ..., w^{2n-2})$ with

$$\frac{\partial}{\partial y_1} = Y$$
$$\frac{\partial}{\partial x^1} = X.$$

In particular, the theorem is true if $n = 1$. Assume the theorem is true for symplectic manifolds of dimension $2(n-1)$. If we let $\omega' = \omega - dx^1 \wedge dy_1$ then since $d\omega' = 0$ and hence

$$\langle X, \omega' \rangle = \mathcal{L}_X \omega' = \langle Y, \omega' \rangle = \mathcal{L}_Y \omega' = 0$$

we conclude that ω' can be expressed as a 2-form in the $w^1, ..., w^{2n-2}$ variables alone. Furthermore,

$$0 \neq \omega^n = (\omega - dx^1 \wedge dy_1)^n$$
$$= \pm n dx^1 \wedge dy_1 \wedge (\omega')^n$$

from which it follows that ω' is the pullback of a form nondegenerate form ϖ on $\mathbb{R}^{2n-2}$. To be exact if we let the coordinate chart given by $(x^1, y_1, w^1, ..., w^{2n-2})$ by denoted by ψ and let pr be the projection $\mathbb{R}^{2n} = \mathbb{R}^2 \times \mathbb{R}^{2n-1} \to \mathbb{R}^{2n-1}$ then $\omega' = (pr \circ \psi)^* \varpi$. Thus the induction hypothesis says that ω' has the form

$\omega' = \sum_{i=2}^{n} dx^i \wedge dy_i$ for some functions x^i, y_i with $i = 2, ..., n$. It is easy to see that the construction implies that in some neighborhood of p the full set of functions x^i, y_i with $i = 1, ..., n$ form the desired symplectic chart. ■

An atlas $\mathcal{A}$ of symplectic charts is called a symplectic atlas. A chart (U, φ) is called compatible with the symplectic atlas $\mathcal{A}$ if for every $(\psi_\alpha, U_\alpha) \in \mathcal{A}$ we have

$$(\varphi \circ \psi^{-1})^* \omega_0 = \omega_0$$

for the canonical symplectic $\omega_{can} = \sum_{i=1}^{n} du^i \wedge du^{i+n}$ defined on $\psi_\alpha(U \cap U_\alpha) \subset \mathbb{R}^{2n}$ using standard rectangular coordinates u^i.

28.7 Poisson Brackets and Hamiltonian vector fields

Definition 28.9 (on forms) *The **Poisson bracket** of two 1-forms is defined to be*

$$\{\alpha, \beta\}_\pm = \mp\flat[\sharp\alpha, \sharp\beta]$$

where the musical symbols refer to the maps $\omega^\sharp$ and $\omega_\flat$. This puts a Lie algebra structure on the space of 1-forms $\Omega^1(M) = \mathfrak{X}^*(M)$.

Definition 28.10 (on functions) *The **Poisson bracket** of two smooth functions is defined to be*

$$\{f, g\}_\pm = \pm\omega(\sharp df, \sharp dg) = \pm\omega(X_f, X_g)$$

This puts a Lie algebra structure on the space $\mathcal{F}(M)$ of smooth function on the symplectic M. It is easily seen (using $dg = \iota_{X_g}\omega$) that $\{f, g\}_\pm = \pm L_{X_g} f = \mp L_{X_f} g$ which shows that $f \mapsto \{f, g\}$ is a derivation for fixed g. The connection between the two Poisson brackets is

$$d\{f, g\}_\pm = \{df, dg\}_\pm.$$

Let us take canonical coordinates so that $\omega = \sum_{i=1}^{n} dx^i \wedge dy_i$. If $X_p = \sum_{i=1}^{n} dx^i(X)\frac{\partial}{\partial x^i} + \sum_{i=1}^{n} dy_i(X)\frac{\partial}{\partial y_i}$ and $v_p = dx^i(v_p)\frac{\partial}{\partial x^i} + dy_i(v_p)\frac{\partial}{\partial y_i}$ then using the Einstein summation convention we have

$$\begin{aligned}
&\omega_\flat(X)(v_p) \\
&= \omega(dx^i(X)\frac{\partial}{\partial x^i} + dy_i(X)\frac{\partial}{\partial y_i}, dx^i(v_p)\frac{\partial}{\partial x^i} + dy_i(v_p)\frac{\partial}{\partial y_i}) \\
&= (dx^i(X)dy_i - dy_i(X)dx^i)(v_p)
\end{aligned}$$

so we have

Lemma 28.3 $\omega_\flat(X_p) = \sum_{i=1}^{n} dx^i(X)dy_i - dy_i(X)dx^i = \sum_{i=1}^{n}(-dy_i(X)dx^i + dx^i(X)dy_i)$

Corollary 28.1 *If* $\alpha = \sum_{i=1}^{n} \alpha(\frac{\partial}{\partial x^i})dx^i + \sum_{i=1}^{n} \alpha(\frac{\partial}{\partial y_i})dy^i$ *then* $\omega^\sharp(\alpha) = \sum_{i=1}^{n} \alpha(\frac{\partial}{\partial y_i})\frac{\partial}{\partial x^i} - \sum_{i=1}^{n} \alpha(\frac{\partial}{\partial x^i})\frac{\partial}{\partial y_i}$

An now for the local formula:

Corollary 28.2 $\{f,g\} = \sum_{i=1}^{n}(\frac{\partial f}{\partial x^i}\frac{\partial g}{\partial y_i} - \frac{\partial f}{\partial y_i}\frac{\partial g}{\partial x^i})$

Proof. $df = \frac{\partial f}{\partial x^i}dx^i + \frac{\partial f}{\partial y_i}dy_i$ and $dg = \frac{\partial g}{\partial x^j}dx^j + \frac{\partial g}{\partial y_i}dy_i$ so $\sharp df = \frac{\partial f}{\partial y_i}\frac{\partial}{\partial x^i} - \frac{\partial f}{\partial x^i}\frac{\partial}{\partial y_i}$ and similarly for dg. Thus (using the summation convention again);

$$\begin{aligned}
\{f,g\} &= \omega(\sharp df, \sharp dg) \\
&= \omega(\frac{\partial f}{\partial y_i}\frac{\partial}{\partial x^i} - \frac{\partial f}{\partial x^i}\frac{\partial}{\partial y_i}, \frac{\partial g}{\partial y_i}\frac{\partial}{\partial x^j} - \frac{\partial g}{\partial x^j}\frac{\partial}{\partial y_i}) \\
&= \frac{\partial f}{\partial x^i}\frac{\partial g}{\partial y_i} - \frac{\partial f}{\partial y_i}\frac{\partial g}{\partial x^i}
\end{aligned}$$

■

A main point about Poison Brackets is

Theorem 28.5 *f is constant along the orbits of X_g if and only if $\{f,g\} = 0$. In fact,*

$$\tfrac{d}{dt} g \circ \mathrm{Fl}_t^{X_f} = 0 \iff \quad \{f,g\} = 0 \quad \iff \tfrac{d}{dt} f \circ \mathrm{Fl}_t^{X_g} = 0$$

Proof. $\frac{d}{dt} g \circ \mathrm{Fl}_t^{X_f} = (\mathrm{Fl}_t^{X_f})^* L_{X_f} g = (\mathrm{Fl}_t^{X_f})^*\{f,g\}$. Also use $\{f,g\} = -\{g,f\}$. ■

The equations of motion for a Hamiltonian H are

$$\frac{d}{dt} f \circ \mathrm{Fl}_t^{X_H} = \pm\{f \circ \mathrm{Fl}_t^{X_H}, H\}_\pm = \mp\{H, f \circ \mathrm{Fl}_t^{X_H}\}_\pm$$

which is true by the following simple computation

$$\begin{aligned}
\frac{d}{dt} f \circ \mathrm{Fl}_t^{X_H} &= \frac{d}{dt}(\mathrm{Fl}_t^{X_H})^* f = (\mathrm{Fl}_t^{X_H})^* L_{X_H} f \\
&= L_{X_H}(f \circ \mathrm{Fl}_t^{X_H}) = \{f \circ \mathrm{Fl}_t^{X_H}, H\}_\pm.
\end{aligned}$$

Notation 28.2 *¿From now on we will use only $\{.,.\}_+$ unless otherwise indicated and shall write $\{.,.\}$ for $\{.,.\}_+$.*

Definition 28.11 *A Hamiltonian system is a triple (M, ω, H) where M is a smooth manifold, ω is a symplectic form and H is a smooth function $H : M \to \mathbb{R}$.*

The main example, at least from the point of view of mechanics, is the cotangent bundle of a manifold which is discussed below. From a mechanical point of view the Hamiltonian function controls the dynamics and so is special.

Let us return to the general case of a symplectic manifold M, ω

Definition 28.12 *Now if $H : M \to \mathbb{R}$ is smooth then we define the* ***Hamiltonian vector field*** *X_H with energy function H to be $\omega^\sharp dH$ so that by definition $\iota_{X_H}\omega = dH$.*

Definition 28.13 *A vector field X on M, ω is called a* ***locally Hamiltonian vector field*** *or a* ***symplectic vector field*** *if and only if $L_X\omega = 0$.*

If a symplectic vector field is complete then we have that $(\mathrm{Fl}_t^X)^*\omega$ is defined for all $t \in \mathbb{R}$. Otherwise, for any relatively compact open set U the restriction Fl_t^X to U is well defined for all $t \leq b(U)$ for some number depending only on U. Thus $(\mathrm{Fl}_t^X)^*\omega$ is defined on U for $t \leq b(U)$. Since U can be chosen to contain any point of interest and since M can be covered by relatively compact sets, it will be of little harm to write $(\mathrm{Fl}_t^X)^*\omega$ even in the case that X is not complete.

Lemma 28.4 *The following are equivalent:*

1. *X is* ***symplectic vector field****, i.e. $L_X\omega = 0$*
2. *$\iota_X\omega$ is closed*
3. *$(\mathrm{Fl}_t^X)^*\omega = \omega$*
4. *X is locally a Hamiltonian vector field.*

Proof. (1)$\Longleftrightarrow$ (4) by the Poincaré lemma. Next, notice that $L_X\omega = d \circ \iota_X\omega + \iota_X \circ d\omega = d \circ \iota_X\omega$ so we have (2)$\Longleftrightarrow$(1). The implication (2)$\Longleftrightarrow$(3) follows from Theorem 7.7. ∎

Proposition 28.2 *We have the following easily deduced facts concerning Hamiltonian vector fields:*

1. The H is constant along integral curves of X_H
2. The flow of X_H is a local symplectomorphism. That is $\mathrm{Fl}_t^{X_H\,*}\omega = \omega$

Notation 28.3 *Denote the set of all Hamiltonian vector fields on M, ω by $\mathcal{H}(\omega)$ and the set of all symplectic vector fields by $\mathcal{SP}(\omega)$*

Proposition 28.3 *The set $\mathcal{SP}(\omega)$ is a Lie subalgebra of $\mathfrak{X}(M)$. In fact, we have $[\mathcal{SP}(\omega), \mathcal{SP}(\omega)] \subset \mathcal{H}(\omega) \subset \mathfrak{X}(M)$.*

Proof. Let $X, Y \in \mathcal{SP}(\omega)$. Then

$$\begin{aligned}
[X,Y]\lrcorner\omega &= \mathcal{L}_X Y\lrcorner\omega = \mathcal{L}_X(Y\lrcorner\omega) - Y\lrcorner\mathcal{L}_X\omega \\
&= d(X\lrcorner Y\lrcorner\omega) + X\lrcorner d(Y\lrcorner\omega) - 0 \\
&= d(X\lrcorner Y\lrcorner\omega) + 0 + 0 \\
&= -d(\omega(X,Y)) = -X_{\omega(X,Y)}\lrcorner\omega
\end{aligned}$$

and since ω in nondegenerate we have $[X,Y] = X_{-\omega(X,Y)} \in \mathcal{H}(\omega)$. ∎

28.8 Configuration space and Phase space

Consider the cotangent bundle of a manifold Q with projection map

$$\pi : T^*Q \to Q$$

and define the **canonical 1-form** $\theta \in T^*(T^*Q)$ by

$$\theta : v_{\alpha_p} \mapsto \alpha_p(T\pi \cdot v_{\alpha_p})$$

where $\alpha_p \in T_p^*Q$ and $v_{\alpha_p} \in T_{\alpha_p}(T_p^*Q)$. In local coordinates this reads

$$\theta_0 = \sum p_i dq^i.$$

Then $\omega_{T^*Q} = -d\theta$ is a symplectic form that in natural coordinates reads

$$\omega_{T^*Q} = \sum dq^i \wedge dp_i$$

Lemma 28.5 *θ is the unique 1-form such that for any $\beta \in \Omega^1(Q)$ we have*

$$\beta^*\theta = \beta$$

where we view β as $\beta : Q \to T^*Q$.

Proof: $\beta^*\theta(v_q) = \theta|_{\beta(q)}(T\beta \cdot v_q) = \beta(q)(T\pi \circ T\beta \cdot v_q) = \beta(q)(v_q)$ since $T\pi \circ T\beta = T(\pi \circ \beta) = T(\mathrm{id}) = \mathrm{id}$.

The cotangent lift T^*f of a diffeomorphism $f : Q_1 \to Q_2$ is defined by the commutative diagram

$$\begin{array}{ccc} T^*Q_1 & \overset{T^*f}{\longleftarrow} & T^*Q_2 \\ \downarrow & & \downarrow \\ Q_1 & \overset{f}{\to} & Q_2 \end{array}$$

and is a symplectic map; i.e. $(T^*f)^*\omega_0 = \omega_0$. In fact, we even have $(T^*f)^*\theta_0 = \theta_0$.

The triple (T^*Q, ω_{T^*Q}, H) is a Hamiltonian system for any choice of smooth function. The most common form for H in this case is $\frac{1}{2}K + V$ where K is a Riemannian metric that is constructed using the mass distribution of the bodies modelled by the system and V is a smooth potential function which, in a conservative system, depends only on $\mathbf{q}$ when viewed in natural cotangent bundle coordinates q^i, p_i.

Now we have $\sharp dg = \frac{\partial g}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial g}{\partial q^i}\frac{\partial}{\partial p_i}$ and introducing the $\pm$ notation one more time we have

$$\begin{aligned}
\{f, g\}_{\pm} &= \pm\omega_{T^*Q}(\sharp df, \sharp dg) = \pm df(\sharp dg) = \pm df(\frac{\partial g}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial g}{\partial q^i}\frac{\partial}{\partial p_i}) \\
&= \pm(\frac{\partial g}{\partial p_i}\frac{\partial f}{\partial q^i} - \frac{\partial g}{\partial q^i}\frac{\partial f}{\partial p_i}) \\
&= \pm(\frac{\partial f}{\partial q^i}\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q^i})
\end{aligned}$$

Thus letting

$$\mathrm{Fl}_t^{X_H}(q_0^1,...,q_0^n,p_0^1,...,p_0^n) = (q^1(t),...,q^n(t),p^1(t),...,p^n(t))$$

the equations of motions read

$$\begin{aligned}\frac{d}{dt}f(q(t),p(t)) &= \frac{d}{dt}f\circ \mathrm{Fl}_t^{X_H} = \{f\circ \mathrm{Fl}_t^{X_H}, H\} \\ &= \frac{\partial f}{\partial q^i}\frac{\partial H}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial H}{\partial q^i}.\end{aligned}$$

Where we have abbreviated $f\circ \mathrm{Fl}_t^{X_H}$ to just f. In particular, if $f = q^i$ and $f = p_i$ then

$$\begin{aligned}\dot{q}^i(t) &= \frac{\partial H}{\partial p_i} \\ \dot{p}_i(t) &= -\frac{\partial H}{\partial q^i}\end{aligned}$$

which should be familiar.

28.9 Transfer of symplectic structure to the Tangent bundle

Case I: a (pseudo) Riemannian manifold

If Q, g is a (pseudo) Riemannian manifold then we have a map $\mathrm{g}^\flat : TQ \to T^*Q$ defined by

$$\mathrm{g}^\flat(v)(w) = \mathrm{g}(v,w)$$

and using this we can define a symplectic form ϖ_0 on TQ by

$$\varpi_0 = \left(\mathrm{g}^\flat\right)^*\omega$$

(Note that $d\varpi_0 = d(\mathrm{g}^{\flat *}\omega) = \mathrm{g}^{\flat *}d\omega = 0$.) In fact, ϖ_0 is exact since ω is exact:

$$\begin{aligned}\varpi_0 &= \left(\mathrm{g}^\flat\right)^*\omega \\ &= \left(\mathrm{g}^\flat\right)^* d\theta = d\left(\mathrm{g}^{\flat *}\theta\right).\end{aligned}$$

Let us write $\Theta_0 = \mathrm{g}^{\flat *}\theta$. Locally we have

$$\begin{aligned}\Theta_0(x,v)(v_1,v_2) &= \mathrm{g}_x(v,v_1) \text{ or} \\ \Theta_0 &= \sum \mathrm{g}_{ij}\dot{q}^i dq^j\end{aligned}$$

and also

$$\begin{aligned}&\varpi_0(x,v)((v_1,v_2),((w_1,w_2)))\\&=\mathrm{g}_x(w_2,v_1)-\mathrm{g}_x(v_2,w_1)+D_x\mathrm{g}_x(v,v_1)\cdot w_1-D_x\mathrm{g}_x(v,w_1)\cdot v_1\end{aligned}$$

which in classical notation (and for finite dimensions) looks like

$$\varpi_h=\mathrm{g}_{ij}dq^i\wedge d\dot{q}^j+\sum\frac{\partial \mathrm{g}_{ij}}{\partial q^k}\dot{q}^i dq^j\wedge dq^k$$

Case II: Transfer of symplectic structure by a Lagrangian function.

Definition 28.14 *Let $L:TQ\to Q$ be a Lagrangian on a manifold Q. We say that L is **regular** or **non-degenerate** at $\xi\in TQ$ if in any canonical coordinate system $(q,\dot{q})$ whose domain contains ξ, the matrix*

$$\left[\frac{\partial^2 L}{\partial\dot{q}^i\partial\dot{q}^j}(q(\xi),\dot{q}(\xi))\right]$$

*is non-degenerate. L is called **regular** or **nondegenerate** if it is regular at all points in TQ.*

We will need the following general concept:

Definition 28.15 *Let $\pi_E:E\to M$ and $\pi_F:F\to M$ be two vector bundles. A map $L:E\to F$ is called a **fiber preserving map** if the following diagram commutes:*

$$\begin{array}{ccccc}E & & \stackrel{L}{\to} & & F\\ \pi_E & \searrow & & \swarrow & \pi_F\\ & & M & & \end{array}$$

We do not require that the map L be linear on the fibers and so in general L is not a vector bundle morphism.

Definition 28.16 *If $L:E\to F$ is a fiber preserving map then if we denote the restriction of L to a fiber E_p by L_p define the **fiber derivative***

$$\mathbf{F}L:E\to\mathrm{Hom}(E,F)$$

by $\mathbf{F}L:e_p\mapsto Df|_p(e_p)$ for $e_p\in E_p$.

In our application of this concept, we take F to be the trivial bundle $Q\times\mathbb{R}$ over Q so $\mathrm{Hom}(E,F)=\mathrm{Hom}(E,\mathbb{R})=T^*Q$.

Lemma 28.6 *A Lagrangian function $L:TQ\to\mathbb{R}$ gives rise to a fiber derivative $\mathbf{F}L:TQ\to T^*Q$. The Lagrangian is nondegenerate if and only if $\mathbf{F}L$ is a diffeomorphism.*

Definition 28.17 *The form ϖ_L is defined by*

$$\varpi_L = (\mathbf{F}L)^* \omega$$

Lemma 28.7 *ω_L is a symplectic form on TQ if and only if L is nondegenerate (i.e. if $\mathbf{F}L$ is a diffeomorphism).*

Observe that we can also define $\theta_L = (\mathbf{F}L)^* \theta$ so that $d\theta_L = d(\mathbf{F}L)^* \theta = (\mathbf{F}L)^* d\theta = (\mathbf{F}L)^* \omega = \varpi_L$ so we see that ω_L is exact (and hence closed a required for a symplectic form).

Now in natural coordinates we have

$$\varpi_L = \frac{\partial^2 L}{\partial \dot{q}^i \partial q^j} dq^i \wedge dq^j + \frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j} dq^i \wedge d\dot{q}^j$$

as can be verified using direct calculation.

The following connection between the transferred forms ϖ_L and ϖ_0 and occasionally not pointed out in some texts.

Theorem 28.6 *Let V be a smooth function on a Riemannian manifold M, h. If we define a Lagrangian by $L = \frac{1}{2}h - V$ then the Legendre transformation $\mathbf{F}L :: TQ \to T^*Q$ is just the map $g^\flat$ and hence $\varpi_L = \varpi_h$.*

Proof. We work locally. Then the Legendre transformation is given by

$$\begin{array}{l} q^i \mapsto q^i \\ \dot{q}^i \mapsto \frac{\partial L}{\partial \dot{q}^i} \end{array} .$$

But since $L(\dot{\mathbf{q}}, \dot{\mathbf{q}}) = \frac{1}{2}\mathrm{g}(\dot{\mathbf{q}}, \dot{\mathbf{q}}) - V(q)$ we have $\frac{\partial L}{\partial \dot{q}^i} = \frac{\partial}{\partial \dot{q}^i} \frac{1}{2}\mathrm{g}_{kl}\dot{q}^l \dot{q}^k = \mathrm{g}_{il}\dot{q}^l$ which together with $q^i \mapsto q^i$ is the coordinate expression for $\mathrm{g}^\flat$:

$$\begin{array}{l} q^i \mapsto q^i \\ \dot{q}^i \mapsto \mathrm{g}_{il}\dot{q}^l \end{array}$$

■

28.10 Coadjoint Orbits

Let G be a Lie group and consider $\mathrm{Ad}^\dagger : G \to GL(\mathfrak{g}^*)$ and the corresponding coadjoint action as in example 28.5. For every $\xi \in \mathfrak{g}^*$ we have a Left invariant 1-form on G defined by

$$\theta^\xi = \xi \circ \omega_G$$

where ω_G is the canonical $\mathfrak{g}$-valued 1-form (the Maurer-Cartan form). Let the G_ξ be the isotropy subgroup of G for a point $\xi \in \mathfrak{g}^*$ under the coadjoint action. Then it is standard that orbit $G \cdot \xi$ is canonically diffeomorphic to the orbit space G/G_ξ and the map $\phi_\xi : g \mapsto g \cdot \xi$ is a submersion onto . Then we have

Theorem 28.7 *There is a unique symplectic form* Ω^{ξ} *on* $G/G_{\xi} \cong G \cdot \xi$ *such that* $\phi_{\xi}^{*}\Omega^{\xi} = d\theta^{\xi}$.

Proof: If such a form as Ω^{ξ} exists as stated then we must have

$$\Omega^{\xi}(T\phi_{\xi}.v, T\phi_{\xi}.w) = d\theta^{\xi}(v,w) \text{ for all } v,w \in T_gG$$

We will show that this in fact *defines* Ω^{ξ} as a symplectic form on the orbit $G\cdot\xi$. First of all notice that by the structure equations for the Maurer-Cartan form we have for $v,w \in T_eG = \mathfrak{g}$

$$\begin{aligned} d\theta^{\xi}(v,w) &= \xi(d\omega_G(v,w)) = \xi(\omega_G([v,w])) \\ &= \xi(-[v,w]) = \mathrm{ad}^{\dagger}(v)(\xi)(w) \end{aligned}$$

¿From this we see that

$$\mathrm{ad}^{\dagger}(v)(\xi) = 0 \Longleftrightarrow v \in \mathrm{Null}(\left. d\theta^{\xi}\right|_e)$$

where $\mathrm{Null}(\left. d\theta^{\xi}\right|_e) = \{v \in \mathfrak{g} : \left. d\theta^{\xi}\right|_e(v,w) \text{ for all } w \in \mathfrak{g}\}$. On the other hand, $G_{\xi} = \ker\{g \longmapsto \mathrm{Ad}_g^{\dagger}(\xi)\}$ so $\mathrm{ad}^{\dagger}(v)(\xi) = 0$ if and only if $v \in T_eG_{\xi} = \mathfrak{g}_{\xi}$.

Now notice that since $d\theta^{\xi}$ is left invariant we have that $\mathrm{Null}(\left. d\theta^{\xi}\right|_g) = TL_g(\mathfrak{g}_{\xi})$ which is the tangent space to the coset gG_{ξ} and which is also $\ker \left. T\phi_{\xi}\right|_g$. Thus we conclude that

$$\mathrm{Null}(\left. d\theta^{\xi}\right|_g) = \ker \left. T\phi_{\xi}\right|_g.$$

It follows that we have a natural isomorphism

$$T_{g\cdot\xi}(G\cdot\xi) = \left. T\phi_{\xi}\right|_g(T_gG) \approx T_gG/(TL_g(\mathfrak{g}_{\xi}))$$

Another view: Let the vector field on $G\cdot\xi$ corresponding to $v,w \in \mathfrak{g}$ generated by the action be denoted by $v^{\dagger}$and $w^{\dagger}$. Then we have $\Omega^{\xi}(\xi)(v^{\dagger}, w^{\dagger}) := \xi(-[v,w])$ at $\xi \in G\cdot\xi$ and then extend to the rest of the points of the orbit by equivariance:

$$\Omega^{\xi}(g\cdot\xi)(v^{\dagger}, w^{\dagger}) = \overset{\dagger}{\underset{g}{\mathrm{Ad}}}(\xi(-[v,w]))$$

28.11 The Rigid Body

In what follows we will describe the rigid body rotating about one of its points in three different versions. The basic idea is that we can represent the configuration space as a subset of $\mathbb{R}^{3N}$ with a very natural kinetic energy function. But this space is also isomorphic to the rotation group $SO(3)$ and we can transfer the kinetic energy metric over to $SO(3)$ and then the evolution of the system is given by geodesics in $SO(3)$ with respect to this metric. Next we take advantage of the fact that the tangent bundle of $SO(3)$ is trivial to transfer the setup over to a trivial bundle. But there are two natural ways to do this and we explore the relation between the two.

28.11.1 The configuration in $\mathbb{R}^{3N}$

Let us consider a rigid body to consist of a set of point masses located in $\mathbb{R}^3$ at points with position vectors $\mathbf{r}_1(t), ...\mathbf{r}_N(t)$ at time t. Thus $\mathbf{r}_i = (x_1, x_2, x_3)$ is the coordinates of the i-th point mass. Let $m_1, ..., m_N$ denote the masses of the particles. To say that this set of point masses is rigid is to say that the distances $|\mathbf{r}_i - \mathbf{r}_j|$ are constant for each choice of i and j. Let us assume for simplicity that the body is in a state of uniform rectilinear motion so that by re-choosing our coordinate axes if necessary we can assume that the there is one of the point masses at the origin of our coordinate system at all times. Now the set of all possible configurations is some submanifold of $\mathbb{R}^{3N}$ which we denote by M. Let us also assume that at least 3 of the masses, say those located at $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_2$ are situated so that the position vectors $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_2$ form a basis of $\mathbb{R}^3$. For convenience let $\mathbf{r}$ and $\dot{\mathbf{r}}$ be abbreviations for $(\mathbf{r}_1(t), ..., \mathbf{r}_N(t))$ and $(\dot{\mathbf{r}}_1(t), ..., \dot{\mathbf{r}}_N(t))$.The correct kinetic energy for the system of particles forming the rigid body is $\frac{1}{2}K(\dot{\mathbf{r}}, \dot{\mathbf{r}})$ where the kinetic energy metric K is

$$K(\mathbf{v}, \mathbf{w}) = m_1\mathbf{v}_1 \cdot \mathbf{w}_1 + \cdots + m_N\mathbf{v}_N \cdot \mathbf{w}_N.$$

Since there are no other forces on the body other than those that constrain the body to be rigid the Lagrangian for M is just $\frac{1}{2}K(\dot{\mathbf{r}}, \dot{\mathbf{r}})$ and the evolution of the point in M representing the body is a geodesic when we use as Hamiltonian K and the symplectic form pulled over from T^*M as described previously.

28.11.2 Modelling the rigid body on $S\mathrm{O}(3)$

Let $\mathbf{r}_1(0), ...\mathbf{r}(0)_N$ denote the initial positions of our point masses. Under these condition there is a unique matrix valued function $g(t)$ with values in $S\mathrm{O}(3)$ such that $\mathbf{r}_i(t) = g(t)\mathbf{r}_i(0)$. Thus the motion of the body is determined by the curve in $S\mathrm{O}(3)$ given by $t \mapsto g(t)$. In fact, we can map $S\mathrm{O}(3)$ to the set of all possible configurations of the points making up the body in a 1-1 manner by letting $\mathbf{r}_1(0) = \xi_1, ...\mathbf{r}(0)_N = \xi_N$ and mapping $\Phi : g \mapsto (g\xi_1, ..., g\xi_N) \in M \subset \mathbb{R}^{3N}$. If we use the map Φ to transfer this over to $TS\mathrm{O}(3)$ we get

$$k(\xi, \upsilon) = K(T\Phi \cdot \xi, T\Phi \cdot \upsilon)$$

for $\xi, \upsilon \in TS\mathrm{O}(3)$. Now k is a Riemannian metric on $S\mathrm{O}(3)$ and in fact, k is a left invariant metric:

$$k(\xi, \upsilon) = k(TL_g\xi, TL_g\upsilon) \text{ for all } \xi, \upsilon \in TS\mathrm{O}(3).$$

Exercise 28.2 *Show that* k *really is left invariant. Hint: Consider the map* $\mu_{g_0} : (\mathbf{v_1}, \cdots, \mathbf{v_N}) \mapsto (g_0\mathbf{v_1}, \cdots, g_0\mathbf{v_N})$ *for* $g_0 \in \mathrm{SO}(3)$ *and notice that* $\mu_{g_0} \circ \Phi = \Phi \circ L_{g_0}$ *and hence* $T\mu_{g_0} \circ T\Phi = T\Phi \circ TL_{g_0}$.

Now by construction, the Riemannian manifolds M, K and $S\mathrm{O}(3), \mathrm{k}$ are isometric. Thus the corresponding path $g(t)$ in $S\mathrm{O}(3)$ is a geodesic with respect to the left invariant metric k. Our Hamiltonian system is now $(TS\mathrm{O}(3), \Omega_k, k)$ where Ω_k is the Legendre transformation of the canonical symplectic form Ω on $T^*S\mathrm{O}(3)$

28.11.3 The trivial bundle picture

Recall that we the Lie algebra of SO(3) is the vector space of skew-symmetric matrices $\mathfrak{su}(3)$. We have the two trivializations of the tangent bundle TSO(3) given by

$$\mathrm{triv}_L(v_g) = (g, \omega_G(v_g)) = (g, g^{-1}v_g)$$
$$\mathrm{triv}_R(v_g) = (g, \omega_G(v_g)) = (g, v_g g^{-1})$$

with inverse maps $\mathrm{SO}(3) \times \mathfrak{so}(3) \to T\mathrm{SO}(3)$ given by

$$(g, B) \mapsto TL_g B$$
$$(g, B) \mapsto TR_g B$$

Now we should be able to represent the system in the trivial bundle $\mathrm{SO}(3) \times \mathfrak{so}(3)$ via the map $\mathrm{triv}_L(v_g) = (g, \omega_G(v_g)) = (g, g^{-1}v_g)$. Thus we let k_0 be the metric on $\mathrm{SO}(3) \times \mathfrak{so}(3)$ coming from the metric k. Thus by definition

$$\mathrm{k}_0((g, v), (g, w)) = \mathrm{k}(TL_g v, TL_g w) = \mathrm{k}_e(v, w)$$

where $v, w \in \mathfrak{so}(3)$ are skew-symmetric matrices.

28.12 The momentum map and Hamiltonian actions

Remark 28.3 *In this section all Lie groups will be assumed to be connected.*

Suppose that (a connected Lie group) G acts on M, ω as a group of symplectomorphisms.

$$\sigma : G \times M \to M$$

Then we say that σ is a **symplectic G-action** . Since G acts on M we have for every $v \in \mathfrak{g}$ the fundamental vector field $X^v = v^\sigma$. The fundamental vector field will be symplectic (locally Hamiltonian). Thus every one-parameter group g^t of G induces a symplectic vector field on M. Actually, it is only the infinitesimal action that matters at first so we define

Definition 28.18 *Let M be a smooth manifold and let $\mathfrak{g}$ be the Lie algebra of a connected Lie group G. A linear map $\sigma' : v \mapsto X^v$ from $\mathfrak{g}$ into $\mathfrak{X}(M)$ is called a **$\mathfrak{g}$-action** if*

$$[X^v, X^w] = -X^{[v,w]} \text{ or}$$
$$[\sigma'(v), \sigma'(w)] = -\sigma'([v, w]).$$

*If M, ω is symplectic and the $\mathfrak{g}$-action is such that $\mathcal{L}_{X^v}\omega = 0$ for all $v \in \mathfrak{g}$ we say that the action is a **symplectic g-action**.*

Definition 28.19 *Every symplectic action $\sigma : G \times M \to M$ induces a $\mathfrak{g}$-action $d\sigma$ via*

$$d\sigma : v \mapsto X^v$$

$$\text{where } X^v(x) = \left.\frac{d}{dt}\right|_0 \sigma(\exp(tv), x).$$

In some cases, we may be able to show that for all v the symplectic field X^v is a full fledged Hamiltonian vector field. In this case associated to each $v \in \mathfrak{g}$ there is a Hamiltonian function $J_v = J_{X^v}$ with corresponding Hamiltonian vector field equal to X^v and J_v is determined up to a constant by $X^v = \sharp dJ_{X^v}$. Now $\iota_{X^v}\omega$ is always closed since $d\iota_{X^v}\omega = \mathcal{L}_{X^v}\omega$. When is it possible to define J_v for every $v \in \mathfrak{g}$?

Lemma 28.8 *Given a symplectic $\mathfrak{g}$-action $\sigma' : v \mapsto X^v$ as above, there is a linear map $v \mapsto J_v$ such that $X^v = \sharp dJ_v$ for every $v \in \mathfrak{g}$ if and only if $\iota_{X^v}\omega$ is exact for all $v \in \mathfrak{g}$.*

Proof. If $H_v = H_{X^v}$ exists for all v then $dJ_{X^v} = \omega(X^v, .) = \iota_{X^v}\omega$ for all v so $\iota_{X^v}\omega$ is exact for all $v \in \mathfrak{g}$. Conversely, if for every $v \in \mathfrak{g}$ there is a smooth function h_v with $dh_v = \iota_{X^v}\omega$ then $X^v = \sharp dh_v$ so h_v is Hamiltonian for X^v. Now let $v_1, ..., v_n$ be a basis for $\mathfrak{g}$ and define $J_{v_i} = h_{v_i}$ and extend linearly. ■

Notice that the property that $v \mapsto J_v$ is linear means that we can define a map $J : M \to \mathfrak{g}^*$ by

$$J(x)(v) = J_v(x)$$

and this is called a **momentum map** .

Definition 28.20 *A symplectic G-action σ (resp. $\mathfrak{g}$-action σ') on M such that for every $v \in \mathfrak{g}$ the vector field X^v is a Hamiltonian vector field on M is called a **Hamiltonian G-action** (resp. Hamiltonian $\mathfrak{g}$-action).*

We can thus associate to every Hamiltonian action at least one momentum map-this being unique up to an additive constant.

Example 28.6 *If G acts on a manifold Q by diffeomorphisms then G lifts to an action on the cotangent bundle T^*M which is automatically symplectic. In fact, because $\omega_0 = d\theta_0$ is exact the action is also a Hamiltonian action. The Hamiltonian function associated to an element $v \in \mathfrak{g}$ is given by*

$$J_v(x) = \theta_0\left(\left.\frac{d}{dt}\right|_0 \exp(tv) \cdot x\right).$$

Definition 28.21 *If G (resp. $\mathfrak{g}$) acts on M in a symplectic manner as above such that the action is Hamiltonian and such that we may choose a momentum map J such that*

$$J_{[v,w]} = \{J_v, J_w\}$$

*where $J_v(x) = J(x)(v)$ then we say that the action is a **strongly Hamiltonian** G-action (resp. $\mathfrak{g}$-action).*

Example 28.7 *The action of example 28.6 is strongly Hamiltonian.*

We would like to have a way to measure of whether a Hamiltonian action is strong or not. Essentially we are just going to be using the difference $J_{[v,w]} - \{J_v, J_w\}$ but it will be convenient to introduce another view which we postpone until the next section where we study "Poisson structures".

PUT IN THEOREM ABOUT MOMENTUM CONSERVATION!!!!

What is a momentum map in the cotangent case? Pick a fixed point $\alpha \in T^*Q$ and consider the map $\Phi_\alpha : G \to T^*Q$ given by $\Phi_\alpha(g) = g \cdot \alpha = g^{-1*}\alpha$. Now consider the pullback of the canonical 1-form $\Phi_\alpha^*\theta_0$.

Lemma 28.9 *The restriction* $\Phi_\alpha^*\theta_0|_{\mathfrak{g}}$ *is an element of* $\mathfrak{g}^*$ *and the map* $\alpha \mapsto \Phi_\alpha^*\theta_0|_{\mathfrak{g}}$ *is the momentum map.*

Proof. We must show that $\Phi_\alpha^*\theta_0|_{\mathfrak{g}}(v) = H_v(\alpha)$ for all $v \in \mathfrak{g}$. Does $\Phi_\alpha^*\theta_0|_{\mathfrak{g}}(v)$ live in the right place? Let $g_v^t = \exp(vt)$. Then

$$\begin{aligned}(T_e\Phi_\alpha v) &= \left.\frac{d}{dt}\right|_0 \Phi_\alpha(\exp(vt)) \\ &= \left.\frac{d}{dt}\right|_0 (\exp(-vt))^*\alpha \\ &\left.\frac{d}{dt}\right|_0 \exp(vt)\cdot\alpha\end{aligned}$$

We have

$$\begin{aligned}\Phi_\alpha^*\theta_0|_{\mathfrak{g}}(v) &= \theta_0|_{\mathfrak{g}}(T_e\Phi_\alpha v) \\ = \theta_0(\left.\frac{d}{dt}\right|_0 &\exp(vt)\cdot\alpha) = J_v(\alpha)\end{aligned}$$

∎

Definition 28.22 *Let G act on a symplectic manifold M, ω and suppose that the action is Hamiltonian. A momentum map J for the action is said to be equivariant with respect to the coadjoint action if $J(g \cdot x) = \mathrm{Ad}_{g^{-1}}^* J(x)$.*

Chapter 29

Poisson Geometry

Life is good for only two things, discovering mathematics and teaching mathematics
–Siméon Poisson

29.1 Poisson Manifolds

In this chapter we generalize our study of symplectic geometry by approaching things from the side of a Poisson bracket.

Definition 29.1 *A* ***Poisson structure*** *on an associative algebra $\mathcal{A}$ is a Lie algebra structure with bracket denoted by $\{.,.\}$ such for a fixed $a \in \mathcal{A}$ that the map $x \mapsto \{a, x\}$ is a derivation of the algebra. An associative algebra with a Poisson structure is called a* ***Poisson algebra*** *and the bracket is called a* ***Poisson bracket*** .

We have already seen an example of a Poisson structure on the algebra $\mathfrak{F}(M)$ of smooth functions on a symplectic manifold. Namely,

$$\{f, g\} = \omega(\omega^\sharp df, \omega^\sharp dg).$$

By the Darboux theorem we know that we can choose local coordinates $(q^1, ..., q^n, p_1, ..., p_n)$ on a neighborhood of any given point in the manifold. Recall also that in such coordinates we have

$$\omega^\sharp df = \sum_{i=1}^{n} \frac{\partial f}{\partial p_i} \frac{\partial}{\partial q^i} - \sum_{i=1}^{n} \frac{\partial f}{\partial q^i} \frac{\partial}{\partial p_i}$$

sometimes called the symplectic gradient. It follows that

$$\sum_{i=1}^{n} (\frac{\partial f}{\partial q^i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q^i})$$

Definition 29.2 *A smooth manifold with a Poisson structure on is algebra of smooth functions is called a* ***Poisson manifold****.*

So every symplectic n-manifold gives rise to a Poisson structure. On the other hand, there are Poisson manifolds that are not so by virtue of being a symplectic manifold.

Now if our manifold is finite dimensional then every derivation of $\mathfrak{F}(M)$ is given by a vector field and since $g \mapsto \{f, g\}$ is a derivation there is a corresponding vector field X_f. Since the bracket is determined by these vector field and since vector fields can be defined locally (recall the presheaf $\mathfrak{X}_M$) we see that a Poisson structure is also a locally defined structure. In fact, $U \mapsto \mathfrak{F}_M(U)$ is a presheaf of Poisson algebras.

Now if we consider the map $w : \mathfrak{F}_M \to \mathfrak{X}_M$ defined by $\{f, g\} = w(f) \cdot g$ we see that $\{f, g\} = w(f) \cdot g = -w(g) \cdot f$ and so $\{f, g\}(p)$ depends only on the differentials df, dg of f and g. Thus we have a tensor $B(.,.) \in \Gamma \bigwedge^2 TM$ such that $B(df, dg) = \{f, g\}$. In other words, $B_p(.,.)$ is a symmetric bilinear map $T_p^*M \times T_p^*M \to \mathbb{R}$. Now any such tensor gives a bundle map $B^\sharp : T^*M \mapsto T^{**}M = TM$ by the rule $B^\sharp(\alpha)(\beta) = B(\beta, \alpha)$ for $\beta, \alpha \in T_p^*M$ and any $p \in M$. In other words, $B(\beta, \alpha) = \beta(B^\sharp(\alpha))$ for all $\beta \in T_p^*M$ and arbitrary $p \in M$. The 2-vector B is called the Poisson tensor for the given Poisson structure. B is also sometimes called a cosymplectic **structure** for reasons that we will now explain.

If M, ω is a symplectic manifold then the map $\omega_\flat : TM \to T^*M$ can be inverted to give a map $\omega^\sharp : T^*M \to TM$ and then a form $W \in \bigwedge^2 TM$ defined by $\omega^\sharp(\alpha)(\beta) = W(\beta, \alpha)$ (here again β, α must be in the same fiber). Now this form can be used to define a Poisson bracket by setting $\{f, g\} = W(df, dg)$ and so W is the corresponding Poisson tensor. But notice that

$$\begin{aligned}\{f, g\} &= W(df, dg) = \omega^\sharp(dg)(df) = df(\omega^\sharp(dg)) \\ &= \omega(\omega^\sharp df, \omega^\sharp dg)\end{aligned}$$

which is just the original Poisson bracket defined in the symplectic manifold M, ω.

Given a Poisson manifold $M, \{.,.\}$ we can always define $\{.,.\}_-$ by $\{f, g\}_- = \{g, f\}$. Since we some times refer to a Poisson manifold $M, \{.,.\}$ by referring just to the space we will denote M with the opposite Poisson structure by M^-.

A Poisson map is map $\phi : M, \{.,.\}_1 \to N, \{.,.\}_2$ is a smooth map such that $\phi^*\{f, g\} = \{\phi^* f, \phi^* g\}$ for all $f, g \in \mathfrak{F}(M)$.

For any subset S of a Poisson manifold let S_0 be the set of functions from $\mathfrak{F}(M)$ that vanish on S. A submanifold S of a Poisson manifold $M, \{.,.\}$ is called **coisotropic** if S_0 closed under the Poisson bracket. A Poisson manifold is called symplectic if the Poisson tensor B is non-degenerate since in this case we can use $B^\sharp$ to define a symplectic form on M. A Poisson manifold admits a (singular) foliation such that the leaves are symplectic. By a theorem of A. Weinstien we can locally in a neighborhood of a point p find a coordinate system

(q^i, p_i, w^i) centered at p and such that

$$B = \sum_{i=1}^{k} \frac{\partial}{\partial q^i} \wedge \frac{\partial}{\partial p_i} + \frac{1}{2} \sum_{i,j} a^{ij}() \frac{\partial}{\partial w^i} \wedge \frac{\partial}{\partial w^j}$$

where the smooth functions depend only on the w's. vanish at p. Here k is the dimension of the leave through p. The rank of the map $B^\sharp$ on T_p^*M is k.

Now to give a typical example let $\mathfrak{g}$ be a Lie algebra with bracket $[.,.]$ and $\mathfrak{g}^*$ its dual. Choose a basis $e_1, ..., e_n$ of $\mathfrak{g}$ and the corresponding dual basis $\epsilon^1, ..., \epsilon^n$ for $\mathfrak{g}^*$. With respect to the basis $e_1, ..., e_n$ we have

$$[e_i, e_j] = \sum C_{ij}^k e_k$$

where C_{ij}^k are the structure constants.

For any functions $f, g \in \mathfrak{F}(\mathfrak{g}^*)$ we have that df_α, dg_α are linear maps $\mathfrak{g}^* \to \mathbb{R}$ where we identify $T_\alpha \mathfrak{g}^*$ with $\mathfrak{g}^*$. This means that df_α, dg_α can be considered to be in $\mathfrak{g}$ by the identification $\mathfrak{g}^{**} = \mathfrak{g}$. Now define the $\pm$ Poisson structure on $\mathfrak{g}^*$ by

$$\{f, g\}_\pm(\alpha) = \pm\alpha([df_\alpha, dg_\alpha])$$

Now the basis $e_1, ..., e_n$ is a coordinate system y on $\mathfrak{g}^*$ by $y_i(\alpha) = \alpha(e_i)$.

Proposition 29.1 *In terms of this coordinate system the Poisson bracket just defined is*

$$\{f, g\}_\pm = \pm \sum_{i=1}^{n} B_{ij} \frac{\partial f}{\partial y_i} \frac{\partial g}{\partial y_j}$$

where $B_{ij} = \sum C_{ij}^k y_k$.

Proof. We suppress the $\pm$ and compute:

$$\begin{aligned} \{f, g\} &= [df, dg] = [\sum \frac{\partial f}{\partial y_i} dy_i, \sum \frac{\partial g}{\partial y_j} dy_j] \\ &= \sum \frac{\partial f}{\partial y_i} \frac{\partial g}{\partial y_j} [dy_i, dy_j] = \sum \frac{\partial f}{\partial y_i} \frac{\partial g}{\partial y_j} \sum C_{ij}^k y_k \\ &= \sum_{i=1}^{n} B_{ij} \frac{\partial f}{\partial y_i} \frac{\partial g}{\partial y_j} \end{aligned}$$

■

Chapter 30

Quantization

UNDER CONSTRUCTION

30.1 Operators on a Hilbert Space

A bounded linear map between complex Hilbert spaces $A : \mathcal{H}_1 \to \mathcal{H}_2$ is said to be of finite rank if and only if the image $A(\mathcal{H}_1) \subset \mathcal{H}_2$ is finite dimensional. Let us denote the set of all such finite rank maps by $\mathcal{F}(\mathcal{H}_1, \mathcal{H}_2)$. If $\mathcal{H}_1 = \mathcal{H}_2$ we write $\mathcal{F}(\mathcal{H})$. The set of bounded linear maps is denoted $\mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ and is a Banach space with norm given by $\|A\|_{\mathcal{B}(\mathcal{H}_1,\mathcal{H}_2)} := \sup\{\|Av\| : \|v\| = 1\}$. Let $\mathcal{B}(\mathcal{H})$ be the algebra of bounded operators that is then a Banach space with the norm given by $\|A\|_{\mathcal{B}(\mathcal{H})} := \sup\{\|Av\| : \|v\| = 1\}$. The reader will recall that $\mathcal{B}(\mathcal{H})$ is in fact a Banach algebra meaning that in we have

$$\|AB\|_{\mathcal{B}(\mathcal{H})} \le \|A\|_{\mathcal{B}(\mathcal{H})} \|B\|_{\mathcal{B}(\mathcal{H})}.$$

We will abbreviate $\|A\|_{\mathcal{B}(\mathcal{H})}$ to just $\|A\|$ when convenient. Recall that the adjoint of an element $A \in \mathcal{B}(\mathcal{H})$ is that unique element A^* defined by $\langle A^*v, w\rangle = \langle v, Aw\rangle$ for all $v, w \in \mathcal{H}$. One can prove that in fact we have

$$\|A^*A\| = \|A\|^2$$

One can define on $\mathcal{B}(\mathcal{H})$ two other important topologies. Namely, the **strong topology** is given by the family of seminorms $A \mapsto \|Av\|$ for various $v \in \mathcal{H}$ and the **weak topology** given by the family of seminorms $A \mapsto \langle Av, w\rangle$ for various $v, w \in \mathcal{H}$. Because of this we will use phrases like "norm-closed", "strongly closed" or "weakly closed" etc.

Recall that for any set $W \subset \mathcal{H}$ we define $W^\perp := \{v : \langle v, w\rangle = 0$ for all $w \in W\}$.The following fundamental lemma is straightforward to prove.

Lemma 30.1 *For any $A \in \mathcal{B}(\mathcal{H})$ we have*

$$\overline{\operatorname{img}(A^*)} = \ker(A)^\perp$$
$$\ker(A^*) = \operatorname{img}(A)^\perp.$$

Lemma 30.2 *$\mathcal{F}(\mathcal{H})$ has the following properties:*
1) $F \in \mathcal{F}(\mathcal{H}) \Longrightarrow A \circ F \in \mathcal{F}(\mathcal{H})$ and $F \circ A \in \mathcal{F}(\mathcal{H})$ for all $A \in \mathcal{B}(\mathcal{H})$.
2) $F \in \mathcal{F}(\mathcal{H}) \Longrightarrow F^* \in \mathcal{F}(\mathcal{H})$.
*Thus $\mathcal{F}(\mathcal{H})$ is a two sided ideal of $\mathcal{B}(\mathcal{H})$ closed under *.*

Proof. Obvious. ■

Definition 30.1 *An linear map $A \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ is called compact if $A(\overline{B(0,1)})$ is relatively compact (has compact closure) in $\mathcal{H}_2$. The set of all compact operators $\mathcal{H}_1 \to \mathcal{H}_2$, is denoted $\mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$ or if $\mathcal{H}_1 = \mathcal{H}_2$ by $\mathcal{K}(\mathcal{H})$.*

Lemma 30.3 *$\mathcal{F}(\mathcal{H}_1, \mathcal{H}_2)$ is dense in $\mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$ and $\mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$ is closed in $\mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$.*

Proof. Let $A \in \mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$. For every small $\epsilon > 0$ there is a finite subset $\{y_1, y_2, .., y_k\} \subset \overline{B(0,1)}$ such that the $\epsilon-$balls with centers $y_1, y_2, .., y_k$ cover $A(\overline{B(0,1)})$. Now let P be the projection onto the span of $\{Ay_1, Ay_2, .., Ay_k\}$. We leave it to the reader to show that $PA \in \mathcal{F}(\mathcal{H}_1, \mathcal{H}_2)$ and $\|PA - A\| < \epsilon$. It follows that $\mathcal{F}(\mathcal{H}_1, \mathcal{H}_2)$ is dense in $\mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$.

Let $\{A_n\}_{n>0}$ be a sequence in $\mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$ that converges to some $A \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$. We want to show that $A \in \mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$. Let $\epsilon > 0$ be given and choose $n_0 > 0$ so that $\|A_n - A\| < \epsilon$. Now A_{n_0} is compact so $A_{n_0}(\overline{B(0,1)})$ is relatively compact. It follows that there is some finite subset $N \subset \overline{B(0,1)}$ such that

$$A_{n_0}(\overline{B(0,1)}) \subset \bigcup_{a \in N} B(a, \epsilon).$$

Then for any $y \in \overline{B(0,1)}$ there is an element $a \in N$ such that

$$\begin{aligned}
&\|A(y) - A(a)\| \\
&\leq \|A(y) - A_{n_0}(y)\| + \|A_{n_0}(y) - A_{n_0}(a)\| + \|A_{n_0}(a) - A(a)\| \\
&\leq 3\epsilon
\end{aligned}$$

and it follows that $A(\overline{B(0,1)})$ is relatively compact. Thus $A \in \mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$. ■

Corollary 30.1 *$\mathcal{K}(\mathcal{H})$ is a two sided ideal in $\mathcal{B}(\mathcal{H})$.*

We now give two examples (of types) of compact operators each of which is prototypical in its own way.

Example 30.1 *For every continuous function $K(.,.) \in L^2([0,1] \times [0,1])$ gives rise to a compact integral operator $\widetilde{K}$ on $L^2([0,1])$ defined by*

$$(A_K f)(x) := \int_{[0,1]\times[0,1]} K(x,y) f(y) dy.$$

Example 30.2 *If $\mathcal{H}$ is separable and we fix a Hilbert basis $\{e_i\}$ then we can define an operator by prescribing its action on the basis elements. For every sequence of numbers $\{\lambda_i\} \subset \mathbb{C}$ such that $\lambda_i \to 0$ we get an operator $A_{\{\lambda_i\}}$ defined by*

$$A_{\{\lambda_i\}} e_i = \lambda_i e_i.$$

This operator is the norm limit of the finite rank operators A_n defined by

$$A_n(e_i) = \begin{cases} \lambda_i e_i \text{ if } i \leq n \\ \\ 0 \text{ if } i > n \end{cases}.$$

*Thus $A_{\{y_i\}}$ is compact by lemma 30.3. This type of operator is called a **diagonal operator** (with respect to the basis $\{e_i\}$).*

30.2 C*-Algebras

Definition 30.2 *A **C* algebra of (bounded) operators** on $\mathcal{H}$ is a norm-closed subalgebra $\mathfrak{A}$ of $\mathcal{B}(\mathcal{H})$ such that*

$$A^* \in \mathfrak{A} \text{ for all } A \in \mathfrak{A}.$$

A trivial example is the space $\mathcal{B}(\mathcal{H})$ itself but beyond that we have the following simple but important example:

Example 30.3 *Consider the case where $\mathcal{H}$ is $L^2(X, \mu)$ for some compact space X and where the measure μ is a positive measure (i.e. non-negative) such that $\mu(U) > 0$ for all nonempty open sets U. Now for every function $f \in C(X)$ we have a multiplication operator $M_f : L^2(X, \mu) \to L^2(X, \mu)$ defined by $M_f(g) = fg$. The map $f \mapsto M_f$ is an algebra monomorphism from $C(X)$ into $\mathcal{B}(\mathcal{H}) := \mathcal{B}(L^2(X, \mu))$ such that $\overline{f} \mapsto M_f^* = M_{\overline{f}}$. Thus we identify $C(X)$ with a subspace of $\mathcal{B}(\mathcal{H})$ which is in fact a C^* algebra of operators on $\mathcal{H}$. This is an example of a **commutative** C^*-algebra.*

Example 30.4 (!) *Using corollary 30.1 one can easily see that $\mathcal{K}(\mathcal{H})$ is a C^* algebra of operators on $\mathcal{H}$. The fact that $\mathcal{K}(\mathcal{H})$ is self adjoint (closed under adjoint $A \mapsto A^*$) follows from the self adjointness of the algebra $\mathcal{F}(\mathcal{H})$.*

Definition 30.3 *A C^* algebra of operators on $\mathcal{H}$ is called separable if it has a countable dense subset.*

Proposition 30.1 *The C^* algebra $\mathcal{B}(\mathcal{H})$ itself is separable if and only if $\mathcal{H}$ is finite dimensional.*

Remark 30.1 *If one gives $\mathcal{B}(\mathcal{H})$ the* strong *topology then $\mathcal{B}(\mathcal{H})$ is separable if and only if $\mathcal{H}$ is separable.*

Proposition 30.2 *The algebra of multipliers $\mathfrak{M} \cong C(X)$ from example 30.3 is separable if and only if X is a separable compact space.*

In the case that $\mathcal{H}$ is finite dimensional we may as well assume that $\mathcal{H} = \mathbb{C}^n$ and then we can also identify $\mathcal{B}(\mathcal{H})$ with the algebra of matrices $\mathbb{M}_{n\times n}(\mathbb{C})$ where now A^* refers to the conjugate transpose of the matrix A. On can also verify that in this case the "operator" norm of a matrix A is given by the maximum of the eigenvalues of the self-adjoint matrix A^*A. Of course one also has

$$\|A\| = \|A^*A\|^{1/2}.$$

30.2.1 Matrix Algebras

It will be useful to study the finite dimensional case in order to gain experience and put the general case in perspective.

30.3 Jordan-Lie Algebras

In this section we will not assume that algebras are necessarily associative or commutative. We also do not always require an algebra to have a unity. If we consider the algebra $\mathbb{M}_{n\times n}(\mathbb{R})$ of $n \times n$ matrices and define a new product $\diamond$ by

$$A \diamond B := \frac{1}{2}(AB + BA)$$

then the resulting algebra $\mathbb{M}_{n\times n}(\mathbb{R}), \diamond$ is not associative and is an example of a so called Jordan algebra. This kind of algebra plays an important role in quantization theory. The general definition is as follows.

Definition 30.4 *A* ***Jordan algebra*** $\mathfrak{A}$,$\diamond$ *is a commutative algebra such that*

$$A \diamond (B \diamond A^2) = (A \diamond B) \diamond A^2.$$

In the most useful cases we have more structure which abstracts as the following:

Definition 30.5 *A Jordan Morphism* $h : \mathfrak{A},\diamond \to \mathfrak{B},\diamond$ *between Jordan algebras is a linear map satisfying* $h(A \diamond B) = h(A) \diamond h(B)$ *for all* $A, B \in \mathfrak{A}$.

Definition 30.6 *A* ***Jordan-Lie algebra*** *is a real vector space* $\mathfrak{A}_{\mathbb{R}}$ *together with two bilinear products* $\diamond$ *and* $\{.,.\}$ *such that*

1)

$$A \diamond B = B \diamond A$$
$$\{A, B\} = -\{B, A\}$$

2) For each $A \in \mathfrak{A}_{\mathbb{R}}$ *the map* $B \mapsto \{A, B\}$ *is a derivation for both* $\mathfrak{A}_{\mathbb{R}}, \diamond$ *and* $\mathfrak{A}_{\mathbb{R}}, \{.,.\}$.

3) There is a constant $\hbar$ *such that we have the following* ***associator identity:***

$$(A \diamond B) \diamond C - (A \diamond B) \diamond C = \frac{1}{4}\hbar^2\{\{A, , C\}, B\}.$$

Observe that this definition actually says that a **Jordan-Lie algebra** is actually two algebras coupled by the associator identity and the requirements concerning the derivation property. The algebra $\mathfrak{A}_{\mathbb{R}}, \diamond$ is a Jordan algebra and $\mathfrak{A}_{\mathbb{R}}, \{.,.\}$ is a Lie algebra.

Notice that if a Jordan-Lie algebra $\mathfrak{A}_{\mathbb{R}}$ is associative then by definition it is a Poisson algebra.

Because we have so much structure we will be interested in maps that preserve some or all of the structure and this leads us the following definition:

Definition 30.7 *A* ***Jordan*** *(resp.* ***Poisson****) morphism $h : \mathfrak{A}_{\mathbb{R}} \to \mathfrak{B}_{\mathbb{R}}$ between Jordan-Lie algebras is a Jordan morphism (resp. Lie algebra morphism) on the underlying algebras $\mathfrak{A}_{\mathbb{R}}, \diamond$ and $\mathfrak{B}_{\mathbb{R}}, \diamond$ (resp. underlying Lie algebras $\mathfrak{A}_{\mathbb{R}}, \{.,.\}$ and $\mathfrak{B}_{\mathbb{R}}, \{.,.\}$). A map $h : \mathfrak{A}_{\mathbb{R}} \to \mathfrak{B}_{\mathbb{R}}$ that is simultaneously a Jordan morphism and a Poisson morphism is called a (Jordan-Lie)* ***morphism****. In each case we also have the obvious definitions for isomorphism.*

Chapter 31

General Relativity

Of the basic forces in nature that are still thought to be fundamental, gravity has been studied the longest and yet it is still at the frontier of physics. This not because of because of the fairly recent advent of General Relativity (GR) but more because of the difficulty science is currently facing with respect to providing a coherent quantum theory of gravity that retains the successful aspects of General Relativity. What is needed is a quantum version of General Relativity and hopefully one that can be unified with the Gauge theories that describe the other three basic forces. In any case, one cannot ignore classical General Relativity which Penrose classifies as a "superb theory". When judged in terms of mathematical beauty and consistency of logical structure Gravity has only classical (relativistic) electromagnetism as a competitor. It is interesting to note that the Gravitational physics of GR has been tested against experiment and has pass time and again with flying colors. The most famous include an explanation of the precession of the perihelion of the orbit of mercury about the sun as well as the observed ability of gravitation to bend light rays. As almost every educated person knows, General Relativity is largely the intellectual product of one man; Albert Einstein. On the other hand, one of the predictions of the theory is that of gravitational waves and nothing of the sort has been directly detected as of the writing of this book[1]. General relativity deals with the physics of extremely large regions of the universe. When the whole universe is considered by physics it is generally referred to as cosmology. Cosmology uses GR to make predictions about the origin of the universe. In fact, because of the prediction of an expanding universe one could say that it was GR that first gave mathematical reason to expect that the universe did not exist in an infinite past and from all eternity into the future.

General Relativity predicts the existence of exotic regions of spacetime called Black Holes and in fact one model of the beginning of time is that of a sort of reverse Black Hole (a White hole) which is, of course, the famous "big bang". A

[1] A sort of indirect confirmation of the existence of gravity wave was provided by an analysis of the orbital period of the Hulse-Taylor binary pulsar. Hulse and Taylor were awarded a Nobel Prise in 1993 for the discovery of this pulsar.

simple model of a black hole is described by a "spherically symmetric" Lorentz manifold whose metric is called the Schwarzschild metric. It is now believed that the central core of our own galaxy is a black hole. There are many strange and exotic aspects to the space (or spacetime) surrounding a black whole. Perhaps the simplest thing we can say about black holes before our mathematical analysis is that gravity is so strong near the black hole that once matter gets close enough it can never escape. Even light cannot escape its pull of gravity near a black hole. Actually, as we shall see, the word "pull" is does not appropriately describe the true nature of gravity. In fact gravity is not force in the ordinary sense of the word. What is gravity then? The answer is that gravity is the manifestation of the curvature of spacetime itself. More succinctly; "gravity is geometry". It is perhaps less well known that Newton's theory of gravity (highly accurate within is range of applicability) can also be formulated in terms of geometry. The difference lies not only in the fact that GR win the experimental contest but also in the fact that if the geometrically reformulated version of the Newtonian theory is compared point by point with GR then it become clear how much more elegant and intellectually satisfying is the theory of Einstein.

Chapter 32

Appendix A

Topological Spaces

In this section we briefly introduce the basic notions from point set topology together with some basic examples. We include this section only as a review and a reference since we expect that the reader should already have a reasonable knowledge of point set topology. In the Euclidean coordinate plane $\mathbb{R}^n$ consisting of all n-tuples of real numbers $(x_1, x_2, ...x_n)$ we have a natural notion of distance between two points. The formula for the distance between two points $p_1 = (x_1, x_2, ...x_n)$ and $p_2 = (y_1, y_2, ..., y_n)$ is simply

$$d(p_1, p_2) = \sqrt{\sum (x_i - y_i)^2}. \tag{A.1}$$

The set of all points of distance less than ϵ from a given point p_0 in the plain is denoted $B(p_0, \epsilon)$, i.e.

$$B(p_0, \epsilon) = \{p \in \mathbb{R}^n : d(p, p_0) < \epsilon\}. \tag{A.2}$$

The set $B(p_0, \epsilon)$ is call the **open ball** of radius ϵ and center p_0. A subset S of $\mathbb{R}^2$ is called **open** if every one of its points is the center of an open ball completely contained inside S. The set of all open subsets of the plane has the property that the union of any number of open sets is still open and the intersection of any finite number of open sets is still open. The abstraction of this situation leads to the concept of a topological space.

Definition A.1 *A set X together with a family $\mathfrak{T}$ of subsets of X is called a topological space if the family $\mathfrak{T}$ has the following three properties.*

1. *$X \in \mathfrak{T}$ and $\emptyset \in \mathfrak{T}$.*
2. *If U_1 and U_2 are both in $\mathfrak{T}$ then $U_1 \cap U_2 \in \mathfrak{T}$ also.*
3. *If $\{U_\alpha\}_{\alpha \in A}$ is any sub-family of $\mathfrak{T}$ indexed by a set A then the union $\bigcup_{\alpha \in A} U_\alpha$ is also in $\mathfrak{T}$.*

In the definition above the family of subsets $\mathfrak{T}$ is called a **topology** on X and the sets in $\mathfrak{T}$ are called **open** sets. The compliment $U^c := X \setminus U$ of an open set U is called **closed** set. The reader is warned that a generic set may be neither open nor closed. Also, some subsets of X might be both open and closed (consider X itself and the empty set). A topology $\mathfrak{T}_2$ is said to be **finer** than a topology $\mathfrak{T}_1$ if $\mathfrak{T}_1 \subset \mathfrak{T}_2$ and in this case we also say that $\mathfrak{T}_1$ is **coarser** than $\mathfrak{T}_2$. We also say that the topology $\mathfrak{T}_1$ is **weaker** than $\mathfrak{T}_2$ and that $\mathfrak{T}_2$ is **stronger** than $\mathfrak{T}_1$.

Neither one of these topologies is generally very interesting but we shall soon introduce much richer topologies. A fact worthy of note in this context is the fact that if $X, \mathfrak{T}$ is a topological space and $S \subset X$ then S inherits a topological structure from X. Namely, a topology on S (called the **relative topology**) is given by

$$\mathfrak{T}_S = \{\text{all sets of the form } S \cap T \text{ where } \; T \in \mathfrak{T}\} \tag{A.3}$$

In this case we say that S is a topological subspace of X.

Definition A.2 *A map between topological spaces $f : X \to Y$ is said to be **continuous at** $p \in X$ if for any open set O containing $f(p)$ there is an open set U containing $p \in X$ such that $f(U) \subset O$. A map $f : X \to Y$ is said to be **continuous** if it is continuous at each point $p \in X$.*

Proposition A.1 *$f : X \to Y$ is continuous if and only if $f^{-1}(O)$ is open for every open set $O \subset Y$.*

Definition A.3 *A subset of a topological space is called **closed** if it is the compliment of an open set.*

Closed sets enjoy properties complimentary to those of open sets:

1. The whole space X and the empty set $\emptyset$ are both closed.

2. The intersection of any family of closed sets is a closed set.

3. The union of a finite number of closed sets is closed.

Since the intersection of closed sets is again a closed every set, the subset $S \subset X$ is contained in a closed set which is the smallest of all closed sets containing S. This **closure** $\overline{S}$ is the intersection of all closed subsets containing S:

$$\overline{S} = \bigcap_{S \subset F} F.$$

Similarly, the **interior** of a set S is the largest open set contained in S and is denoted by $\overset{\circ}{S}$. A point $p \in S \subset X$ is called an **interior point** of S if there is an open set containing p and contained in S. The interior of S is just the set of all its interior points. It may be shown that $\overset{\circ}{S} = (\overline{S^c})^c$

Definition A.4 *The (topological)* ***boundary*** *of a set* $S \subset X$ *is* $\partial S := \overline{S} \cap \overline{S^c}$ *and*

We say that a set $S \subset X$ is **dense** in X if $\overline{S} = X$.

Definition A.5 *A subset of a topological space* $X, \mathfrak{T}$ *is called* ***clopen*** *if it is both open and closed.*

Definition A.6 *A topological space* X *is called* ***connected*** *if it is not the union of two proper clopen set. Here, proper means not* X *or* $\emptyset$. *A topological space* X *is called* ***path connected*** *if for every pair of points* $p, q \in X$ *there is a continuous map* $c : [a, b] \to X$ *(a path) such that* $c(a) = q$ *and* $c(b) = p$. *(Here* $[a, b] \subset \mathbb{R}$ *is endowed with the relative topology inherited from the topology on* $\mathbb{R}$.*)*

Example A.1 *The unit sphere* S^2 *is a topological subspace of the Euclidean space* $\mathbb{R}^3$.

Let X be a set and $\{\mathfrak{T}_\alpha\}_{\alpha \in A}$ any family of topologies on X indexed by some set A. The the intersection

$$\mathfrak{T} = \bigcap_{\alpha \in A} \mathfrak{T}_\alpha$$

is a topology on X. Furthermore, $\mathfrak{T}$ is coarser that every $\mathfrak{T}_\alpha$.

Given any family $\mathfrak{F}$ of subsets of X there exists a weakest (coarsest) topology containing all sets of $\mathfrak{F}$. We will denote this topology by $\mathfrak{T}(\mathfrak{F})$.

One interesting application of this is the following; Given a family of maps $\{f_\alpha\}$ from a set S to a topological space $Y, \mathfrak{T}_Y$ there is a coarsest topology on S such that all of the maps f_α are continuous. This topology will be denoted $\mathfrak{T}_{\{f_\alpha\}}$ and is called the topology generated by the family of maps $\{f_\alpha\}$.

Definition A.7 *If* X *and* Y *are topological spaces then we define the* ***product topology*** *on* $X \times Y$ *as the topology generated by the projections* $pr_1 : X \times Y \to X$ *and* $pr_2 : X \times Y \to Y$.

Definition A.8 *If* $\pi : X \to Y$ *is a surjective map where* X *is a topological space but* Y *is just a set. Then the* ***quotient topology*** *is the topology generated by the map* π. *In particular, if* $A \subset X$ *we may form the set of equivalence classes* X/A *where* $x \sim y$ *if both are in* A *or they are equal. The the map* $x \mapsto [x]$ *is surjective and so we may form the quotient topology on* X/A.

Let X be a topological space and $x \in X$. A family of subsets $\mathcal{B}_x$ all of which contain x is called an **open neighborhood base** at x if every open set containing x contains (as a superset) an set from $\mathcal{B}_x$. If X has a countable open base at each $x \in X$ we call X **first countable**.

A subfamily $\mathcal{B}$ is called a **base** for a topology $\mathfrak{T}$ on X if the topology $\mathfrak{T}$ is exactly the set of all unions of elements of $\mathcal{B}$. If X has a countable base for its given topology we say that X is a **second countable** topological space.

By considering balls of rational radii and rational centers one can see that $\mathbb{R}^n$ is first and second countable.

A.0.1 Separation Axioms

Another way to classify topological spaces is according to the following scheme:

(**Separation Axioms**)

A topological space $X, \mathfrak{T}$ is called a T_0 space if given $x, y \in X$, $x \neq y$, there exists either an open set containing x, but not y or the other way around (We don't get to choose which one).

A topological space $X, \mathfrak{T}$ is called T_1 if whenever given any x,$y \in X$ there is an open set containing x but not y (and the other way around;we get to do it either way).

A topological space $X, \mathfrak{T}$ is called T_2 or **Hausdorff** if whenever given any two points x, $y \in X$ there are disjoint open sets U_1 and U_2 with $x \in U_1$ and $y \in U_2$.

A topological space $X, \mathfrak{T}$ is called T_3 or **regular** if whenever given a closed set $F \subset X$ and a point $x \in X \backslash F$ there are disjoint open sets U_1 and U_2 with $x \in U_1$ and $F \subset U_2$

A topological space $X, \mathfrak{T}$ is called T_4 or **normal** if given any two disjoint closed subsets of X, say F_1 and F_2, there are two disjoint open sets U_1 and U_2 with $F_1 \subset U_1$ and $F_2 \subset U_2$.

Lemma A.1 (Urysohn) *Let X be normal and $F, G \subset X$ closed subsets with $F \cap G = \emptyset$. Then there exists a continuous function $f : X \to [0,1] \subset \mathbb{R}$ such that $f(F) = 0$ and $f(G) = 1$.*

A open cover of topological space X (resp. subset $S \subset X$) a collection of open subsets of X, say$\{U_\alpha\}$, such that $X = \bigcup U_\alpha$ (resp. $S \subset \bigcup U_\alpha$). For example the set of all open disks of radius $\epsilon > 0$ in the plane covers the plane. A finite cover consists of only a finite number of open sets.

Definition A.9 *A topological space X is called **compact** if every open cover of X can be reduced to a finite open cover by eliminating some (possibly an infinite number) of the open sets of the cover. A subset $S \subset X$ is called **compact** if it is compact as a topological subspace (i.e. with the relative topology).*

Proposition A.2 *The continuous image of a compact set is compact.*

A.0.2 Metric Spaces

If the set X has a notion of distance attached to it then we can get an associated topology. This leads to the notion of a metric space.

A set X together with a function $d : X \times X \to \mathbb{R}$ is called a **metric space** if

$d(x, x) \geq 0$ for all $x \in X$

$d(x, y) = 0$ if and only if $x = y$

$d(x, z) \leq d(x, y) + d(y, z)$ for any $x, y, z \in X$ (*this is called the triangle inequality*).

The function d is called a **metric** or a **distance function**.

Imitating the situation in the plane we can define the notion of an open ball $B(p_0, \epsilon)$ with center p_0 and radius ϵ. Now once we have the metric then we have a topology; we define a subset of a metric space X, d to be open if every point of S is an interior point where a point $p \in S$ is called an **interior point** of S if there is some ball $B(p, \epsilon)$ with center p and (sufficiently small) radius $\epsilon > 0$ completely contained in S. The family of all of these metrically defined open sets forms a topology on X which we will denote by $\mathfrak{T}_d$. It is easy to see that any $B(p, \epsilon)$ is open according to our definition.

If $f : X, d \to Y, \rho$ is a map of metric spaces then f is continuous at $x \in X$ if and only if for every $\epsilon > 0$ there is a $\delta(\epsilon) > 0$ such that if $d(x', x) < \delta(\epsilon)$ then $\rho(f(x'), f(x)) < \epsilon$.

Definition A.10 *A sequence of elements $x_1, x_2, ...$ of a metric space X, d is said to* ***converge*** *to p if for every $\epsilon > 0$ there is an $N(\epsilon) > 0$ such that if $k > N(\epsilon)$ then $x_k \in B(p, \epsilon)$. A sequence $x_1, x_2, ...$ is called a* ***Cauchy sequence*** *is for every $\epsilon > 0$ there is an $N(\epsilon) > 0$ such that if $k, l > N(\epsilon)$ then $d(x_k, x_l) < \epsilon$. A metric space X, d is said to be* ***complete*** *if every Cauchy sequence also converges.*

A map $f : X, d \to Y, \rho$ of metric spaces is continuous at $x \in X$ if and only if for every sequence x_i converging to x, the sequence $y_i := f(x_i)$ converges to $f(x)$.

A.1 B. Attaching Spaces and Quotient Topology

Suppose that we have a topological space X and a surjective set map $f : X \to S$ onto some set S. We may endow S with a natural topology according to the following recipe. A subset $U \subset S$ is defined to be open if and only if $f^{-1}(U)$ is an open subset of X. This is particularly useful when we have some equivalence relation on X which allows us to consider the set of equivalence classes $X/\sim$. In this case we have the canonical map $\varrho : X \to X/\sim$ that takes $x \in X$ to its equivalence class $[x]$. The quotient topology is then given as before by the requirement that $U \subset S$ is open if and only if and only if $\varrho^{-1}(U)$ is open in X. A common application of this idea is the identification of a subspace to a point. Here we have some subspace $A \subset X$ and the equivalence relation is given by the following two requirements:

$$\begin{array}{rl} \text{If } x \in X \backslash A & \text{then } x \sim y \text{ only if } x = y \\ \text{If } x \in A & \text{then } x \sim y \text{ for any } y \in A \end{array}$$

In other words, every element of A is identified with every other element of A. We often denote this space by X/A.

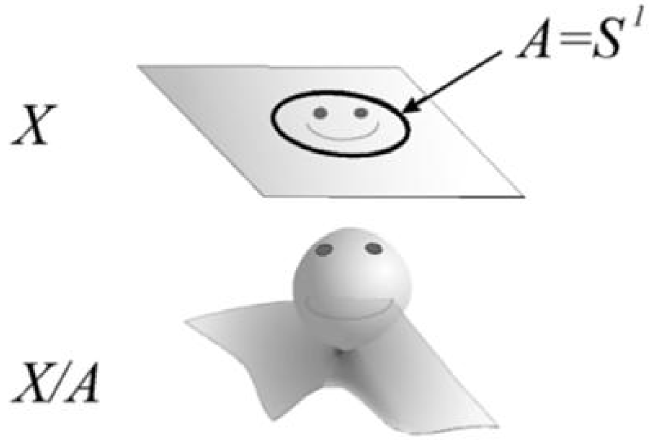
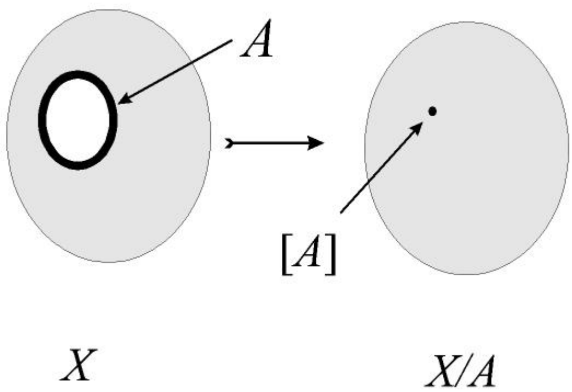


Figure A.1: creation of a "hole"

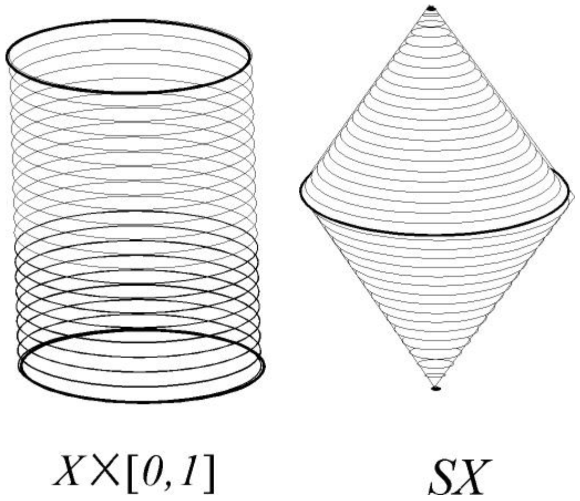


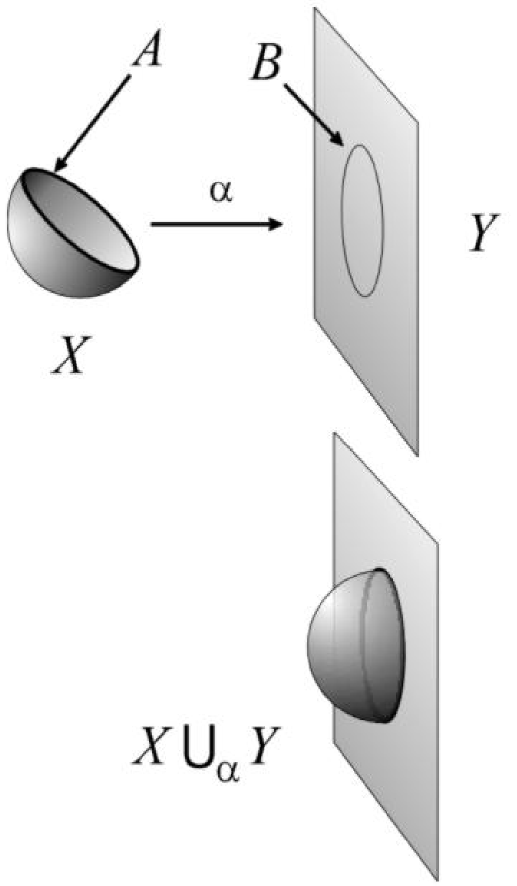
A hole is removed by identification

It is not difficult to verify that if X is Hausdorff (resp. normal) and A is closed then X/A is Hausdorff (resp. normal). The identification of a subset to a point need not simplify the topology but may also complicate the topology as shown in the figure.

An important example of this construction is the suspension. If X is a topological space then we define its suspension SX to be $(X \times [0,1])/A$ where $A := (X \times \{0\}) \cup (X \times \{1\})$. For example it is easy to see that $SS^1 \cong S^2$. More generally, $SS^{n-1} \cong S^n$.

Consider two topological spaces X and Y and subset $A \subset X$ a closed subset. Suppose that we have a map $\alpha : A \to B \subset Y$. Using this map we may define an equivalence relation on the disjoint union $X \bigsqcup Y$ that is given by requiring that $x \sim \alpha(x)$ for $x \in A$. The resulting topological space is denoted $X \cup_\alpha Y$.


$X \times [0,1]$
SX


A
B
α
Y
X
$X \cup_\alpha Y$

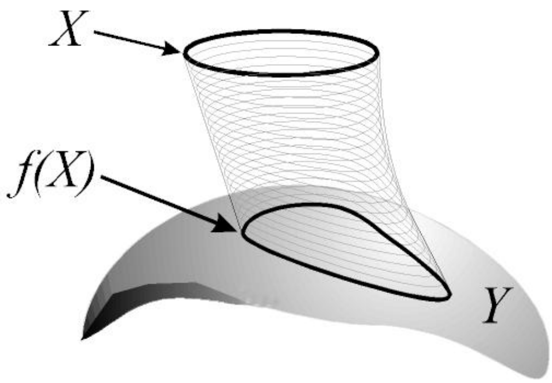


Figure A.2: Mapping Cylinder

Attaching a 2-cell

Another useful construction is that of a the mapping cylinder of a map $f : X \to Y$. First we transfer the map to a map on the base $X \times \{0\}$ of the cylinder $X \times I$ by

$$f(x, 0) := f(x)$$

and then we form the quotient $Y \cup_f (X \times I)$. We denote this quotient by M_f and call it the mapping cylinder of f.

A.2 C. Topological Vector Spaces

We shall outline some of the basic definitions and theorems concerning topological vector spaces.

Definition A.11 *A* ***topological vector space (TVS)*** *is a vector space* V *with a Hausdorff topology such that the addition and scalar multiplication operations are (jointly) continuous.*

Definition A.12 *Recall that a neighborhood of a point p in a topological space is a subset that has a nonempty interior containing p. The set of all neighborhoods that contain a point x in a topological vector space* V *is denoted* $\mathcal{N}(x)$.

The families $\mathcal{N}(x)$ for various x satisfy the following neighborhood axioms

1. Every set that contains a set from $\mathcal{N}(x)$ is also a set from $\mathcal{N}(x)$
2. If N_i is a family of sets from $\mathcal{N}(x)$ then $\bigcap_i N_i \in \mathcal{N}(x)$
3. Every $N \in \mathcal{N}(x)$ contains x

4. If $V \in \mathcal{N}(x)$ then there exists $W \in \mathcal{N}(x)$ such that for all $y \in W$, $V \in \mathcal{N}(y)$.

Conversely, let X be some set. If for each $x \in X$ there is a family $\mathcal{N}(x)$ of subsets of X that satisfy the above neighborhood axioms then there is a uniquely determined topology on X for which the families $\mathcal{N}(x)$ are the neighborhoods of the points x. For this a subset $U \subset X$ is open if and only if for each $x \in U$ we have $U \in \mathcal{N}(x)$.

Definition A.13 *A sequence x_n in a TVS is call a Cauchy sequence if and only if for every neighborhood U of 0 there is a number N_U such that $x_l - x_k \in U$ for all $k, l \geq N_U$.*

Definition A.14 *A relatively nice situation is when V has a norm that induces the topology. Recall that a **norm** is a function $\|\| : v \mapsto \|v\| \in \mathbb{R}$ defined on V such that for all $v, w \in \mathsf{V}$ we have*

1. $\|v\| \geq 0$ and $\|v\| = 0$ if and only if $v = 0$,
2. $\|v + w\| \leq \|v\| + \|w\|$,
3. $\|\alpha v\| = |\alpha| \, \|v\|$ for all $\alpha \in \mathbb{R}$.

   In this case we have a metric on V given by $\operatorname{dist}(v, w) := \|v - w\|$. A **seminorm** is a function $\|\| : v \mapsto \|v\| \in \mathbb{R}$ such that 2) and 3) hold but instead of 1) we require only that $\|v\| \geq 0$.

Definition A.15 *A **normed space** V is a TVS that has a metric topology given by a norm. That is the topology is generated by the family of all open balls*

$$B_{\mathsf{V}}(x, r) := \{x \in \mathsf{V} : \|x\| < r\}.$$

Definition A.16 *A linear map $\ell : \mathsf{V} \to \mathsf{W}$ between normed spaces is called bounded if and only if there is a constant C such that for all $v \in \mathsf{V}$ we have $\|\ell v\|_{\mathsf{W}} \leq C \|v\|_{\mathsf{V}}$. If ℓ is bounded then the smallest such constant C is*

$$\|\ell\| := \sup \frac{\|\ell v\|_{\mathsf{W}}}{\|v\|_{\mathsf{V}}} = \sup\{\|\ell v\|_{\mathsf{W}} : \|v\|_{\mathsf{V}} \leq 1\}$$

The set of all bounded linear maps $\mathsf{V} \to \mathsf{W}$ is denoted $\mathcal{B}(\mathsf{V},\mathsf{W})$. The vector space $\mathcal{B}(\mathsf{V},\mathsf{W})$ is itself a normed space with the norm given as above.

Definition A.17 *A **locally convex topological vector space** V is a TVS such that it's topology is generated by a family of seminorms $\{\|.\|_\alpha\}_\alpha$. This means that we give V the weakest topology such that all $\|.\|_\alpha$ are continuous. Since we have taken a TVS to be Hausdorff we require that the family of seminorms is **sufficient** in the sense that for each $x \in \mathsf{V}$ we have $\bigcap\{x : \|x\|_\alpha = 0\} = \emptyset$. A locally convex topological vector space is sometimes called a locally convex space and so we abbreviate the latter to **LCS**.*

Example A.2 *Let Ω be an open set in $\mathbb{R}^n$ or any manifold. For each $x \in \Omega$ define a seminorm ρ_x on $C(\Omega)$ by $\rho_x(f) = f(x)$. This family of seminorms makes $C(\Omega)$ a topological vector space. In this topology convergence is pointwise convergence. Also, $C(\Omega)$ is not complete with this TVS structure.*

Definition A.18 *An LCS that is complete (every Cauchy sequence converges) is called a **Frechet space**.*

Definition A.19 *A complete normed space is called a **Banach space**.*

Example A.3 *Suppose that X, μ is a σ-finite measure space and let $p \geq 1$. The set $L^p(X, \mu)$ of all with respect to measurable functions $f : X \to \mathbb{C}$ such that $\int |f|^p \, d\mu \leq \infty$ is a Banach space with the norm $\|f\| := \left(\int |f|^p \, d\mu\right)^{1/p}$. Technically functions equal almost everywhere $d\mu$ must be identified.*

Example A.4 *The space $C_b(\Omega)$ of bounded continuous functions on Ω is a Banach space with norm given by $\|f\|_\infty := \sup_{x\in\Omega} |f(x)|$.*

Example A.5 *Once again let Ω be an open subset of $\mathbb{R}^n$. For each compact $K \subset\subset \Omega$ we have a seminorm on $C(\Omega)$ defined by $f \mapsto \|f\|_K := \sup_{x\in K} |f(x)|$. The corresponding convergence is the uniform convergence on compact subsets of Ω. It is often useful to notice that the same topology can be obtained by using $\|f\|_{K_i}$ obtained from a countable sequence of nested compact sets $K_1 \subset K_2 \subset \ldots$ such that*

$$\bigcup K_n = \Omega.$$

Such a sequence is called an exhaustion of Ω.

If we have topological vector space V and a closed subspace S, then we can form the quotient V/S. The quotient can be turned in to a normed space by introducing as norm

$$\|[x]\|_{\mathrm{V/S}} := \inf_{v\in[x]} \|v\|.$$

If S is not closed then this only defines a seminorm.

Theorem A.1 *If* V *is Banach space and a closed subspace* S *a closed (linear) subspace then* V/S *is a Banach space with the above defined norm.*

Proof. Let x_n be a sequence in V such that $[x_n]$ is a Cauchy sequence in V/S. Choose a subsequence such that $\|[x_n] - [x_{n+1}]\| \leq 1/2^n$ for $n = 1, 2, \ldots..$ Setting s_1 equal to zero we find $s_2 \in \mathrm{S}$ such that $\|x_1 - (x_2 + s_2)\|$ and continuing inductively define a sequence s_i such that such that $\{x_n + s_n\}$ is a Cauchy sequence in V. Thus there is an element $y \in \mathrm{V}$ with $x_n + s_n \to y$. But since the quotient map is norm decreasing the sequence $[x_n + s_n] = [x_n]$ must also converge;

$$[x_n] \to [y].$$

■

Remark A.1 *It is also true that if S is closed and* V/S *is a Banach space then so is* V.

A.2.1 Hilbert Spaces

Definition A.20 *A Hilbert space $\mathcal{H}$ is a complex vector space with a Hermitian inner product $\langle .,. \rangle$. A Hermitian inner product is a bilinear form with the following properties:*
1) $\langle v, w\rangle = \overline{\langle v, w\rangle}$
2) $\langle v, \alpha w_1 + \beta w_2\rangle = \alpha\langle v, w_1\rangle + \beta\langle v, w_2\rangle$
3) $\langle v, v\rangle \geq 0$ and $\langle v, v\rangle = 0$ only if $v = 0$.

One of the most fundamental properties of a Hilbert space is the projection property.

Theorem A.2 *If K is a convex, closed subset of a Hilbert space $\mathcal{H}$, then for any given $x \in \mathcal{H}$ there is a unique element $p_K(x) \in \mathcal{H}$ which minimizes the distance $\|x - y\|$ over $y \in K$. That is*

$$\|x - p_K(x)\| = \inf_{y\in K} \|x - y\|.$$

If K is a closed linear subspace then the map $x \mapsto p_K(x)$ is a bounded linear operator with the projection property $p_K^2 = p_K$.

Definition A.21 *For any subset $S \in \mathcal{H}$ we have the orthogonal compliment $S^\perp$ defined by*

$$S^\perp = \{x \in \mathcal{H} : \langle x, s\rangle = 0 \text{ for all } s \in S\}.$$

$S^\perp$ is easily seen to be a linear subspace of $\mathcal{H}$. Since $\ell_s : x \mapsto \langle x, s\rangle$ is continuous for all s and since

$$S^\perp = \cap_s \ell_s^{-1}(0)$$

we see that $S^\perp$ is closed. Now notice that since by definition

$$\|x - P_s x\|^2 \leq \|x - P_s x - \lambda s\|^2$$

for any $s \in S$ and any real λ we have $\|x - P_s x\|^2 \leq \|x - P_s x\|^2 - 2\lambda\langle x - P_s x, s\rangle + \lambda^2 \|s\|^2$. Thus we see that $p(\lambda) := \|x - P_s x\|^2 - 2\lambda\langle x - P_s x, s\rangle + \lambda^2 \|s\|^2$ is a polynomial in λ with a minimum at $\lambda = 0$. This forces $\langle x - P_s x, s\rangle = 0$ and so we see that $x - P_s x$. From this we see that any $x \in \mathcal{H}$ can be written as $x = x - P_s x + P_s x = s + s^\perp$. On the other hand it is easy to show that $S^\perp \cap S = 0$. Thus we have $\mathcal{H} = S \oplus S^\perp$ for any closed linear subspace $S \subset \mathcal{H}$. In particular the decomposition of any x as $s + s^\perp \in S \oplus S^\perp$ is unique.

Appendix B

The Language of Category Theory

Category theory provides a powerful means of organizing our thinking in mathematics. Some readers may be put off by the abstract nature of category theory. To such readers I can only say that it is not really difficult to catch on to the spirit of category theory and the payoff in terms of organizing mathematical thinking is considerable. I encourage these readers to give it a chance. In any case, it is not strictly necessary for the reader to be completely at home with category theory before going further into the book. In particular, physics and engineering students may not be used to this kind of abstraction and should simply try to gradually get used to the language of categories. Feel free to defer reading this appendix on Category theory until it seems necessary

Roughly speaking, category theory is an attempt at clarifying structural similarities that tie together different parts of mathematics. A category has "objects" and "morphisms". The prototypical category is just the category **Set** which has for its objects ordinary *sets* and for its morphisms *maps* between sets. The most important category for differential geometry is what is sometimes called the "smooth category" consisting of smooth manifolds and smooth maps. (The definition of these terms given in the text proper but roughly speaking smooth means differentiable.)

Now on to the formal definition of a category.

Definition B.1 *A* ***category*** $\mathfrak{C}$ *is a collection of objects* $\mathrm{Ob}(\mathfrak{C}) = \{X, Y, Z, ...\}$ *and for every pair of objects* X, Y *a set* $\mathrm{Hom}_{\mathfrak{C}}(X, Y)$ *called the set of* ***morphisms*** *from* X *to* Y*. The family of all such morphisms will be denoted* $\mathrm{Mor}(\mathfrak{C})$*. In addition, a category is required to have a* ***composition*** *law which is defined as a map* $\circ : \mathrm{Hom}_{\mathfrak{C}}(X, Y) \times \mathrm{Hom}_{\mathfrak{C}}(Y, Z) \to \mathrm{Hom}_{\mathfrak{C}}(X, Z)$ *such that for every three objects* $X, Y, Z \in Obj(\mathfrak{C})$ *the following axioms hold:*

Axiom B.1 (Cat1) $\mathrm{Hom}_{\mathfrak{C}}(X, Y)$ *and* $\mathrm{Hom}_{\mathfrak{C}}(Z, W)$ *are disjoint unless* $X = Z$ *and* $Y = W$ *in which case* $\mathrm{Hom}_{\mathfrak{C}}(X, Y) = \mathrm{Hom}_{\mathfrak{C}}(Z, W)$.

Axiom B.2 (Cat2) *The composition law is associative:* $f \circ (g \circ h) = (f \circ g) \circ h$.

Axiom B.3 (Cat3) *Each set of morphisms of the form* $\mathrm{Hom}_{\mathfrak{C}}(X, X)$ *must contain a necessarily element* id_X*, the identity element, such that* $f \circ \mathrm{id}_X = f$ *for any* $f \in \mathrm{Hom}_{\mathfrak{C}}(X, Y)$ *(and any* Y*), and* $\mathrm{id}_X \circ f = f$ *for any* $f \in \mathrm{Hom}_{\mathfrak{C}}(Y, X)$.

Notation B.1 *A morphism is sometimes written using an arrow. For example, if* $f \in \mathrm{Hom}_{\mathfrak{C}}(X, Y)$ *we would indicate this by writing* $f : X \to Y$ *or by* $X \xrightarrow{f} Y$.

The notion of category is typified by the case where the objects are sets and the morphisms are maps between the sets. In fact, subject to putting extra structure on the sets and the maps, this will be almost the only type of category we shall need to talk about. On the other hand there are plenty of interesting categories of this type. Examples include the following.

1. **Grp**: The objects are groups and the morphisms are group homomorphisms.
2. **Rng** : The objects are rings and the morphisms are ring homomorphisms.
3. **Lin**$_{\mathbb{F}}$: The objects are vector spaces over the field $\mathbb{F}$ and the morphisms are linear maps. This category is referred to as the linear category or the vector space category
4. **Top**: The objects are topological spaces and the morphisms are continuous maps.
5. **Man**^{r}**:** The category of C^r−differentiable manifolds and C^r−maps: One of the main categories discussed in this book. This is also called the smooth or differentiable category especially when $r = \infty$.

Notation B.2 *If for some morphisms* $f_i : X_i \to Y_i$ *,* $(i = 1, 2)$*,* $g_X : X_1 \to X_2$ *and* $g_Y : Y_1 \to Y_2$ *we have* $g_Y \circ f_1 = f_2 \circ g_X$ *then we express this by saying that the following diagram "commutes":*

$$\begin{array}{cccccc} & & f_1 & & \\ & X_1 & \rightarrow & Y_1 & \\ g_X & \downarrow & & \downarrow & g_Y \\ & X_2 & \rightarrow & Y_2 & \\ & & f_2 & & \end{array}$$

Similarly, if $h \circ f = g$ *we say that the diagram*

$$\begin{array}{cccc} & f & & \\ X & \rightarrow & Y & \\ & \searrow & \downarrow & h \\ g & & Z & \end{array}$$

commutes. More generally, tracing out a path of arrows in a diagram corresponds to composition of morphisms and to say that such a diagram commutes is to say that the compositions arising from two paths of arrows that begin and end at the same object are equal.

Definition B.2 *Suppose that $f : X \to Y$ is a morphism from some category $\mathfrak{C}$. If f has the property that for any two (parallel) morphisms $g_1, g_2 : Z \to X$ we always have that $f \circ g_1 = f \circ g_2$ implies $g_1 = g_2$, i.e. if f is "left cancellable", then we call f a **monomorphism**. Similarly, if $f : X \to Y$ is "right cancellable" we call f an **epimorphism**. A morphism that is both a monomorphism and an epimorphism is called an **isomorphism.** If the category needs to be specified then we talk about a $\mathfrak{C}$-**monomorphism,** $\mathfrak{C}$-**epimorphism** and so on).*

In some cases we will use other terminology. For example, an isomorphism in the smooth category is called a **diffeomorphism**. In the linear category, we speak of linear maps and linear isomorphisms. Morphisms which comprise $\mathrm{Hom}_{\mathfrak{C}}(X, X)$ are also called endomorphisms and so we also write $\mathrm{End}_{\mathfrak{C}}(X) := \mathrm{Hom}_{\mathfrak{C}}(X, X)$. The set of all isomorphisms in $\mathrm{Hom}_{\mathfrak{C}}(X, X)$ is sometimes denoted by $\mathrm{Aut}_{\mathfrak{C}}(X)$ and these morphisms are called **automorphisms**.

We single out the following: In many categories like the above we can form a sort of derived category that uses the notion of pointed space and pointed map. For example, we have the "pointed topological category" . A pointed topological space is an topological space X together with a distinguished point p. Thus a typical object in the pointed topological category would be written (X, p). A morphism $f : (X, p) \to (W, q)$ is a continuous map such that $f(p) = q$.

B.0.2 Functors

A **functor** F is a pair of maps both denoted by the same letter F that map objects and morphisms from one category to those of another

$$F : \mathrm{Ob}(\mathfrak{C}_1) \to \mathrm{Ob}(\mathfrak{C}_2)$$
$$F : \mathrm{Mor}(\mathfrak{C}_1) \to \mathrm{Mor}(\mathfrak{C}_2)$$

such that composition and identity morphisms are respected: This means that for a morphism $f : X \to Y$, the morphism

$$F(f) : F(X) \to F(Y)$$

is a morphism in the second category and we must have

1. $F(\mathrm{id}_{\mathfrak{C}_1}) = \mathrm{id}_{\mathfrak{C}_2}$
2. If $f : X \to Y$ and $g : Y \to Z$ then $F(f) : F(X) \to F(Y)$, $F(g) : F(Y) \to F(Z)$ and
$$F(g \circ f) = F(g) \circ F(f).$$

Example B.1 *Let* $\mathbf{Lin}_{\mathbb{R}}$ *be the category whose objects are real vector spaces and whose morphisms are real linear maps. Similarly, let* $\mathbf{Lin}_{\mathbb{C}}$ *be the category of complex vector spaces with complex linear maps. To each real vector space* V *we can associate the complex vector space* $\mathbb{C} \otimes_{\mathbb{R}} \mathrm{V}$ *called the complexification of* V *and to each linear map of real vector spaces* $\ell : \mathrm{V} \to \mathrm{W}$ *we associate the complex extension* $\ell_{\mathbb{C}} : \mathbb{C} \otimes_{\mathbb{R}} \mathrm{V} \to \mathbb{C} \otimes_{\mathbb{R}} \mathrm{W}$. *Here,* $\mathbb{C} \otimes_{\mathbb{R}} \mathrm{V}$ *is easily thought of as the vector space* V *where now complex scalars are allowed. Elements of* $\mathbb{C} \otimes_{\mathbb{R}} \mathrm{V}$ *are generated by elements of the form* $c \otimes v$ *where* $c \in \mathbb{C}$, $v \in \mathrm{V}$ *and we have* $i(c \otimes v) = ic \otimes v$ *where* $i = \sqrt{-1}$. *The map* $\ell_{\mathbb{C}} : \mathbb{C} \otimes_{\mathbb{R}} \mathrm{V} \to \mathbb{C} \otimes_{\mathbb{R}} \mathrm{W}$ *is defined by the requirement* $\ell_{\mathbb{C}}(c \otimes v) = c \otimes \ell v$. *Now the assignments*

$$\ell \mapsto \ell_{\mathbb{C}}$$
$$\mathrm{V} \mapsto \mathbb{C} \otimes_{\mathbb{R}} \mathrm{V}$$

define a functor from $\mathbf{Lin}_{\mathbb{R}}$ *to* $\mathbf{Lin}_{\mathbb{C}}$.

Remark B.1 *In practice, complexification amounts to simply allowing complex scalars. For instance, we might just write cv instead of* $c \otimes v$.

Actually, what we have defined here is a **covariant functor**. A **contravariant functor** is defined similarly except that the order of composition is reversed so that instead of **Funct2** above we would have $\mathcal{F}(g \circ f) = \mathcal{F}(f) \circ \mathcal{F}(g)$. An example of a contravariant functor is the dual vector space functor which is a functor from the category of vector spaces $\mathbf{Lin}_{\mathbb{R}}$ to itself which sends each space V to its dual V^* and each linear map to its dual (or transpose). Under this functor a morphism

$$\mathrm{V} \quad \overset{L}{\to} \quad \mathrm{W}$$

is sent to the morphism

$$\mathrm{V}^* \quad \overset{L^*}{\leftarrow} \quad \mathrm{W}^*$$

Notice the arrow reversal.

Remark B.2 *One of the most important functors for our purposes is the* ***tangent functor*** *defined in section* **??**. *Roughly speaking this functor replaces differentiable maps and spaces by their linear parts.*

Example B.2 *Consider the category of real vector spaces and linear maps. To every vector space* V *we can associate the dual of the dual* V^{**}. *This is a* covariant *functor which is the composition of the dual functor with itself:*

$$\begin{array}{ccccc} V & & W^* & & V^{**} \\ A \downarrow & \mapsto & A^* \downarrow & \mapsto & A^{**} \downarrow \\ W & & V^* & & W^{**} \end{array}$$

B.0.3 Natural transformations

Now suppose we have two functors

$$\begin{aligned} \digamma_1 &: Ob(\mathfrak{C}_1) \to Ob(\mathfrak{C}_2) \\ \digamma_1 &: Mor(\mathfrak{C}_1) \to Mor(\mathfrak{C}_2) \end{aligned}$$

and

$$\begin{aligned} \digamma_2 &: Ob(\mathfrak{C}_1) \to Ob(\mathfrak{C}_2) \\ \digamma_2 &: Mor(\mathfrak{C}_1) \to Mor(\mathfrak{C}_2) \end{aligned}$$

A natural transformation $\mathcal{T}$ from $\digamma_1$ to $\digamma_2$ is an assignment to each object X of $\mathfrak{C}_1$ a morphism $\mathcal{T}(X) : \digamma_1(X) \to \digamma_2(X)$ such that for every morphism $f : X \to Y$ of $\mathfrak{C}_1$, the following diagram commutes:

$$\begin{array}{ccccc} & & \mathcal{T}(X) & & \\ & \digamma_1(X) & \to & \digamma_2(X) & \\ \digamma_1(f) & \downarrow & & \downarrow & \digamma_2(f) \\ & \digamma_1(Y) & \to & \digamma_2(Y) & \\ & & \mathcal{T}(Y) & & \end{array}$$

A common first example is the natural transformation ι between the identity functor $I : \mathbf{Lin}_{\mathbb{R}} \to \mathbf{Lin}_{\mathbb{R}}$ and the double dual functor $** : \mathbf{Lin}_{\mathbb{R}} \to \mathbf{Lin}_{\mathbb{R}}$:

$$\begin{array}{ccccc} & & \iota(\mathrm{V}) & & \\ & \mathrm{V} & \to & \mathrm{V}^{**} & \\ f & \downarrow & & \downarrow & f^{**} \\ & \mathrm{W} & \to & \mathrm{W}^{**} & \\ & & \iota(\mathrm{W}) & & \end{array} .$$

The map $\mathrm{V} \to \mathrm{V}^{**}$ sends a vector to a linear function $\widetilde{v} : \mathrm{V}^* \to \mathbb{R}$ defined by $\widetilde{v}(\alpha) := \alpha(v)$ (the hunter becomes the hunted so to speak). If there is an inverse natural transformation $\mathcal{T}^{-1}$ in the obvious sense, then we say that $\mathcal{T}$ is a natural isomorphism and for any object $X \in \mathfrak{C}_1$ we say that $\digamma_1(X)$ is naturally isomorphic to $\digamma_2(X)$. The natural transformation just defined is easily checked to have an inverse so is a natural isomorphism. The point here is not just that V is isomorphic to V^{**} in the category $\mathbf{Lin}_{\mathbb{R}}$ but that the isomorphism exhibited is natural. It works for all the spaces V in a uniform way that involves no special choices. This is to be contrasted with the fact that V is isomorphic to V^* where the construction of such an isomorphism involves an arbitrary choice of a basis.

Appendix C

Primer for Manifold Theory

After imposing rectilinear coordinates on a Euclidean space E^n (such as the plane E^2) we identify Euclidean space with $\mathbb{R}^n$, the vector space of $n-$tuples of numbers. In fact, since a Euclidean space in this sense is an object of intuition (at least in 2d and 3d) some reader may insist that to be sure such a space of points really exists that we should in fact start with $\mathbb{R}^n$ and "forget" the origin and all the vector space structure while retaining the notion of point and distance. The coordinatization of Euclidean space is then just a "remembering" of this forgotten structure. Thus our coordinates arise from a map $x : E^n \to \mathbb{R}^n$ which is just the identity map. This approach has much to recommend it but there is at least one regrettable aspect to this approach which is the psychological effect that occurs when we impose other coordinates on our system and then introduce differentiable manifolds as abstract geometric objects that support coordinate systems. It might seem that this is a big abstraction and when the definitions of charts and atlases and so on appear a certain notational fastidiousness sets in that somehow creates a psychological gap between open set in $\mathbb{R}^n$ and the abstract space that we coordinatize. But what is now lost from sight is that we have already been dealing with an abstract manifold! Namely, E^3 which support many coordinate systems such as spherical coordinates. Putting coordinates on space, even the rectangular coordinates which allows us to identify E^3 with $\mathbb{R}^3$ is already the basic idea involved in the notion of a differentiable manifold. The idea of a differentiable manifold is a natural idea that becomes over complicated when we are forced to make exact definitions. As a result of the nature of these definition the student is faced with a pedagogy that teaches notation, trains one to examine each expression for logical set theoretic self consistency, but fails to teach geometric intuition. Having made this complaint, the author must confess that he too will use the modern notation and will not stray far from standard practice. These remarks are meant to encourage the student to stop and seek the simplest most intuitive viewpoint whenever feeling overwhelmed by notation. The student is encouraged to experiment with abbreviated personal notation when checking calculations and to draw diagrams and schematics that encode the geometric ideas whenever possible. "The picture writes the equations".

So, as we said, after imposing rectilinear coordinates on a Euclidean space E^n (such as the plane E^2 space E^3) we identify Euclidean space with $\mathbb{R}^n$, the vector space of $n-$tuples of numbers.

We will envision there to be a copy of $\mathbb{R}^n$ at each of its points $p \in \mathbb{R}^n$ which is denoted $\mathbb{R}^n_p$. The elements of $\mathbb{R}^n_p$ are to be thought of as vectors based at p, that is, the "tangent vectors" at p. These tangent spaces are related to each other by the obvious notion of vectors being parallel (this is exactly what is not generally possible for tangents spaces of a general manifold). For the standard basis vectors e_j (relative to the coordinates x_i) taken as being based at p we often write $\frac{\partial}{\partial x_i}\Big|_p$ and this has the convenient second interpretation as a differential operator acting on smooth functions defined near $p \in \mathbb{R}^n$. Namely,

$$\frac{\partial}{\partial x_i}\Big|_p f = \frac{\partial f}{\partial x_i}(p).$$

An n-tuple of smooth functions $X^1, ..., X^n$ defines a smooth vector field $X = \sum X^i \frac{\partial}{\partial x_i}$ whose value at p is $\sum X^i(p) \left.\frac{\partial}{\partial x_i}\right|_p$. Thus a vector field assigns to each p in its domain, an open set U, a vector $\sum X^i(p) \left.\frac{\partial}{\partial x_i}\right|_p$ at p. We may also think of vector field as a differential operator via

$$f \mapsto Xf \in C^\infty(U)$$
$$(Xf)(p) := \sum X^i(p) \frac{\partial f}{\partial x_i}(p)$$

Example C.1 *$X = y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y}$ is a vector field defined on $U = \mathbb{R}^2 - \{0\}$ and $(Xf)(x, y) = y\frac{\partial f}{\partial x}(x, y) - x\frac{\partial f}{\partial y}(x, y)$.*

Notice that we may certainly add vector fields defined over the same open set as well as multiply by functions defined there:

$$(fX + gY)(p) = f(p)X(p) + g(p)X(p)$$

The familiar expression $df = \frac{\partial f}{\partial x_1}dx_1 + \cdots + \frac{\partial f}{\partial x_n}dx_n$ has the intuitive interpretation expressing how small changes in the variables of a function give rise to small changes in the value of the function. Two questions should come to mind. First, "what does 'small' mean and how small is small enough?" Second, "which direction are we moving in the coordinate" space? The answer to these questions lead to the more sophisticated interpretation of df as being a linear functional on each tangent space. Thus we must choose a direction v_p at $p \in \mathbb{R}^n$ and then $df(v_p)$ is a number depending linearly on our choice of vector v_p. The definition is determined by $dx_i(e_j) = \delta_{ij}$. In fact, this shall be the basis of our definition of df at p. We want

$$Df|_p \left(\frac{\partial}{\partial x_i}\Big|_p \right) := \frac{\partial f}{\partial x_i}(p).$$

Now any vector at p may be written $v_p = \sum_{i=1}^{n} v^i \left.\frac{\partial}{\partial x_i}\right|_p$ which invites us to use v_p as a differential operator (at p):

$$v_p f := \sum_{i=1}^{n} v^i \frac{\partial f}{\partial x_i}(p) \in \mathbb{R}$$

This consistent with our previous statement about a vector field being a differential operator simply because $X(p) = X_p$ is a vector at p for every $p \in U$. This is just the directional derivative. In fact we also see that

$$\begin{aligned} Df|_p(v_p) &= \sum_j \frac{\partial f}{\partial x_j}(p) dx_j \left(\sum_{i=1}^{n} v^i \left.\frac{\partial}{\partial x_i}\right|_p \right) \\ &= \sum_{i=1}^{n} v^i \frac{\partial f}{\partial x_i}(p) = v_p f \end{aligned}$$

so that our choices lead to the following definition:

Definition C.1 *Let f be a smooth function on an open subset U of $\mathbb{R}^n$. By the symbol df we mean a family of maps $Df|_p$ with p varying over the domain U of f and where each such map is a linear functional of tangent vectors based at p given by $Df|_p(v_p) = v_p f = \sum_{i=1}^{n} v^i \frac{\partial f}{\partial x_i}(p)$.*

Definition C.2 *More generally, a smooth 1-form α on U is a family of linear functionals $\alpha_p : T_p\mathbb{R}^n \to \mathbb{R}$ with $p \in U$ that is smooth is the sense that $\alpha_p(\left.\frac{\partial}{\partial x_i}\right|_p)$ is a smooth function of p for all i.*

¿From this last definition it follows that if $X = X^i \frac{\partial}{\partial x_i}$ is a smooth vector field then $\alpha(X)(p) := \alpha_p(X_p)$ defines a smooth function of p. Thus an alternative way to view a $1-$form is as a map $\alpha : X \mapsto \alpha(X)$ that is defined on vector fields and linear over the algebra of smooth functions $C^\infty(U)$:

$$\alpha(fX + gY) = f\alpha(X) + g\alpha(Y).$$

C.0.4 Fixing a problem

Now it is at this point that we want to destroy the privilege of the rectangular coordinates and express our objects in an arbitrary coordinate system smoothly related to the existing coordinates. This means that for any two such coordinate systems, say $u^1, ..., u^n$ and $y^1, ..., y^n$ we want to have the ability to express fields and forms in either system and have for instance

$$X^i_{(y)} \frac{\partial}{\partial y_i} = X = X^i_{(u)} \frac{\partial}{\partial u_i}$$

for appropriate functions $X^i_{(y)}$, $X^i_{(u)}$. This equation only makes sense on the overlap of the domains of the coordinate systems. To be consistent with the chain rule we must have

$$\frac{\partial}{\partial y^i} = \frac{\partial u^j}{\partial y^i}\frac{\partial}{\partial u^j}$$

which then forces the familiar transformation law:

$$\sum \frac{\partial u^j}{\partial y^i} X^i_{(y)} = X^i_{(u)}$$

We think of $X^i_{(y)}$ and $X^i_{(u)}$ as referring to or representing the same geometric reality from two different coordinate systems. No big deal right? OK, how about the fact that there is this underlying abstract space that we are coordinatizing? That too is no big deal. We were always doing it in calculus anyway. What about the fact that the coordinate systems aren't defined as a 1-1 correspondence with the points of the space unless we leave out some points in some coordinates. For example, when using polar coordinates we leave out the origin and the axis we are measuring from to avoid ambiguity in θ and in order to have a nice open domain. Well if this is all fine then we may as well imagine other abstract spaces that support coordinates in this way. In fact, we don't have to look far for an example. Any surface such as the sphere will do. We can talk about 1-forms like say $\alpha = \theta d\phi + \phi \sin(\theta) d\theta$, or a vector field tangent to the sphere $\theta \sin(\phi)\frac{\partial}{\partial \theta} + \theta^2 \frac{\partial}{\partial \phi}$ and so on (just pulling things out of a hat). We just have to be clear about how these arise and most of all how to change to a new coordinate expression for the same object in a different coordinate system. This is the approach of tensor analysis. An object called a 2-tensor T is represented in two different coordinate systems as for instance

$$T = \sum T^{ij}_{(y)} \frac{\partial}{\partial y^i} \otimes \frac{\partial}{\partial y^j}$$
$$T = \sum T^{ij}_{(u)} \frac{\partial}{\partial u^i} \otimes \frac{\partial}{\partial u^j}$$

where all we really need to know for many purposes is the transformation law for coordinate changes:

$$T^{ij}_{(y)} = \sum_{r,s} T^{rs}_{(u)} \frac{\partial y^i}{\partial u^r}\frac{\partial y^i}{\partial u^s}.$$

Then either expression is referring to the same abstract tensor T. This is just a preview but it highlights the approach wherein a transformation laws play a defining role. Eventually, this leads to the abstract notion of a G-bundle.

Appendix D

Modules, Multilinear Algebra

A great many people think they are thinking when they are merely rearranging their prejudices.

-William James (1842-1910)

Synopsis: Multilinear maps, tensors, tensor fields.

The set of all vector fields on a manifold is a vector space over the real numbers but not only can we add vector fields and scale by numbers but we may also scale by smooth functions. We say that the vector fields form a module over the ring of smooth functions. A module is similar to a vector space with the differences stemming from the use of elements of a ring R as the scalars rather than the field of complex $\mathbb{C}$ or real numbers $\mathbb{R}$. For a module, one still has $1w = w$, $0w = 0$ and $-1w = -w$. Of course, every vector space is also a module since the latter is a generalization of the notion of vector space.

Definition D.1 *Let* R *be a ring. A left* R-***module*** *(or a left* ***module*** *over* R*) is an abelian group* $W, +$ *together with an operation* $\mathsf{R} \times W \to W$ *written* $(a, w) \mapsto aw$ *such that*

1) $(a + b)w = aw + bw$ *for all* $a, b \in \mathsf{R}$ *and all* $w \in W$*,*

2) $a(w_1 + w_2) = aw_1 + aw_2$ *for all* $a \in \mathsf{R}$ *and all* $w_2, w_1 \in W$*.*

A right R-***module*** *is defined similarly with the multiplication of the right so that*

1) $w(a + b) = wa + wb$ *for all* $a, b \in \mathsf{R}$ *and all* $w \in W$*,*

2) $(w_1 + w_2)a = w_1a + w_2a$ *for all* $a \in \mathsf{R}$ *and all* $w_2, w_1 \in W$*. .*

If the ring is commutative (the usual case for us) then we may right $aw = wa$ and consider any right module as a left module and visa versa. Even if the ring is not commutative we will usually stick to left modules and so we drop the reference to "left" and refer to such as R-modules.

Remark D.1 *We shall often refer to the elements of* R *as* ***scalars****.*

Example D.1 *An abelian group* $A, +$ *is a* $\mathbb{Z}$ *module and a* $\mathbb{Z}$*-module is none other than an abelian group. Here we take the product of* $n \in \mathbb{Z}$ *with* $x \in A$ *to be* $nx := x + \cdots + x$ *if* $n \geq 0$ *and* $nx := -(x + \cdots + x)$ *if* $n < 0$ *(in either case we are adding* $|n|$ *terms).*

Example D.2 *The set of all* $m \times n$ *matrices with entries being elements of a commutative ring* R *(for example real polynomials) is an* R*-module.*

Example D.3 *The module of all module homomorphisms of a module* W *onto another module* M *is a module and is denoted* $\mathrm{Hom}_{\mathsf{R}}(\mathrm{W}, \mathrm{M})$ *or* $L_{\mathsf{R}}(\mathrm{W}, \mathrm{M})$.

Example D.4 *Let* V *be a vector space and* $\ell : \mathrm{V} \to \mathrm{V}$ *a linear operator. Using this one operator we may consider* V *as a module over the ring of polynomials* $\mathbb{R}[t]$ *by defining the "scalar" multiplication by*

$$p(t)v := p(\ell)v$$

for $p \in \mathbb{R}[t]$.

Since the ring is usually fixed we often omit mention of the ring. In particular, we often abbreviate $L_{\mathsf{R}}(\mathrm{W}, \mathrm{M})$ to $L(\mathrm{W}, \mathrm{M})$. Similar omissions will be made without further mention. Also, since every real (resp. complex) Banach space E is a vector space and hence a module over $\mathbb{R}$ (resp. $\mathbb{C}$) we must distinguish between the bounded linear maps which we have denoted up until now as $L(\mathsf{E}; \mathsf{F})$ and the linear maps that would be denoted the same way in the context of modules. Our convention will be the following:

Definition D.2 ((convention)) *In case the modules in question are presented as infinite dimensional topological vector spaces, say* E *and* F *we will let* $L(\mathsf{E}; \mathsf{F})$ *continue to mean the space of bounded linear operator unless otherwise stated.*

A **submodule** is defined in the obvious way as a subset $S \subset \mathrm{W}$ that is closed under the operations inherited from W so that S itself is a module. The intersection of all submodules containing a subset $A \subset \mathrm{W}$ is called the **submodule generated by** A and is denoted $\langle A \rangle$ and A is called a **generating set**. If $\langle A \rangle = \mathrm{W}$ for a finite set A, then we say that W is finitely generated.

Let S be a submodule of W and consider the quotient abelian group W/S consisting of cosets, that is sets of the form $[v] := v + S = \{v + x : x \in S\}$ with addition given by $[v] + [w] = [v + w]$. We define a scalar multiplication by elements of the ring R by $a[v] := [av]$ respectively. In this way, W/S is a module called a quotient module.

Definition D.3 *Let* W_1 *and* W_2 *be modules over a ring* R. *A map* $L : W_1 \to W_2$ *is called* ***module homomorphism*** *if*

$$L(aw_1 + bw_2) = aL(w_1) + bL(w_2).$$

By analogy with the case of vector spaces, which module theory includes, we often characterize a module homomorphism L *by saying that* L *is* ***linear over*** R.

A real (resp. complex) vector space is none other than a module over the field of real numbers $\mathbb{R}$ (resp. complex numbers $\mathbb{C}$). In fact, most of the modules we encounter will be either vector spaces or spaces of sections of some vector bundle.

Many of the operations that exist for vector spaces have analogues in the modules category. For example, the direct sum of modules is defined in the obvious way. Also, for any module homomorphism $L : \mathrm{W}_1 \to \mathrm{W}_2$ we have the usual notions of kernel and image:

$$\begin{aligned} \ker L &= \{v \in \mathrm{W}_1 : L(v) = 0\} \\ \mathrm{img}(L) &= L(\mathrm{W}_1) = \{w \in \mathrm{W}_2 : w = Lv \text{ for some } v \in \mathrm{W}_1\}. \end{aligned}$$

These are submodules of W_1 and W_2 respectively.

On the other hand, modules are generally not as simple to study as vector spaces. For example, there are several notions of dimension. The following notions for a vector space all lead to the same notion of dimension. For a completely general module these are all potentially different notions:

1. Length of the longest chain of submodules

$$0 = \mathrm{W}_n \subsetneq \cdots \subsetneq \mathrm{W}_1 \subsetneq \mathrm{W}$$

2. The cardinality of the largest linearly independent set (see below).

3. The cardinality of a basis (see below).

For simplicity in our study of dimension, let us now assume that R is commutative.

Definition D.4 *A set of elements $w_1, ..., w_k$ of a module are said to be* ***linearly dependent*** *if there exist ring elements $r_1, ..., r_k \in \mathsf{R}$ not all zero, such that $r_1 w_1 + \cdots + r_k w_k = 0$. Otherwise, they are said to be* ***linearly independent****. We also speak of the set $\{w_1, ..., w_k\}$ as being a* ***linearly independent set****.*

So far so good but it is important to realize that just because $w_1, ..., w_k$ are linearly independent doesn't mean that we may write each of these w_i as a linear combination of the others. It may even be that some element w forms a linearly dependent set since there may be a nonzero r such that $rw = 0$ (such a w is said to have **torsion**).

If a linearly independent set $\{w_1, ..., w_k\}$ is maximal in size then we say that the module has **rank** k. Another strange possibility is that a maximal linearly independent set may not be a generating set for the module and hence may not be a basis in the sense to be defined below. The point is that although for an arbitrary $w \in \mathrm{W}$ we must have that $\{w_1, ..., w_k\} \cup \{w\}$ is linearly dependent and hence there must be a nontrivial expression $rw + r_1 w_1 + \cdots + r_k w_k = 0$, it does **not** follow that we may solve for w since r may not be an invertible element of the ring (i.e. it may not be a **unit**).

Definition D.5 *If B is a generating set for a module* W *such that every element of* W *has a unique expression as a* finite R*-linear combination of elements of B then we say that B is a **basis** for* W.

Definition D.6 *If an R-module has a basis then it is referred to as a **free module**.*

If a module over a (commutative) ring R has a basis then the number of elements in the basis is called the **dimension** and must in this case be the same as the rank (the size of a maximal linearly independent set). If a module W is free with basis $w_1, ..., w_n$ then we have an isomorphism $\mathsf{R}^n \cong \mathrm{W}$ given by

$$(r_1, ..., r_n) \mapsto r_1 w_1 + \cdots + r_n w_n.$$

Exercise D.1 *Show that every finitely generated R-module is the homomorphic image of a free module.*

It turns out that just as for vector spaces the cardinality of a basis for a free module W is the same as that of every other basis for W. The cardinality of any basis for a free module W is called the dimension of W. If R is a field then every module is free and is a vector space by definition. In this case, the current definitions of dimension and basis coincide with the usual ones.

The ring R is itself a free R-module with standard basis given by $\{1\}$. Also, $\mathsf{R}^n := \mathsf{R} \times \cdots \times \mathsf{R}$ is a free module with standard basis $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$ where, as usual $\mathbf{e}_i := (0, ..., 1, ..., 0)$; the only nonzero entry being in the i-th position.

Definition D.7 *Let k be a commutative ring, for example a field such as $\mathbb{R}$ or $\mathbb{C}$. A ring A is called a k-**algebra** if there is a ring homomorphism $\mu : \mathsf{k} \to \mathsf{R}$ such that the image $\mu(\mathsf{k})$ consists of elements that commute with everything in A. In particular, A is a module over k.*

Example D.5 *The ring $\mathcal{C}^\infty_M(U)$ is an $\mathbb{R}$-algebra.*

We shall have occasion to consider A-modules where A is an algebra over some k. In this context the elements of A are still called scalars but the elements of $\mathsf{k} \subset \mathsf{A}$ will be referred to as **constants**.

Example D.6 *For an open set $U \subset M$ the set vector fields $\mathfrak{X}_M(U)$ is a vector space over $\mathbb{R}$ but it is also a module over the $\mathbb{R}$-algebra $\mathcal{C}^\infty_M(U)$. So for all $X, Y \in \mathfrak{X}_M(U)$ and all $f, g \in \mathcal{C}^\infty_M(U)$ we have*

1. $f(X + Y) = fX + fY$
2. $(f + g)X = fX + gX$
3. $f(gX) = (fg)X$

Similarly, $\mathfrak{X}_M^*(U) = \Gamma(U, T^*M)$ is also a module over $\mathcal{C}_M^\infty(U)$ that is naturally identified with the module dual $\mathfrak{X}_M(U)^*$ by the pairing $(\theta, X) \mapsto \theta(X)$. Here $\theta(X) \in \mathcal{C}_M^\infty(U)$ and is defined by $p \mapsto \theta_p(X_p)$. The set of all vector fields $\mathcal{Z} \subset \mathfrak{X}(U)$ that are zero at a fixed point $p \in U$ is a submodule in $\mathfrak{X}(U)$. If $U, (x^1, ..., x^n)$ is a coordinate chart then the set of vector fields

$$\frac{\partial}{\partial x^1}, ..., \frac{\partial}{\partial x^n}$$

is a basis (over $\mathcal{C}_M^\infty(U)$) for the module $\mathfrak{X}(U)$. Similarly,

$$dx^1, ..., dx^n$$

is a basis for $\mathfrak{X}_M^*(U)$. It is important to realize that if U is not a coordinate chart domain then it may be that $\mathfrak{X}_M(U)$ and $\mathfrak{X}_M(U)^*$ have no basis. In particular, we should not expect $\mathfrak{X}(M)$ to have a basis in the general case.

Example D.7 *The sections of any vector bundle over a manifold M form a $C^\infty(M)$-module denoted $\Gamma(E)$. Let $E \to M$ be a trivial vector bundle of finite rank n. Then there exists a basis of vector fields $E_1, ..., E_n$ for the module $\Gamma(E)$. Thus for any section X there exist unique functions f^i such that*

$$X = \sum f^i E_i.$$

In fact, since E is trivial we may as well assume that $E = M \times \mathbb{R}^n \overset{pr_1}{\to} M$. Then for any basis $e_1, ..., e_n$ for $\mathbb{R}^n$ we may take

$$E_i(x) := (x, e_i).$$

Definition D.8 *Let V_i, $i = 1, ..., k$ and W be modules over a ring R. A map $\mu : \mathrm{V}_1 \times \cdots \times \mathrm{V}_k \to \mathrm{W}$ is called **multilinear** (k-multilinear) if for each i, $1 \leq i \leq k$ and each fixed $(w_1, ..., \widehat{w_i}, ..., w_k) \in \mathrm{V}_1 \times \cdots \times \widehat{\mathrm{V}_1} \times \cdots \times \mathrm{V}_k$ we have that the map*

$$v \mapsto \mu(w_1, ..., \underset{i-th}{v}, ..., w_k),$$

obtained by fixing all but the i-th variable, is a module homomorphism. In other words, we require that μ be R- linear in each slot separately. The set of all multilinear maps $\mathrm{V}_1 \times \cdots \times \mathrm{V}_k \to \mathrm{W}$ is denoted $L_{\mathsf{R}}(\mathrm{V}_1, ..., \mathrm{V}_k; \mathrm{W})$. If $\mathrm{V}_1 = \cdots = \mathrm{V}_k = \mathrm{V}$ then we abbreviate this to $L^k_{\mathsf{R}}(\mathrm{V}; \mathrm{W})$.

The space of multilinear maps $L_{\mathsf{R}}(\mathrm{V}_1, ..., \mathrm{V}_k; \mathrm{W})$ is itself an R-module in a fairly obvious way. For $a, b \in \mathsf{R}$ and $\mu_1, \mu_2 \in L_{\mathsf{R}}(\mathrm{V}_1, ..., \mathrm{V}_k; \mathrm{W})$ then $a\mu_1 + b\mu_2$ is defined in the usual way.

Purely algebraic results

In this section we intend all modules to be treated strictly as modules. Thus we do not require multilinear maps to be bounded. In particular, $L_{\mathsf{R}}(\mathrm{V}_1, ..., \mathrm{V}_k; \mathrm{W})$ does not refer to bounded multilinear maps even if the modules are coincidentally Banach spaces. We shall comment on how thing look in the Banach space category in a later section.

Definition D.9 *The dual of an* $\mathsf{R}-$*module* W *is the module* $\mathrm{W}^* := \mathrm{Hom}_{\mathsf{R}}(\mathrm{W},\mathsf{R})$ *of all* R*-linear functionals on* W*.*

Any element of W can be though of as an element of $\mathrm{W}^{**} := \mathrm{Hom}_{\mathsf{R}}(\mathrm{W}^*,\mathsf{R})$ according to $w(\alpha) := \alpha(w)$. This provides a map $\mathrm{W} \hookrightarrow \mathrm{W}^{**}$ an if this map is an isomorphism then we say that W is **reflexive** .

If W is reflexive then we are free to identify W with W^{**}.

Exercise D.2 *Show that if* W *is a free with finite dimension then* W *is reflexive. We sometimes write* $w\lrcorner(\alpha) = \langle w, \alpha\rangle = \langle \alpha, w\rangle$*.*

There is a bit of uncertainty about how to use the word "tensor". On the one hand, a tensor is a certain kind of multilinear mapping. On the other hand, a tensor is an element of a tensor product (defined below) of several copies of a module and its dual. For finite dimensional vector spaces these two viewpoints turn out to be equivalent as we shall see but since we are also interested in infinite dimensional spaces we must make a terminological distinction. We make the following slightly nonstandard definition:

Definition D.10 *Let* V *and* W *be* R*-modules. A* W*-valued* $({}^{r}{}_{s})$*-**tensor map** on* V *is a multilinear mapping of the form*

$$\Lambda : \underbrace{\mathrm{V}^* \times \mathrm{V}^* \cdots \times \mathrm{V}^*}_{r-\text{times}} \times \underbrace{\mathrm{V} \times \mathrm{V} \times \cdots \times \mathrm{V}}_{s-\text{times}} \to \mathrm{W}.$$

The set of all tensor maps into W will be denoted $T^{r}{}_{s}(\mathrm{V};\mathrm{W})$. Similarly, a W-valued $({}_{s}{}^{r})$-**tensor map** on V is a multilinear mapping of the form

$$\Lambda : \underbrace{\mathrm{V} \times \mathrm{V} \times \cdots \times \mathrm{V}}_{s-\text{times}} \times \underbrace{\mathrm{V}^* \times \mathrm{V}^* \cdots \times \mathrm{V}^*}_{r-\text{times}} \to \mathrm{W}$$

and the corresponding space of all such is denoted $T_{s}{}^{r}(\mathrm{V};\mathrm{W})$.

There is, of course, a natural isomorphism $T_{s}{}^{r}(\mathrm{V};\mathrm{W}) \cong T^{r}{}_{s}(\mathrm{V};\mathrm{W})$ induced by the map $\mathrm{V}^{s} \times \mathrm{V}^{*r} \cong \mathrm{V}^{*r} \times \mathrm{V}^{s}$ given on homogeneous elements by $v \otimes \omega \mapsto \omega \otimes v$. (Warning) In the presence of an inner product there is another possible isomorphism here given by $v \otimes \omega \mapsto \flat v \otimes \sharp\omega$. **This map is a "transpose" map and just as we do not identify a matrix with its transpose we do not generally identify individual elements under this isomorphism.**

Remark D.2 *The reader may have wondered about the possibly of multilinear maps were the covariant and contravariant variables are interlaced such as* $\Upsilon : \mathrm{V} \times \mathrm{V}^* \times \mathrm{V} \times \mathrm{V}^* \times \mathrm{V}^* \to \mathrm{W}$*. Of course, such things exist and this example would be an element of what we might denote by* $T_{1}{}^{1}{}_{1}{}^{2}(\mathrm{V};\mathrm{W})$*. But we can agree to associate to each such object an unique element of* $T^{3}{}_{2}(\mathrm{V};\mathrm{W})$ *by simple keeping the order among the covariant variable and among the contravariant variable but shifting all covariant variables to the left of the contravariant variables. Some authors have insisted on the need to avoid this consolidation for reasons which we will show to be unnecessary below.*

Notation D.1 *For the most part we shall be needing only* $T^{r}{}_{s}(\mathrm{V};\mathrm{W})$ *and so we agree abbreviate this to* $T^r_s(\mathrm{V};\mathrm{W})$ *and call the elements* (r,s)*-**tensor maps**. So by convention*

$$T^r_s(\mathrm{V};\mathrm{W}) := T^{r}{}_{s}(\mathrm{V};\mathrm{W})$$
$$\text{but}$$
$$T^r_s(\mathrm{V};\mathrm{W}) \neq T_{r}{}^{s}(\mathrm{V};\mathrm{W}).$$

Elements of $T^r_0(\mathrm{V})$ are said to be of **contravariant type** and of degree r and those in $T^0_s(\mathrm{V})$ are of **covariant type** (and degree s). If $r,s>0$ then the elements of $T^r_s(\mathrm{V})$ are called **mixed tensors (of tensors of mixed type)** with contravariant degree r and covariant degree s.

Remark D.3 *An* $\mathbb{R}$ *-valued* (r,s)*-**tensor map** is usually just called an* (r,s)*-**tensor** but as we mentioned above, there is another common meaning for this term which is equivalent in the case of finite dimensional vector spaces. The word* tensor *is also used to mean "tensor field" (defined below). The context will determine the proper meaning.*

Remark D.4 *The space* $T^r_s(\mathrm{V};\mathsf{R})$ *is sometimes denoted by* $T^r_s(\mathrm{V};\mathsf{R})$ *(or even* $T^r_s(\mathrm{V})$ *in case* $\mathsf{R}=\mathbb{R}$*) but we reserve this notation for another space defined below which is canonically isomorphic to* $T^r_s(\mathrm{V};\mathsf{R})$ *in case* V *is free with finite dimension.*

Definition D.11 *If* $\Upsilon_1 \in T^{k_1}_{l_1}(\mathrm{V};\mathsf{R})$ *and* $\Upsilon_2 \in T^{k_2}_{l_2}(\mathrm{V};\mathsf{R})$ *then* $\Upsilon_1 \otimes \Upsilon_2 \in T^{k_1+k_2}_{l_1+l_2}(\mathrm{V};\mathsf{R})$ *where*

$$\begin{aligned}&(\Upsilon_1 \otimes \Upsilon_2)(\theta_1,...,\theta_{k_1+k_2},\mathrm{v}_1,...,\mathrm{v}_{l_1+l_2})\\ &=\Upsilon_1(\theta_1,...,\theta_{k_1},\mathrm{v}_1,,...,\mathrm{v}_{l_1})\Upsilon_2(\theta_{k_1+1},...,\theta_{k_2},\mathrm{v}_{l_1+1},,...,\mathrm{v}_{l_2}).\end{aligned}$$

Remark D.5 *We call this type of tensor product the **"map" tensor product** in case we need to distinguish it from the tensor product defined below.*

Now suppose that V is free with finite dimension n. Then there is a basis $f_1,...,f_n$ for V with dual basis $f^1,...,f^n$. Now we have $\mathrm{V}^* = T^0_1(\mathrm{V};\mathsf{R})$. Also, we may consider $f_i \in \mathrm{V}^{**} = T^0_1(\mathrm{V}^*;\mathsf{R})$ and then, as above, take tensor products to get elements of the form

$$f_{i_1}\otimes\cdots\otimes f_{i_r}\otimes f^{j_1}\otimes\cdots\otimes f^{j_s}.$$

These are multilinear maps in $T^r_s(\mathrm{V};\mathsf{R})$ by definition:

$$\begin{aligned}&(f_{i_1}\otimes\cdots\otimes f_{i_r}\otimes f^{j_1}\otimes\cdots\otimes f^{j_s})(\alpha_1,...,\alpha_r,v_1,...,v_s)\\ &=\alpha_1(f_{i_1})\cdots\alpha_r(f_{i_r})f^{j_1}(v_1)\cdots f^{j_s}(v_s).\end{aligned}$$

There are $n^s n^r$ such maps that together form a basis for $T^r_s(\mathrm{V};\mathsf{R})$ which is therefore also free. Thus we may write any tensor map $\Upsilon \in T^r_s(\mathrm{V};\mathsf{R})$ as a sum

$$\Upsilon = \sum \Upsilon^{i_1\ldots i_r}{}_{j_1,\ldots j_s}\, f_{i_1}\otimes\cdots\otimes f_{i_r}\otimes f^{j_1}\otimes\cdots\otimes f^{j_s}$$

and the scalars $\Upsilon^{i_1 \ldots i_r}{}_{j_1,\ldots j_s} \in \mathsf{R}$

We shall be particularly interested in the case where all the modules are real (or complex) vector spaces such as the tangent space at a point on a smooth manifold. As we mention above, we will define a space $T^r_s(\mathsf{V};\mathsf{R})$ for each (r, s) that is canonically isomorphic to $T^r_s(\mathsf{V};\mathsf{R})$. There will be a product $\otimes T^r_s(\mathsf{V};\mathsf{R}) \times T^p_q(\mathsf{V};\mathsf{R}) \to T^{r+p}_{s+q}(\mathsf{V};\mathsf{R})$ for these spaces also and this will match up with the current definition under the canonical isomorphism.

Example D.8 *The inner product (or "dot product") on the Euclidean vector space* $\mathbb{R}^n$ *given for vectors* $\tilde{\mathrm{v}} = (v_1, ..., v_n)$ *and* $\tilde{\mathrm{w}} = (w_1, ..., w_n)$ *by*

$$(\vec{v}, \vec{w}) \mapsto \langle \vec{v}, \vec{w} \rangle = \sum_{i=1}^{n} v_i w_i$$

is 2*-multilinear (more commonly called* ***bilinear****).*

Example D.9 *For any* n *vectors* $\tilde{\mathrm{v}}_1, ..., \tilde{\mathrm{v}}_n \in \mathbb{R}^n$ *the determinant of* $(\tilde{\mathrm{v}}_1, ..., \tilde{\mathrm{v}}_n)$ *is defined by considering* $\tilde{\mathrm{v}}_1, ..., \tilde{\mathrm{v}}_n$ *as columns and taking the determinant of the resulting* $n \times n$ *matrix. The resulting function is denoted* $\det(\tilde{\mathrm{v}}_1, ..., \tilde{\mathrm{v}}_n)$ *and is* n*-multilinear.*

Example D.10 *Let* $\mathfrak{X}(M)$ *be the* $C^\infty(M)$*-module of vector fields on a manifold* M *and let* $\mathfrak{X}^*(M)$ *be the* $C^\infty(M)$*-module of 1-forms on* M*. The map* $\mathfrak{X}^*(M) \times \mathfrak{X}(M) \to C^\infty(M)$ *given by* $(\alpha, X) \mapsto \alpha(X) \in C^\infty(M)$ *is clearly multilinear (bilinear) over* $C^\infty(M)$*.*

Suppose now that we have two R-modules V_1 and V_2. Let us construct a category $\mathcal{C}_{\mathrm{V}_1 \times \mathrm{V}_2}$ whose objects are bilinear maps $\mathrm{V}_1 \times \mathrm{V}_2 \to \mathrm{W}$ where W varies over all R-modules but V_1 and V_2 are fixed. A morphism from, say $\mu_1 : \mathrm{V}_1 \times \mathrm{V}_2 \to \mathrm{W}_1$ to $\mu_2 : \mathrm{V}_1 \times \mathrm{V}_2 \to \mathrm{W}_2$ is defined to be a map $\ell : \mathrm{W}_1 \to \mathrm{W}_2$ such that the diagram

$$\begin{array}{ccc} & & \mathrm{W}_1 \\ & \nearrow^{\mu_1} & \\ \mathrm{V}_1 \times \mathrm{V}_2 & & \ell \downarrow \\ & \searrow^{\mu_2} & \\ & & \mathrm{W}_2 \end{array}$$

commutes. Suppose that there is an R-module $\mathrm{T}_{\mathrm{V}_1,\mathrm{V}_2}$ together with a bilinear map $\mathrm{t} : \mathrm{V}_1 \times \mathrm{V}_2 \to \mathrm{T}_{\mathrm{V}_1,\mathrm{V}_2}$ that has the following universal property for this category: For every bilinear map $\mu : \mathrm{V}_1 \times \mathrm{V}_2 \to \mathrm{W}$ there is a unique linear map $\widetilde{\mu} : \mathrm{T}_{\mathrm{V}_1,\mathrm{V}_2} \to \mathrm{W}$ such that the following diagram commutes:

$$\begin{array}{ccc} \mathrm{V}_1 \times \mathrm{V}_2 & \xrightarrow{\mu} & \mathrm{W} \\ \mathrm{t} \downarrow & \nearrow_{\widetilde{\mu}} & \\ \mathrm{T}_{\mathrm{V}_1,\mathrm{V}_2} & & \end{array}$$

If such a pair $\mathrm{T}_{\mathrm{V}_1,\mathrm{V}_2}, \mathrm{t}$ exists with this property then it is unique up to isomorphism in $\mathcal{C}_{\mathrm{V}_1 \times \mathrm{V}_2}$. In other words, if $\widehat{\mathrm{t}} : \mathrm{V}_1 \times \mathrm{V}_2 \to \widehat{\mathrm{T}}_{\mathrm{V}_1,\mathrm{V}_2}$ is another object with this universal property then there is a module isomorphism $\mathrm{T}_{\mathrm{V}_1,\mathrm{V}_2} \cong \widehat{\mathrm{T}}_{\mathrm{V}_1,\mathrm{V}_2}$ such that the following diagram commutes:

$$\begin{array}{ccc} & & \mathrm{T}_{\mathrm{V}_1,\mathrm{V}_2} \\ & \nearrow_{\mathrm{t}} & \\ \mathrm{V}_1 \times \mathrm{V}_2 & & \cong\downarrow \\ & \searrow^{\widehat{\mathrm{t}}} & \\ & & \widehat{\mathrm{T}}_{\mathrm{V}_1,\mathrm{V}_2} \end{array}$$

We refer to any such universal object as a tensor product of V_1 and V_2. We will construct a specific tensor product that we denote by $\mathrm{V}_1 \otimes \mathrm{V}_2$ with the corresponding map denoted by $\otimes : \mathrm{V}_1 \times \mathrm{V}_2 \to \mathrm{V}_1 \otimes \mathrm{V}_2$. More generally, we seek a universal object for k-multilinear maps $\mu : \mathrm{V}_1 \times \cdots \times \mathrm{V}_k \to \mathrm{W}$.

Definition D.12 *A module* $\mathrm{T} = \mathrm{T}_{\mathrm{V}_1,\cdots,\mathrm{V}_k}$ *together with a multilinear map* $\mathrm{u} : \mathrm{V}_1 \times \cdots \times \mathrm{V}_k \to \mathrm{T}$ *is called* ***universal for*** k***-multilinear maps on*** $\mathrm{V}_1 \times \cdots \times \mathrm{V}_k$ *if for every multilinear map* $\mu : \mathrm{V}_1 \times \cdots \times \mathrm{V}_k \to \mathrm{W}$ *there is a unique linear map* $\widetilde{\mu} : \mathrm{T} \to \mathrm{W}$ *such that the following diagram commutes:*

$$\begin{array}{ccc} \mathrm{V}_1 \times \cdots \times \mathrm{V}_k & \overset{\mu}{\to} & \mathrm{W} \\ \mathrm{u}\downarrow & \nearrow_{\widetilde{\mu}} & \\ \mathrm{T} & & \end{array},$$

i.e. we must have $\mu = \widetilde{\mu} \circ \mathrm{u}$*. If such a universal object exists it will be called a* ***tensor product*** *of* $\mathrm{V}_1, ..., \mathrm{V}_k$ *and the module itself* $\mathrm{T} = \mathrm{T}_{\mathrm{V}_1,\cdots,\mathrm{V}_k}$ *is also referred to as a tensor product of the modules* $\mathrm{V}_1, ..., \mathrm{V}_k$*.*

Lemma D.1 *If two modules* $\mathrm{T}_1, \mathrm{u}_1$ *and* $\mathrm{T}_2, \mathrm{u}_2$ *are both universal for* k*-multilinear maps on* $\mathrm{V}_1 \times \cdots \times \mathrm{V}_k$ *then there is an isomorphism* $\Phi : \mathrm{T}_1 \to \mathrm{T}_2$ *such that* $\Phi \circ \mathrm{u}_1 = \mathrm{u}_2$*;*

$$\begin{array}{ccc} & \mathrm{V}_1 \times \cdots \times \mathrm{V}_k & \\ \mathrm{u}_1 \swarrow & & \searrow \mathrm{u}_2 \\ \mathrm{T}_1 & \overset{\Phi}{\to} & \mathrm{T}_2 \end{array}.$$

Proof. By the assumption of universality, there are maps u_1 and u_2 such that $\Phi \circ \mathrm{u}_1 = \mathrm{u}_2$ and $\bar{\Phi} \circ \mathrm{u}_2 = \mathrm{u}_1$. We thus have $\bar{\Phi} \circ \Phi \circ \mathrm{u}_1 = \mathrm{u}_1 = \mathrm{id}$ and by the uniqueness part of the universality of u_1 we must have $\bar{\Phi} \circ \Phi = \mathrm{id}$ or $\bar{\Phi} = \Phi^{-1}$. ■

We now show the existence of a tensor product. The specific tensor product of modules $\mathrm{V}_1, ..., \mathrm{V}_k$ that we construct will be denoted by $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$ and the corresponding map will be denoted by

$$\otimes^k : \mathrm{V}_1 \times \cdots \times \mathrm{V}_k \to \mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k.$$

We start out by introducing the notion of a free module on an arbitrary set. If S is just some set, then we may consider the set $F_{\mathsf{R}}(S)$ all finite formal linear combinations of elements of S with coefficients from R. For example, if $a, b, c \in \mathsf{R}$ and $s_1, s_2, s_3 \in S$ then $as_1 + bs_2 + cs_3$ is such a formal linear combination. In general, an element of $F_{\mathsf{R}}(S)$ will be of the form

$$\sum_{s \in S} a_s s$$

where the coefficients a_s are elements of R and all but finitely many are 0. Thus the sums involved are always finite. Addition of two such expressions and multiplication by elements of R are defined in the obvious way;

$$b \sum_{s \in S} a_s s = \sum_{s \in S} b a_s s$$
$$\sum_{s \in S} a_s s + \sum_{s \in S} b_s s = \sum_{s \in S} (a_s + b_s) s.$$

This is all just a way of speaking of functions $a() : S \to \mathsf{R}$ with finite support. It is also just a means of forcing the element of our arbitrary set to be the "basis elements" of a modules. The resulting module $F_{\mathsf{R}}(S)$ is called the **free module generated by** S. For example, the set of all formal linear combinations of the set of symbols $\{\mathbf{i}, \mathbf{j}\}$ over the real number ring, is just a 2 dimensional vector space with basis $\{\mathbf{i}, \mathbf{j}\}$.

Let $\mathrm{V}_1, \cdots, \mathrm{V}_k$ be modules over R and let $\mathrm{F}_{\mathsf{R}}(\mathrm{V}_1 \times \cdots \times \mathrm{V}_k)$ denote the set of all formal linear combinations of elements of the Cartesian product $\mathrm{V}_1 \times \cdots \times \mathrm{V}_k$. For example

$$3(v_1, w) - 2(v_2, w) \in \mathrm{F}_{\mathsf{R}}(\mathrm{V}, \mathrm{W})$$

but it is **not true** that $3(v_1, w) - 2(v_2, w) = 3(v_1 - 2v_2, w)$ since (v_1, w), (v_2, w) and $(v_1 - 2v_2, w)$ are linearly independent by definition. We now define a submodule of $\mathrm{F}_{\mathsf{R}}(\mathrm{V}_1 \times \cdots \times \mathrm{V}_k)$. Consider the set B of all elements of $\mathrm{F}_{\mathsf{R}}(\mathrm{V}_1 \times \cdots \times \mathrm{V}_k)$ that have one of the following two forms:

1.
$$(\mathrm{v}_1, ..., a\mathrm{v}_i, \cdots, \mathrm{v}_k) - a(\mathrm{v}_1, ..., \mathrm{v}_i, \cdots, \mathrm{v}_k)$$
for some $a \in \mathsf{R}$ and some $1 \le i \le k$ and some $(\mathrm{v}_1, ..., \mathrm{v}_i, \cdots, \mathrm{v}_k) \in \mathrm{V}_1 \times \cdots \times \mathrm{V}_k$.

2.
$$(\mathrm{v}_1, ..., \mathrm{v}_i + \mathrm{w}_i, \cdots, \mathrm{v}_k) - (\mathrm{v}_1, ..., \mathrm{v}_i, \cdots, \mathrm{v}_k) - (\mathrm{v}_1, ..., \mathrm{w}_i, \cdots, \mathrm{v}_k)$$
for some $1 \le i \le k$ and some choice of $(\mathrm{v}_1, ..., \mathrm{v}_i, \cdots, \mathrm{v}_k) \in \mathrm{V}_1 \times \cdots \times \mathrm{V}_k$ and $\mathrm{w}_i \in \mathrm{V}_i$.

We now define $\langle B \rangle$ to be the submodule generated by B and then define the tensor product $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$ of the spaces $\mathrm{V}_1, \cdots, \mathrm{V}_k$ to be the quotient module $\mathrm{F}_{\mathsf{R}}(\mathrm{V}_1 \times \cdots \times \mathrm{V}_k)/\langle B \rangle$. Let

$$\pi : \mathrm{F}_{\mathsf{R}}(\mathrm{V}_1 \times \cdots \times \mathrm{V}_k) \to \mathrm{F}_{\mathsf{R}}(\mathrm{V}_1 \times \cdots \times \mathrm{V}_k)/\langle B \rangle$$

be the quotient map and define $v_1 \otimes \cdots \otimes v_k$ to be the image of $(v_1, \cdots, v_k) \in V_1 \times \cdots \times V_k$ under this quotient map. The quotient is the tensor space we were looking for

$$V_1 \otimes \cdots \otimes V_k := F_R(V_1 \times \cdots \times V_k)/\langle B \rangle.$$

To get our universal object we need to define the corresponding map. The map we need is just the composition

$$V_1 \times \cdots \times V_k \hookrightarrow F_R(V_1 \times \cdots \times V_k) \to V_1 \otimes \cdots \otimes V_k.$$

We denote this map by $\otimes^k : V_1 \times \cdots \times V_k \to V_1 \otimes \cdots \otimes V_k$. Notice that $\otimes^k(v_1, \cdots, v_k) = v_1 \otimes \cdots \otimes v_k$. By construction, we have the following facts:

1. $$v_1 \otimes \cdots \otimes a v_i \otimes \cdots \otimes v_k = a v_1 \otimes \cdots \otimes v_i \otimes \cdots \otimes v_k$$

   for any $a \in R$, any $i \in \{1, 2, ..., k\}$ and any $(v_1, ..., v_i, \cdots, v_k) \in V_1 \times \cdots \times V_k$.

2. $$\begin{aligned} &v_1 \otimes \cdots \otimes (v_i + w_i) \otimes \cdots \otimes v_k \\ &= v_1 \otimes \cdots \otimes v_i \otimes \cdots \otimes v_k + v_1 \otimes \cdots \otimes w_i \otimes \cdots \otimes v_k \end{aligned}$$

   any $i \in \{1, 2, ..., k\}$ and for all choices of $v_1, ..., v_k$ and w_i.

   Thus $\otimes^k$ is multilinear.

Definition D.13 *The elements in the image of π, that is, elements that may be written as $v_1 \otimes \cdots \otimes v_k$ for some v_i, are called* ***decomposable*** *.*

Exercise D.3 *Not all elements are decomposable but the decomposable elements generate $V_1 \otimes \cdots \otimes V_k$.*

It may be that the V_i may be modules over more that one ring. For example, any complex vector space is a module over both $\mathbb{R}$ and $\mathbb{C}$. Also, the module of smooth vector fields $\mathfrak{X}_M(U)$ is a module over $C^\infty(U)$ and a module (actually a vector space) over $\mathbb{R}$. Thus it is sometimes important to indicate the ring involved and so we write the tensor product of two R-modules V and W as $V \otimes_R W$. For instance, there is a big difference between $\mathfrak{X}_M(U) \otimes_{C^\infty(U)} \mathfrak{X}_M(U)$ and $\mathfrak{X}_M(U) \otimes_{\mathbb{R}} \mathfrak{X}_M(U)$.

Now let $\otimes^k : V_1 \times \cdots \times V_k \to V_1 \otimes \cdots \otimes V_k$ be natural map defined above which is the composition of the set injection $V_1 \times \cdots \times V_k \hookrightarrow F_R(V_1 \times \cdots \times V_k)$ and the quotient map $F_R(V_1 \times \cdots \times V_k) \to V_1 \otimes \cdots \otimes V_k$. We have seen that this map actually turns out to be a multilinear map.

Theorem D.1 *Given modules* $\mathrm{V}_1, ..., \mathrm{V}_k$*, the space* $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$ *together with the map* $\otimes^k$ *has the following universal property:*

For any k*-multilinear map* $\mu : \mathrm{V}_1 \times \cdots \times \mathrm{V}_k \to W$*, there is a unique linear map* $\widetilde{\mu} : \mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k \to W$ *called the universal map, such that the following diagram commutes:*

$$\begin{array}{ccc} \mathrm{V}_1 \times \cdots \times \mathrm{V}_k & \xrightarrow{\mu} & W \\ \otimes^k \downarrow & \nearrow_{\widetilde{\mu}} & \\ \mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k & & \end{array} .$$

Thus the pair $(\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k,\ \otimes^k)$ is universal for k-multilinear maps on $\mathrm{V}_1 \times \cdots \times \mathrm{V}_k$ and by D.1 if T, u is any other universal pair for k-multilinear map we have $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k \cong \mathrm{T}$. The module $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$ is called *the* **tensor product** of $\mathrm{V}_1, ..., \mathrm{V}_k$.

Proof. Suppose that $\mu : \mathrm{V}_1 \times \cdots \times \mathrm{V}_k \to W$ is multilinear. Since, $\mathrm{F}_{\mathsf{R}}(\mathrm{V}_1 \times \cdots \times \mathrm{V}_k)$ is free there is a unique linear map $M : \mathrm{F}_{\mathsf{R}}(\mathrm{V}_1 \times \cdots \times \mathrm{V}_k) \to W$ such that $M((\mathrm{v}_1, ..., \mathrm{v}_k)) = \mu(\mathrm{v}_1, ..., \mathrm{v}_k)$. Clearly, this map is zero on $\langle B \rangle$ and so there is a factorization $\widetilde{\mu}$ of M through $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$. Thus we always have

$$\widetilde{\mu}(\mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_k) = M((\mathrm{v}_1, ..., \mathrm{v}_k)) = \mu(\mathrm{v}_1, ..., \mathrm{v}_k).$$

It is easy to see that $\widetilde{\mu}$ is unique since a linear map is determined by its action on generators (the decomposable elements generate $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$) ■

Lemma D.2 *We have the following natural isomorphisms:*

1) $(\mathrm{V} \otimes \mathrm{W}) \otimes \mathrm{U} \cong \mathrm{V} \otimes (\mathrm{W} \otimes \mathrm{U}) \cong \mathrm{V} \otimes (\mathrm{W} \otimes \mathrm{U})$ *and under these isomorphisms* $(\mathrm{v} \otimes \mathrm{w}) \otimes \mathrm{u} \longleftrightarrow \mathrm{v} \otimes (\mathrm{w} \otimes \mathrm{u}) \longleftrightarrow \mathrm{v} \otimes \mathrm{w} \otimes \mathrm{u}$.

2) $\mathrm{V} \otimes \mathrm{W} \cong \mathrm{W} \otimes \mathrm{V}$ *and under this isomorphism* $\mathrm{v} \otimes \mathrm{w} \longleftrightarrow \mathrm{w} \otimes \mathrm{v}$.

Proof. We prove (1) and leave (2) as an exercise.

Elements of the form $(\mathrm{v} \otimes \mathrm{w}) \otimes \mathrm{u}$ generate $(\mathrm{V} \otimes \mathrm{W}) \otimes \mathrm{U}$ so any map that sends $(\mathrm{v} \otimes \mathrm{w}) \otimes \mathrm{u}$ to $\mathrm{v} \otimes (\mathrm{w} \otimes \mathrm{u})$ for all $\mathrm{v}, \mathrm{w}, \mathrm{u}$ must be unique. Now we have compositions

$$(\mathrm{V} \times \mathrm{W}) \times \mathrm{U} \overset{\otimes \times \mathrm{id}_{\mathrm{U}}}{\to} (\mathrm{V} \otimes \mathrm{W}) \times \mathrm{U} \overset{\otimes}{\to} (\mathrm{V} \otimes \mathrm{W}) \otimes \mathrm{U}$$

and

$$\mathrm{V} \times (\mathrm{W} \times \mathrm{U}) \overset{\mathrm{id}_{\mathrm{U}} \times \otimes}{\to} (\mathrm{V} \times \mathrm{W}) \otimes \mathrm{U} \overset{\otimes}{\to} \mathrm{V} \otimes (\mathrm{W} \otimes \mathrm{U}).$$

It is a simple matter to check that these composite maps have the same universal property as the map $\mathrm{V} \times \mathrm{W} \times \mathrm{U} \overset{\otimes}{\to} \mathrm{V} \otimes \mathrm{W} \otimes \mathrm{U}$. The result now follows from the existence and essential uniqueness results proven so far (D.1 and D.1). ■

We shall use the first isomorphism and the obvious generalizations to identify $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$ with all legal parenthetical constructions such as $(((\mathrm{V}_1 \otimes \mathrm{V}_2) \otimes \cdots \otimes \mathrm{V}_j) \otimes \cdots) \otimes \mathrm{V}_k$ and so forth. In short, we may construct $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$ by tensoring spaces two at a time. In particular we assume the isomorphisms

$$(\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k) \otimes (\mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_k) \cong \mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k \otimes \mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_k$$

which map $(\mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_k) \otimes (\mathrm{w}_1 \otimes \cdots \otimes \mathrm{w}_k)$ to $\mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_k \otimes \mathrm{w}_1 \otimes \cdots \otimes \mathrm{w}_k$.

Consider the situation where we have module homomorphisms $h_i : \mathrm{W}_i \to \mathrm{V}_i$ for $1 \leq i \leq m$. We may then define a map $T(h_1, ..., h_m) : \mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_m \to \mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_m$ (by using the universal property again) so that the following diagram commutes:

$$\begin{array}{ccc} \mathrm{W}_1 \times \cdots \times \mathrm{W}_m & \overset{h_1 \times \ldots \times h_m}{\rightarrow} & \mathrm{V}_1 \times \cdots \times \mathrm{V}_m \\ \otimes^k \downarrow & & \otimes^k \downarrow \\ \mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_m & \overset{T(h_1, \ldots, h_m)}{\rightarrow} & \mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_m \end{array}.$$

This is functorial in the sense that

$$T(h_1, ..., h_m) \circ T(g_1, ..., g_m) = T(h_1 \circ g_1, ..., h_m \circ g_m)$$

and $T(\mathrm{id}, ..., \mathrm{id}) = \mathrm{id}$. Also, $T(h_1, ..., h_m)$ has the following effect on decomposable elements:

$$T(h_1, ..., h_m)(\mathrm{v}_1 \otimes \cdots \otimes \mathrm{v}_m) = h_1(\mathrm{v}_1) \otimes \cdots \otimes h_m(\mathrm{v}_m).$$

Now we could jump the gun a bit and use the notation $h_1 \otimes \cdots \otimes h_m$ for $T(h_1, ..., h_m)$ but is this the same thing as the element of $\mathrm{Hom}_{\mathsf{R}}(\mathrm{W}_1, \mathrm{V}_1) \otimes \cdots \otimes \mathrm{Hom}_{\mathsf{R}}(\mathrm{W}_m, \mathrm{V}_m)$ which must be denoted the same way? The answer is that in general, these are distinct objects. On the other hand, there is little harm done if context determines which of the two possible meanings we are invoking. Furthermore, we shall see than in many cases, the two meanings actually do coincide.

The following proposition give a basic and often used isomorphism:

Proposition D.1 *For* $\mathsf{R}-$*modules* $\mathrm{W}, \mathrm{V}, \mathrm{U}$ *we have*

$$\mathrm{Hom}_{\mathsf{R}}(\mathrm{W} \otimes \mathrm{V}, \mathrm{U}) \cong L(\mathrm{W}, \mathrm{V}; \mathrm{U})$$

More generally,

$$\mathrm{Hom}_{\mathsf{R}}(\mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_k, \mathrm{U}) \cong L(\mathrm{W}_1, ..., \mathrm{W}_k; \mathrm{U})$$

Proof. This is more or less just a restatement of the universal property of $\mathrm{W} \otimes \mathrm{V}$. One should check that this association is indeed an isomorphism. ∎

Exercise D.4 *Show that if* W *is free with basis* $(f_1, ..., f_n)$ *then* W^* *is also free and has a dual basis* $f^1, ..., f^n$*, that is,* $f^i(f_j) = \delta^i_j$.

Theorem D.2 *If* $\mathrm{V}_1, ..., \mathrm{V}_k$ *are free* $\mathsf{R}-$*modules and if* $(v^j_1, ..., v^j_{n_j})$ *is a basis for* V_j *then set of all decomposable elements of the form* $v^1_{i_1} \otimes \cdots \otimes v^k_{i_k}$ *form a basis for* $\mathrm{V}_1 \otimes \cdots \otimes \mathrm{V}_k$.

Proposition D.2 *There is a unique* R*-module map* $\iota : \mathrm{W}^*_1 \otimes \cdots \otimes \mathrm{W}^*_k \to (\mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_k)^*$ *such that if* $\alpha_1 \otimes \cdots \otimes \alpha_k$ *is a (decomposable) element of* $\mathrm{W}^*_1 \otimes \cdots \otimes \mathrm{W}^*_k$ *then*

$$\iota(\alpha_1 \otimes \cdots \otimes \alpha_k)(w_1 \otimes \cdots \otimes w_k) = \alpha_1(w_1) \cdots \alpha_k(w_k).$$

If the modules are all free then this is an isomorphism.

Proof. If such a map exists, it must be unique since the decomposable elements span $\mathrm{W}_1^* \otimes \cdots \otimes \mathrm{W}_k^*$. To show existence we define a multilinear map

$$\vartheta : \mathrm{W}_1^* \times \cdots \times \mathrm{W}_k^* \times \mathrm{W}_1 \times \cdots \times \mathrm{W}_k \to \mathsf{R}$$

by the recipe

$$(\alpha_1, ..., \alpha_k, w_1, ..., w_k) \mapsto \alpha_1(w_1) \cdots \alpha_k(w_k).$$

By the universal property there must be a linear map

$$\widetilde{\vartheta} : \mathrm{W}_1^* \otimes \cdots \otimes \mathrm{W}_k^* \otimes \mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_k \to \mathsf{R}$$

such that $\widetilde{\vartheta} \circ u = \vartheta$ where u is the universal map. Now define

$$\begin{aligned} &\iota(\alpha_1 \otimes \cdots \otimes \alpha_k)(w_1 \otimes \cdots \otimes w_k) \\ &:= \widetilde{\vartheta}(\alpha_1 \otimes \cdots \otimes \alpha_k \otimes w_1 \otimes \cdots \otimes w_k). \end{aligned}$$

The fact, that ι is an isomorphism in case the W_i are all free follows easily from exercise **??** and theorem D.2. Once we view an element of W_i as a functional from $\mathrm{W}_i^{**} = L(\mathrm{W}_i^*; \mathsf{R})$ we see that the effect of this isomorphism is to change the interpretation of the tensor product to the "map" tensor product in $(\mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_k)^* = L(\mathrm{W}_1 \otimes \cdots \otimes \mathrm{W}_k; \mathsf{R})$. Thus the basis elements match up under ι. ■

Definition D.14 *The k-th tensor power of a module* W *is defined to be*

$$\mathrm{W}^{\otimes k} := \mathrm{W} \otimes \cdots \otimes \mathrm{W}.$$

This module is also denoted $\bigotimes^k(\mathrm{W})$*. We also define the space of* $\binom{r}{\ s}$***-tensors on*** W *:*

$$\bigotimes{}^r{}_s(\mathrm{W}) := \mathrm{W}^{\otimes r} \otimes (\mathrm{W}^*)^{\otimes s}.$$

Similarly, $\bigotimes{}_s{}^r(\mathrm{W}) := (\mathrm{W}^*)^{\otimes s} \otimes \mathrm{W}^{\otimes r}$ *is the space of* $\binom{\ r}{s}$***-tensors on*** W*.*

Again, although we distinguish $\bigotimes{}^r{}_s(\mathrm{W})$ from $\bigotimes{}_s{}^r(\mathrm{W})$ we shall be able to develop things so as to use mostly the space $\bigotimes{}^r{}_s(\mathrm{W})$ and so eventually we take $\bigotimes^r_s(\mathrm{W})$ to mean $\bigotimes{}^r{}_s(\mathrm{W})$.

If $(v_1, ..., v_n)$ is a basis for a module V and $(v^1, ..., v^n)$ the dual basis for V^* then a basis for $\bigotimes{}^r{}_s(\mathrm{V})$ is given by

$$\{v_{i_1} \otimes \cdots \otimes v_{i_r} \otimes v^{j_1} \otimes \cdots \otimes v^{j_s}\}$$

where the index set is the set $\mathcal{I}(r, s, n)$ defined by

$$\mathcal{I}(r, s, n) := \{(i_1, ..., i_r; j_1, ..., j_s) : 1 \le i_k \le n \text{ and } 1 \le j_k \le n\}.$$

Thus $\bigotimes{}^r{}_s(\mathrm{V})$ has dimension $n^r n^s$ (where $n = \dim(\mathrm{V})$).

We restate the universal property in this special case of tensors:

Proposition D.3 (Universal mapping property) *Given a module or vector space* V *over* R*, then* $\bigotimes^r_s(\mathrm{V})$ *has associated with it, a map*

$$\otimes_{\binom{r}{s}} : \underbrace{\mathrm{V}\times\mathrm{V}\times\cdots\times\mathrm{V}}_{r}\times\underbrace{\mathrm{V}^*\times\mathrm{V}^*\times\cdots\times\mathrm{V}^*}_{s}\to\bigotimes{}^r{}_s(\mathrm{V})$$

such that for any multilinear map $\Lambda\in T_r{}^s(\mathrm{V};\mathsf{R})$*;*

$$\Lambda:\underbrace{\mathrm{V}\times\mathrm{V}\times\cdots\times\mathrm{V}}_{r-times}\times\underbrace{\mathrm{V}^*\times\mathrm{V}^*\cdots\times\mathrm{V}^*}_{s-times}\to\mathsf{R}$$

there is a unique linear map $\widetilde{\Lambda}:\bigotimes{}^r{}_s(\mathrm{V})\to\mathsf{R}$ *such that* $\widetilde{\Lambda}\circ\otimes_{\binom{r}{s}}=\Lambda$*. Up to isomorphism, the space* $\bigotimes{}^r{}_s(\mathrm{V})$ *is the unique space with this universal mapping property.*

Corollary D.1 *There is an isomorphism* $(\bigotimes{}^r{}_s(\mathrm{V}))^*\cong T_r{}^s(\mathrm{V})$ *given by* $\widetilde{\Lambda}\mapsto\widetilde{\Lambda}\circ\otimes^r_s$*. (**Warning**: Notice that the* $T^r{}_s(\mathrm{V})$ *occurring here is **not** the default space* $T^r{}_s(\mathrm{V})$ *that we eventually denote by* $T^r_s(\mathrm{V})$*.*

Corollary D.2 $(\bigotimes{}^r{}_s(\mathrm{V}^*))^*=T_r{}^s(\mathrm{V}^*)$

Now along the lines of the map of proposition D.2 we have a homomorphism

$$\iota^r_{.s}:\bigotimes{}^r{}_s(\mathrm{V})\to T^r{}_s(\mathrm{V}) \tag{D.1}$$

given by

$$\begin{aligned}&\iota^r_{.s}\big(\big(\mathrm{v}_1\otimes\cdots\otimes\mathrm{v}_r\otimes\eta^1\otimes\cdots..\otimes\eta^s\big)\big)(\theta^1,...,\theta^r,\mathrm{w}_1,..,\mathrm{w}_s)\\&=\theta_1(\mathrm{v}_1)\theta_2(\mathrm{v}_2)\cdots\theta_l(\mathrm{v}_l)\eta^1(\mathrm{w}_1)\eta^2(\mathrm{w}_2)\cdots\eta^k(\mathrm{w}_k)\end{aligned}$$

$\theta^1,\theta^2,...,\theta^r\in\mathrm{V}^*$ and $\mathrm{w}_1,\mathrm{w}_2,..,\mathrm{w}_k\in\mathrm{V}$. If V is a finite dimensional free module then we have $\mathrm{V}=\mathrm{V}^{**}$. This is the reflexive property.

Definition D.15 *We say that* V *is **totally reflexive** if the homomorphism 8.1 just given is in fact an isomorphism. This happens for free modules:*

Proposition D.4 *For a finite dimensional free module* V *we have a natural isomorphism* $\bigotimes{}^r{}_s(\mathrm{V})\cong T^r{}_s(\mathrm{V})$*. The isomorphism is given by the map* $\iota^r_{.s}$ *(see 8.1)*

Proof. Just to get the existence of a natural isomorphism we may observe that

$$\begin{aligned}\bigotimes{}^r{}_s(\mathrm{V})&=\bigotimes{}^r{}_s(\mathrm{V}^{**})=(\bigotimes{}^r{}_s(\mathrm{V}^*))^*\\&=T_r{}^s(\mathrm{V}^*)=L(\mathrm{V}^{*r},\mathrm{V}^{***s};\mathsf{R})\\&=L(\mathrm{V}^{*r},\mathrm{V}^s;\mathsf{R}):=T^r{}_s(\mathrm{V}).\end{aligned}$$

We would like to take a more direct approach. Since V is free we may take a basis $\{f_1, ..., f_n\}$ and a dual basis $\{f^1, ..., f^n\}$ for V^*. It is easy to see that ι^r_s sends the basis elements of $\bigotimes^r{}_s(V)$ to basis elements of $T^r{}_s(V; \mathsf{R})$ as for example

$$\iota^1_1 : f^i \otimes f_j \mapsto f^i \otimes f_j$$

where only the interpretation of the $\otimes$ changes. On the right side $f^i \otimes f_j$ is by definition the multilinear map $f^i \otimes f_j : (\alpha, v) := \alpha\left(f^i\right) f_j(v)$ ∎

In the finite dimensional case, we will identify $\bigotimes^r{}_s(V)$ with the space $T^r{}_s(V;\mathsf{R}) = L(V^{r*}, V^s; \mathsf{R})$ of r, s-multilinear maps. We may freely think of a decomposable tensor $v_1 \otimes ... \otimes v_r \otimes \eta^1 \otimes ... \otimes \eta^s$ as a *multilinear map* by the formula

$$\begin{aligned}&(v_1 \otimes \cdots \otimes v_r \otimes \eta^1 \otimes \cdots .. \otimes \eta^s) \cdot (\theta^1, ..., \theta^r, w_1, .., w_s) \\ &= \theta_1(v_1)\theta_2(v_2) \cdots \theta_l(v_l)\eta^1(w_1)\eta^2(w_2) \cdots \eta^k(w_k)\end{aligned}$$

U) of smooth vector fields over an open set U in some manifold M. We shall see that for finite dimensional manifolds $T^r_s(\mathfrak{X}(U))$ is naturally isomorphic to the smooth sections of a so called tensor bundle. We take up this important topic shortly.

D.0.5 Contraction of tensors

Consider a tensor of the form $v_1 \otimes v_2 \otimes \eta^1 \otimes \eta^2 \in T^2_2(V)$ we can define the 1,1 contraction of $v_1 \otimes v_2 \otimes \eta^1 \otimes \eta^2$ as the tensor obtained as

$$C^1_1(v_1 \otimes v_2 \otimes \eta^1 \otimes \eta^2) = \eta^1(v_1) v_2 \otimes \eta^2.$$

Similarly we can define

$$C^1_2(v_1 \otimes v_2 \otimes \eta^1 \otimes \eta^2) = \eta^2(v_1) v_2 \otimes \eta^1.$$

In general, we could define C^i_j on "monomials" $v_1 \otimes v_2 ... \otimes v_r \otimes \eta^1 \otimes \eta^2 \otimes ... \otimes \eta^s$ and then extend linearly to all of $T^r_s(V)$. This works fine for V finite dimensional and turns out to give a notion of contraction which is the same a described in the next definition.

Definition D.16 *Let $\{e_1, ..., e_n\} \subset V$ be a basis for V and $\{e^1, ..., e^n\} \subset V^*$ the dual basis. If $\tau \in T^r_s(V)$ we define $C^i_j \tau \in T^{r-1}_{s-1}(V)$*

$$\begin{aligned}&C^i_j \tau(\theta^1, ..., \theta^{r-1}, w_1, .., w_{s-1}) \\ &= \sum_{k=1}^n \tau(\theta^1, ..., \underset{i-th\ position}{e^k}, \ldots, \theta^{r-1}, w_1, \ldots, \underset{j-th\ position}{e_k}, \ldots, w_{s-1}).\end{aligned}$$

It is easily checked that this definition is independent of the basis chosen. In the infinite dimensional case the sum contraction cannot be defined in general to apply to all tensors. However, we can still define contractions on linear combinations of tensors of the form $\mathrm{v}_1 \otimes \mathrm{v}_2 ... \otimes \mathrm{v}_r \otimes \eta^1 \otimes \eta^2 \otimes ... \otimes \eta^s$ as we did above. Returning to the finite dimensional case, suppose that

$$\tau = \sum \tau^{i_1,...,i_r}_{j_1,...,j_s} e_{i_1} \otimes ... \otimes e_{i_r} \otimes e^{j_1} \otimes ... \otimes e^{j_s}.$$

Then it is easy to check that if we define

$$\tau^{i_1,...,i_{r-1}}_{j_1,...,j_{s-1}} = \sum_{k=1}^{n} \tau^{i_1,...,k,....,i_{r-1}}_{j_1,....,k,...,j_{s-1}}$$

where the upper repeated index k is in the i-th position and the lower occurrence of k is in the j-th position then

$$C^i_j \tau = \sum \tau^{i_1,...,i_{r-1}}_{j_1,...,j_{s-1}} e_{i_1} \otimes ... \otimes e_{i_{r-1}} e^{j_1} \otimes ... \otimes e^{j_{s-1}}.$$

Even in the infinite dimensional case the following definition makes sense. The contraction process can be repeated until we arrive at a function.

Definition D.17 *Given* $\tau \in T^r_s$ *and* $\sigma = \mathrm{w}_1 \otimes ... \otimes \mathrm{w}_l \otimes \eta^1 \otimes \eta^2 \otimes ... \otimes \eta^m \in T^l_m$ *a simple tensor with* $l \leq r$ *and* $m \leq s$*, we define the contraction against* σ *by*

$$\begin{aligned} &\sigma \lrcorner \tau(\alpha_1, ..., \alpha_{r-l}, \mathrm{v}_1, ..., \mathrm{v}_{r-l}) \\ &:= C(\tau \otimes (\mathrm{w}_1 \otimes ... \otimes \mathrm{w}_l \otimes \eta^1 \otimes \eta^2 \otimes ... \otimes \eta^m)) \\ &:= \tau(\eta^1, ..., \eta^m, \alpha_1, ..., \alpha_{r-l}, \mathrm{w}_1, ..., \mathrm{w}_l, \mathrm{v}_1, ..., \mathrm{v}_{r-l}). \end{aligned}$$

For a given simple tensor σ *we thus have a linear map* $\sigma \lrcorner : T^r_s \to T^{r-l}_{s-m}$. *For finite dimensional* V *this can be extended to a bilinear pairing between* T^l_m *and* T^r_s

$$T^l_m(\mathrm{V}) \times T^r_s(\mathrm{V}) \to T^{r-l}_{s-m}(\mathrm{V}).$$

D.0.6 Extended Matrix Notation.

$$\mathbf{A} = \sum A^{i_1 \cdots i_r}_{j_1 \cdots j_s} e_{i_1} \otimes \cdots \otimes e_{i_r} \otimes e^{j_1} \otimes \cdots \otimes e^{j_s}$$

is abbreviated to

$$\mathbf{A} = A^I_J e_I e^J$$

or in matrix notation

$$\mathbf{A} = eAe^*$$

In fact, as a multilinear map we can simply let $\mathbf{v}$ denote an $r-$tuple of vectors from V and θ an $s-$tuple of elements of V^*. Then $\mathbf{A}(\theta, \mathbf{v}) = \theta e A e^* \mathbf{v}$ where for example

$$\theta e = \theta_1(e_{i_1}) \cdots \theta_r(e_{i_r})$$

and

$$e^* \mathbf{v} = e^{j_1}(v_1) \cdots e^{j_s}(v_s)$$

so that

$$\theta e A \varepsilon \mathbf{v} = \sum A^{i_1 \ldots i_r}_{j_1 \cdots j_s} \theta_1(e_{i_1}) \cdots \theta_r(e_{i_r}) e^{j_1}(v_1) \cdots e^{j_s}(v_s).$$

Notice that if one takes the convention that objects with indices up are "column vectors" and indices down "row vectors" then to get the order of the matrices correctly the repeated indices should read *down then up* going from left to right. So $A^I_J e_I e^J$ should be changed to $e_I A^I_J e^J$ before it can be interpreted as a matrix multiplication.

Remark D.6 *We can also write* $\triangle^{K'}_I A^I_J \triangle^J_{L'} = A^{K'}_{L'}$ *where* $\triangle^R_{S'} = e^R e'_S$.

D.0.7 $\mathsf{R}-$Algebras

Definition D.18 *Let* R *be a commutative ring. An* $\mathsf{R}-$***algebra*** $\mathfrak{A}$ *is an* $\mathsf{R}-$*module that is also a ring with identity* $1_{\mathfrak{A}}$ *where the ring addition and the module addition coincide; and where*

1) $r(a_1a_2) = (ra_1)a_2 = a_1(ra_2)$ *for all* $a_1, a_2 \in \mathfrak{A}$ *and all* $r \in \mathsf{R}$,

2) $(r_1r_2)(a_1a_2) = (r_1a_1)(r_2a_2)$.

If we also have $(a_1a_2)a_3 = a_1(a_2a_3)$ *for all* $a_1, a_2, a_3 \in \mathfrak{A}$ *we call* $\mathfrak{A}$ *an associative* $\mathsf{R}-$*algebra.*

Definition D.19 *Let* $\mathfrak{A}$ *and* $\mathfrak{B}$ *be* $\mathsf{R}-$*algebras. A module homomorphism* $h : \mathfrak{A} \to \mathfrak{B}$ *that is also a ring homomorphism is called an* $\mathsf{R}-$***algebra homomorphism****. Epimorphism, monomorphism and isomorphism are defined in the obvious way.*

If a submodule $\mathfrak{I}$ of an algebra $\mathfrak{A}$ is also a two sided ideal with respect to the ring structure on $\mathfrak{A}$ then $\mathfrak{A}/\mathfrak{I}$ is also an algebra.

Example D.11 *The set of all smooth functions* $C^\infty(U)$ *is an* $\mathbb{R}-$*algebra (*$\mathbb{R}$ *is the real numbers) with unity being the function constantly equal to* 1.

Example D.12 *The set of all complex* $n \times n$ *matrices is an algebra over* $\mathbb{C}$ *with the product being matrix multiplication.*

Example D.13 *The set of all complex* $n \times n$ *matrices with real polynomial entries is an algebra over the ring of polynomials* $\mathbb{R}[x]$.

Definition D.20 *The set of all endomorphisms of an* $\mathsf{R}-$*module* W *is an* $\mathsf{R}-$*algebra denoted* $End_{\mathsf{R}}(\mathrm{W})$ *and called the* ***endomorphism algebra*** *of* W*. Here, the sum and scalar multiplication is defined as usual and the product is composition. Note that for* $r \in \mathsf{R}$

$$r(f \circ g) = (rf) \circ g = f \circ (rg)$$

where $f, g \in End_{\mathsf{R}}(\mathrm{W})$.

Definition D.21 *A* $\mathbb{Z}$*-graded* R*-algebra is an* R*-algebra with a direct sum decomposition* $\mathfrak{A} = \sum_{i \in \mathbb{Z}} \mathfrak{A}_i$ *such that* $\mathfrak{A}_i \mathfrak{A}_j \subset \mathfrak{A}_{i+j}$.

Definition D.22 *Let $\mathfrak{A} = \sum_{i\in\mathbb{Z}} \mathfrak{A}_i$ and $\mathfrak{B} = \sum_{i\in\mathbb{Z}} \mathfrak{B}_i$ be $\mathbb{Z}$-graded algebras. An R-algebra homomorphism $h : \mathfrak{A} \to \mathfrak{B}$ is a called a $\mathbb{Z}$-graded homomorphism if $h(\mathfrak{A}_i) \subset \mathfrak{B}_i$ for each $i \in \mathbb{Z}$.*

We now construct the tensor algebra on a fixed $\mathsf{R}-$module W. This algebra is important because is universal in a certain sense and contains the symmetric and alternating algebras as homomorphic images. Consider the following situation: $\mathfrak{A}$ is an $\mathsf{R}-$algebra, W an $\mathsf{R}-$module and $\phi : \mathrm{W} \to \mathfrak{A}$ a module homomorphism. If $h : \mathfrak{A} \to \mathfrak{B}$ is an algebra homomorphism then of course $h \circ \phi : \mathrm{W} \to \mathfrak{B}$ is an $\mathsf{R}-$module homomorphism.

Definition D.23 *Let W be an $\mathsf{R}-$module. An $\mathsf{R}-$ algebra $\mathfrak{U}$ together with a map $\phi : \mathrm{W} \to \mathfrak{U}$ is called universal with respect to W if for any $\mathsf{R}-$module homomorphism $\psi : \mathrm{W} \to \mathfrak{B}$ there is a unique algebra homomorphism If $h : \mathfrak{U} \to \mathfrak{B}$ such that $h \circ \phi = \psi$.*

Again if such a universal object exists it is unique up to isomorphism. We now exhibit the construction of this type of universal algebra. First we define $T^0(\mathrm{W}) := \mathsf{R}$ and $T^1(\mathrm{W}) := \mathrm{W}$. Then we define $T^k(\mathrm{W}) := \mathrm{W}^{k\otimes} = \mathrm{W} \otimes \cdots \otimes \mathrm{W}$. The next step is to form the direct sum $T(\mathrm{W}) := \sum{}_{i=0}^{\infty} T^i(\mathrm{W})$. In order to make this a $\mathbb{Z}$-graded algebra we define $T^i(\mathrm{W}) := 0$ for $i < 0$ and then define a product on $T(\mathrm{W}) := \sum{}_{i\in\mathbb{Z}} T^i(\mathrm{W})$ as follows: We know that for $i, j > 0$ there is an isomorphism $\mathrm{W}^{i\otimes} \otimes \mathrm{W}^{j\otimes} \to \mathrm{W}^{(i+j)\otimes}$ and so a bilinear map $\mathrm{W}^{i\otimes} \times \mathrm{W}^{j\otimes} \to \mathrm{W}^{(i+j)\otimes}$ such that

$$\mathrm{w}_1 \otimes \cdots \otimes \mathrm{w}_i \times \mathrm{w}'_1 \otimes \cdots \otimes \mathrm{w}'_j \mapsto \mathrm{w}_1 \otimes \cdots \otimes \mathrm{w}_i \otimes \mathrm{w}'_1 \otimes \cdots \otimes \mathrm{w}'_j.$$

Similarly, we define $T^0(\mathrm{W}) \times \mathrm{W}^{i\otimes} = \mathsf{R} \times \mathrm{W}^{i\otimes} \to \mathrm{W}^{i\otimes}$ by just using scalar multiplication. Also, $\mathrm{W}^{i\otimes} \times \mathrm{W}^{j\otimes} \to 0$ if either i or j is negative. Now we may use the symbol $\otimes$ to denote these multiplications without contradiction and put then together to form an product on $T(\mathrm{W}) := \sum{}_{i\in\mathbb{Z}} T^i(\mathrm{W})$. It is now clear that $T^i(\mathrm{W}) \times T^j(\mathrm{W}) \mapsto T^i(\mathrm{W}) \otimes T^j(\mathrm{W}) \subset T^{i+j}(\mathrm{W})$ where we make the needed trivial definitions for the negative powers $T^i(\mathrm{W}) = 0$, $i < 0$. Thus $T(\mathrm{W})$ is a graded algebra.

D.0.8 Alternating Multilinear Algebra

In this section we make the simplifying assumption that all of the rings we use will have the following property: The sum of the unity element with itself; $1+1$ is invertible. Thus if we use 2 to denote the element $1 + 1$ then we assume the existence of a unique element "1/2" such that $2 \cdot 1/2 = 1$. Thus, in the case of fields, the assumption is that the field is not of characteristic 2. The reader need only worry about two cases:

1. The unity "1" is just the number 1 in some subring of $\mathbb{C}$ (e.g. $\mathbb{R}$ or $\mathbb{Z}$) or
2. the unity "1" refers to some sort of function or section with values in a ring like $\mathbb{C}$, $\mathbb{R}$ or $\mathbb{Z}$ that takes on the constant value 1. For example, in the ring $C^\infty(M)$, the unity is just the constant function 1.

Definition D.24 *A $\mathbb{Z}$-graded algebra is called skew-commutative (or graded commutative) if for $a_i \in A_i$ and $a_j \in A_j$ we have*

$$a_i \cdot a_j = (-1)^{kl} a_j \cdot a_i$$

Definition D.25 *A morphism of degree n from a graded algebra $A = \bigoplus_{i \in \mathbb{Z}} A_i$ to a graded algebra $B = \bigoplus_{i \in \mathbb{Z}} B_i$ is a algebra homomorphism $h : A \to B$ such that $h(A_i) \subset B_{i+n}$.*

Definition D.26 *A super algebra is a $\mathbb{Z}_2$-graded algebra $A = A_0 \oplus A_1$ such that $A_i \cdot A_j \subset A_{i+j \ mod 2}$ and such that $a_i \cdot a_j = (-1)^{kl} a_j \cdot a_i$ for $i, j \in \mathbb{Z}_2$.*

Alternating tensor maps

Definition D.27 *A k-multilinear map $\alpha : \mathrm{W} \times \cdots \times \mathrm{W} \to \mathrm{F}$ is called **alternating** if $\alpha(\mathrm{w}_1, ..., \mathrm{w}_k) = 0$ whenever $\mathrm{w}_i = \mathrm{w}_j$ for some $i \neq j$. The space of all **alternating** k-multilinear maps into F will be denoted by $L^k_{alt}(\mathrm{W}; \mathrm{F})$ or by $L^k_{alt}(\mathrm{W})$ if the ring is either $\mathbb{R}$ or $\mathbb{C}$ and there is no chance of confusion.*

Remark D.7 *Notice that we have moved the k up to make room for the Alt thus $L^k_{alt}(\mathrm{W};\mathsf{R}) \subset L^0_k(\mathrm{W};\mathsf{R})$.*

Thus if $\omega \in L^k_{alt}(\mathrm{V})$, then for any permutation σ of the letters $1, 2, ..., k$ we have

$$\omega(\mathrm{w}_1, \mathrm{w}_2, .., \mathrm{w}_k) = sgn(\sigma)\omega(\mathrm{w}_{\sigma_1}, \mathrm{w}_{\sigma_2}, .., \mathrm{w}_{\sigma_k}).$$

Now given $\omega \in L^r_{alt}(\mathrm{V})$ and $\eta \in L^s_{alt}(\mathrm{V})$ we define their **wedge product** or **exterior product** $\omega \wedge \eta \in L^{r+s}_{alt}(\mathrm{V})$ by the formula

$$\omega \wedge \eta(\mathrm{v}_1, ..., \mathrm{v}_r, \mathrm{v}_{r+1}, ..., \mathrm{v}_{r+s}) := \frac{1}{r!s!} \sum_{\sigma} \mathrm{sgn}(\sigma)\omega(\mathrm{v}_{\sigma_1}, ..., \mathrm{v}_{\sigma_r})\eta(\mathrm{v}_{\sigma_{r+1}}, ..., \mathrm{v}_{\sigma_{r+s}})$$

or by

$$\omega \wedge \eta(\text{“same as above”}) := \sum_{r,s-\text{shuffles } \sigma} \mathrm{sgn}(\sigma)\omega(\mathrm{v}_{\sigma_1}, ..., \mathrm{v}_{\sigma_r})\eta(\mathrm{v}_{\sigma_{r+1}}, ..., \mathrm{v}_{\sigma_{r+s}}).$$

In the latter formula we sum over all permutations such that $\sigma_1 < \sigma_2 < .. < \sigma_r$ and $\sigma_{r+1} < \sigma_{r+2} < .. < \sigma_{r+s}$. This kind of permutation is called an $r, s-$shuffle as indicated in the summation. The most important case is for $\omega, \eta \in L^1_{alt}(\mathrm{V})$ in which case

$$(\omega \wedge \eta)(v, w) = \omega(v)\eta(w) - \omega(w)\eta(v)$$

This is clearly a skew symmetric multi-linear map.

If we use a basis $\varepsilon^1, \varepsilon^2, ..., \varepsilon^n$ for V^* it is easy to show that the set of all elements of the form $\varepsilon^{i_1} \wedge \varepsilon^{i_2} \wedge \cdots \wedge \varepsilon^{i_k}$ with $i_1 < i_2 < ... < i_k$ form a basis for . Thus for any $\omega \in A^k(\mathrm{V})$

$$\omega = \sum_{i_1 < i_2 < ... < i_k} a_{i_1 i_2, .., i_k} \varepsilon^{i_1} \wedge \varepsilon^{i_2} \wedge \cdots \wedge \varepsilon^{i_k}$$

Remark D.8 *In order to facilitate notation we will abbreviate a sequence of k integers, say $i_1, i_2, ..., i_k$, from the set $\{1, 2, ..., \dim(\mathrm{V})\}$ as I and $\varepsilon^{i_1} \wedge \varepsilon^{i_2} \wedge \cdots \wedge \varepsilon^{i_k}$ is written as ε^I. Also, if we require that $i_1 < i_2 < ... < i_k$ we will write $\vec{I}$. We will freely use similar self explanatory notation as we go along with out further comment. For example, the above equation can be written as*

$$\omega = \sum a_{\vec{I}} \varepsilon^{\vec{I}}$$

Lemma D.3 $L^k_{alt}(\mathrm{V}) = 0$ *if* $k > n = \dim(\mathrm{V})$.

Proof. Easy exercise. ■

If one defines $L^0_{alt}(\mathrm{V})$ to be the scalars $\mathbb{K}$ and recalling that $L^1_{alt}(\mathrm{V}) = \mathrm{V}^*$ then the sum

$$L_{alt}(\mathrm{V}) = \bigoplus_{k=0}^{\dim(M)} L^k_{alt}(\mathrm{V})$$

is made into an algebra via the wedge product just defined.

Proposition D.5 *For* $\omega \in L^r_{alt}(V)$ *and* $\eta \in L^s_{alt}(V)$ *we have* $\omega \wedge \eta = (-1)^{rs} \eta \wedge \omega \in L^{r+s}_{alt}(V)$.

The Abstract Grassmann Algebra

We wish to construct a space that is universal with respect to alternating multilinear maps. To this end, consider the tensor space $T^k(\mathrm{W}) := \mathrm{W}^{k\otimes}$ and let A be the submodule of $T^k(\mathrm{W})$ generated by elements of the form

$$\mathrm{w}_1 \otimes \cdots \mathrm{w}_i \otimes \cdots \otimes \mathrm{w}_i \cdots \otimes \mathrm{w}_k.$$

In other words, A is generated by decomposable tensors with two (or more) equal factors. We define the space of k-vectors to be

$$\mathrm{W} \wedge \cdots \wedge \mathrm{W} := \bigwedge^k \mathrm{W} := T^k(\mathrm{W})/\mathrm{A}.$$

Let $\mathrm{A}_k : \mathrm{W} \times \cdots \times \mathrm{W} \to T^k(\mathrm{W}) \to \ngeq^k \mathrm{W}$ be the canonical map composed with projection onto $\bigwedge^k \mathrm{W}$. This map turns out to be an alternating multilinear map. We will denote $\mathrm{A}_k(\mathrm{w}_1, ..., \mathrm{w}_k)$ by $\mathrm{w}_1 \wedge \cdots \wedge \mathrm{w}_k$. The pair $(\bigwedge^k \mathrm{W}, \mathrm{A}_k)$ is universal with respect to alternating k-multilinear maps: Given any alternating k-multilinear map $\alpha : \mathrm{W} \times \cdots \times \mathrm{W} \to \mathrm{F}$, there is a unique linear map $\alpha_\wedge : \bigwedge^k \mathrm{W} \to \mathrm{F}$ such that $\alpha = \alpha_\wedge \circ \mathrm{A}_k$; that is$\bigwedge^k$

$$\begin{array}{ccc} \mathrm{W} \times \cdots \times \mathrm{W} & \xrightarrow{\alpha} & \mathrm{F} \\ \mathrm{A}_k \downarrow & \nearrow_{\alpha_\wedge} & \\ \bigwedge^k \mathrm{W} & & \end{array}$$

commutes. Notice that we also have that $\mathrm{w}_1\wedge\cdots\wedge\mathrm{w}_k$ is the image of $\mathrm{w}_1\otimes\cdots\otimes\mathrm{w}_k$ under the quotient map. Next we define $\bigwedge \mathrm{W} := \sum_{k=0}^{\infty} \bigwedge^k \mathrm{W}$ and impose the multiplication generated by the rule

$$(\mathrm{w}_1 \wedge \cdots \wedge \mathrm{w}_i) \times (\mathrm{w}'_1 \wedge \cdots \wedge \mathrm{w}'_j) \mapsto \mathrm{w}_1 \wedge \cdots \wedge \mathrm{w}_i \wedge \mathrm{w}'_1 \wedge \cdots \wedge \mathrm{w}'_j \in \bigwedge^{i+j} \mathrm{W}.$$

The resulting algebra is called the Grassmann algebra or exterior algebra. If we need to have a $\mathbb{Z}$ grading rather than a $\mathbb{Z}^+$ grading we may define $\bigwedge^k \mathrm{W} := 0$ for $k < 0$ and extend the multiplication in the obvious way.

Notice that since $(\mathrm{w} + \mathrm{v}) \wedge (\mathrm{w} + \mathrm{v}) = 0$, it follows that $\mathrm{w} \wedge \mathrm{v} = -\mathrm{v} \wedge \mathrm{w}$. In fact, any odd permutation of the factors in a decomposable element such as $\mathrm{w}_1 \wedge \cdots \wedge \mathrm{w}_k$, introduces a change of sign:

$$\begin{aligned}&\mathrm{w}_1 \wedge \cdots \wedge \mathrm{w}_i \wedge \cdots \wedge \mathrm{w}_j \wedge \cdots \wedge \mathrm{w}_k \\ &= -\mathrm{w}_1 \wedge \cdots \wedge \mathrm{w}_j \wedge \cdots \wedge \mathrm{w}_i \wedge \cdots \wedge \mathrm{w}_k\end{aligned}$$

Just to make things perfectly clear, we exhibit a short random calculation where the ring is the real numbers $\mathbb{R}$:

$$\begin{aligned}&(2(1 + 2\mathrm{w} + \mathrm{w} \wedge \mathrm{v} + \mathrm{w} \wedge \mathrm{v} \wedge \mathrm{u}) + \mathrm{v} \wedge \mathrm{u}) \wedge (\mathrm{u} \wedge \mathrm{v}) \\ &= (2 + 4\mathrm{w}+2\mathrm{w} \wedge \mathrm{v}+2\mathrm{w} \wedge \mathrm{v} \wedge \mathrm{u} + \mathrm{v} \wedge \mathrm{u}) \wedge (-\mathrm{v} \wedge \mathrm{u}) \\ &= -2\mathrm{v} \wedge \mathrm{u} - 4\mathrm{w} \wedge \mathrm{v} \wedge \mathrm{u} - 2\mathrm{w} \wedge \mathrm{v} \wedge \mathrm{v} \wedge \mathrm{u} - 2\mathrm{w} \wedge \mathrm{v} \wedge \mathrm{u} \wedge \mathrm{v} \wedge \mathrm{u} - \mathrm{v} \wedge \mathrm{u} \wedge \mathrm{v} \wedge \mathrm{u} \\ &= -2\mathrm{v} \wedge \mathrm{u} - 4\mathrm{w} \wedge \mathrm{v} \wedge \mathrm{u} + 0 + 0 = -2\mathrm{v} \wedge \mathrm{u} - 4\mathrm{w} \wedge \mathrm{v} \wedge \mathrm{u}.\end{aligned}$$

Lemma D.4 *If* V *is has rank* n*, then* $\bigwedge^k \mathrm{V} = 0$ *for* $k \geq n$*. If* $f_1, ..., f_n$ *is a basis for* V *then the set*

$$\{f_{i_1} \wedge \cdots \wedge f_{i_k} : 1 \leq i_1 < \cdots < i_k \leq n\}$$

is a basis for $\bigwedge^k \mathrm{V}$ *where we agree that* $f_{i_1} \wedge \cdots \wedge f_{i_k} = 1$ *if* $k = 0$*.*

The following lemma follows easily from the universal property of $\alpha_\wedge : \bigwedge^k \mathrm{W} \to \mathrm{F}$:

Lemma D.5 *There is a natural isomorphism*

$$L^k_{alt}(\mathrm{W}; \mathrm{F}) \cong L(\bigwedge^k \mathrm{W}; \mathrm{F})$$

In particular,

$$L^k_{alt}(\mathrm{W}) \cong \left(\bigwedge^k \mathrm{W}\right)^*.$$

Remark D.9 (Convention) *Let* $\alpha \in L^k_{alt}(\mathrm{W}; \mathrm{F})$*. Because the above isomorphism is so natural it may be taken as an identification and so we sometimes write* $\alpha(v_1, ..., v_k)$ *as* $\alpha(v_1 \wedge \cdots \wedge v_k)$*.*

In the finite dimensional case, we have module isomorphisms

$$\bigwedge^k \mathrm{W}^* \cong \left(\bigwedge^k \mathrm{W}\right)^* \cong L^k_{alt}(\mathrm{W})$$

which extends by direct sum to

$$L_{alt}(\mathrm{W}) \cong \bigwedge \mathrm{W}^*$$

This isomorphism is in fact an exterior algebra isomorphism (we have an exterior product defined for both).

The following table summarizes:

Exterior Products	Isomorphisms that hold in finite dimension	Alternating multilinear maps
$\bigwedge^k \mathrm{W}$	$\downarrow$	
$\bigwedge^k \mathrm{W}^* \cong \left(\bigwedge^k \mathrm{W}\right)^*$	$\cong$	$L^k_{alt}(\mathrm{W})$
$\bigwedge \mathrm{W} = \bigoplus_k \bigwedge^k \mathrm{W}$		
$\bigwedge \mathrm{W}^* = \bigoplus_k \bigwedge^k \mathrm{W}^*$	graded algebra iso. $\cong$	$A(\mathrm{W}) = \bigoplus_k L^k_{alt}(\mathrm{W})$

D.0.9 Orientation on vector spaces

Let V be a finite dimensional vector space. The set of all ordered bases fall into two classes called orientation classes.

Definition D.28 *Two bases are in the same orientation class if the change of basis matrix from one to the other has positive determinant.*

That is, given two frames (bases) in the same class, say $(f_1, ...f_n)$ and $(\widetilde{f}_1, ...\widetilde{f}_n)$ with

$$\widetilde{f}_i = f_j C^j_i$$

then $\det C > 0$ and we say that the frames determine the same orientation. The relation is easily seen to be an equivalence relation.

Definition D.29 *A choice of one of the two orientation classes of frames for a finite dimensional vector space* V *is called an* ***orientation*** *on* V. *The vector space in then said to be oriented.*

Exercise D.5 *Two frames, say* $(f_1, ..., f_n)$ *and* $(\widetilde{f}_1, ..., \widetilde{f}_n)$ *determine the same orientation on* V *if and only if* $f_1 \wedge ... \wedge f_n = a\ \widetilde{f}_1 \wedge ... \wedge \widetilde{f}_n$ *for some positive real number* $a > 0$.

Exercise D.6 *If* σ *is a permutation on* n *letters* $\{1, 2, ...n\}$ *then* $(f_{\sigma 1}, ..., f_{\sigma n})$ *determine the same orientation if and only if* $sgn(\sigma) = +1$.

A top form $\omega \in L^n_{alt}(\mathrm{V})$ determines an orientation on V by the rule $(f_1, ..., f_n) \sim (\widetilde{f_1}, ..., \widetilde{f_n})$ if and only if

$$\omega(f_1, ..., f_n) = \omega(\widetilde{f_1}, ..., \widetilde{f_n}).$$

Furthermore, two top forms $\omega_1, \omega_2 \in L^n_{alt}(\mathrm{V})$ determine the same orientation on V if and only if $\omega_1 = a\omega_2$ for some positive real number $a > 0$.

First of all lets get the basic idea down. Think about this: A function defined on the circle with range in the interval $(0, 1)$ can be thought of in terms of its graph. The latter is a subset, a cross section, of the product $S^1 \times (0, 1)$. Now, what would a similar cross section of the Mobius band signify? This can't be the same thing as before since a continuous cross section of the Mobius band would have to cross the center and this need not be so for $S^1 \times (0, 1)$. Such a cross section would have to be a sort of twisted function. The Mobius band (with projection onto its center line) provides us with our first nontrivial example of a fiber bundle. The cylinder $S^1 \times (0, 1)$ with projection onto one the factors is a trivial example. Projection onto S^1 gives a topological line bundle while projection onto the interval is a circle bundle. Often what we call the Mobius band will be the slightly different object that is a twisted version of $S^1 \times \mathbb{R}^1$. Namely, the space obtained by identifying one edge of $[0, 1] \times \mathbb{R}^1$ with the other but with a twist.

A fiber bundle is to be though of as a bundle of -or "parameterized family" of-spaces $E_x = \pi^{-1}(x) \subset E$ called the fibers. Nice bundles have further properties that we shall usually assume without explicit mention. The first one is simply that the spaces are Hausdorff and paracompact. The second one is called local triviality. In order to describe this we need the notion of a bundle map and the ensuing notion of equivalence of general fiber bundles.

Most of what we do here will work either for the general topological category or for the smooth category so we once again employ the conventions of 3.7.1.

Definition D.30 *A general C^r- bundle is a triple $\xi = (E, \pi, X)$ where $\pi : E \to M$ is a surjective C^r-map of C^r-spaces (called the bundle projection). For each $p \in X$ the subspace $E_p := \pi^{-1}(p)$is called the fiber over p. The space E is called the **total space** and X is the **base space**. If $S \subset X$ is a subspace we can always form the restricted bundle (E_S, π_S, S) where $E_S = \pi^{-1}(S)$ and $\pi_S = \pi|_S$ is the restriction.*

Definition D.31 *A C^r-**section** of a general bundle $\pi_E : E \to M$ is a C^r-map $s : M \to E$ such that $\pi_E \circ s = \mathrm{id}_M$. In other words, the following diagram must commute:*

$$\begin{array}{ccccc} s & & E & & \\ & \nearrow & \downarrow & \pi_E & \\ M & \to & M & & . \\ & \mathrm{id} & & & \end{array}$$

The set of all C^r-sections of a general bundle $\pi_E : E \to M$ is denoted by $\Gamma^k(M, E)$. We also define the notion of a section over an open set U in M is the obvious way and these are denoted by $\Gamma^k(U, E)$.

Notation D.2 *We shall often abbreviate to just* $\Gamma(U, E)$ *or even* $\Gamma(E)$ *whenever confusion is unlikely. This is especially true in case* $k = \infty$ *(smooth case) or* $k = 0$ *(continuous case).*

Now there are two different ways to treat bundles as a category:

The Category Bun.

Actually, we should define the Categories Bun_k; $k = 0, 1, ..., \infty$. The objects of Bun_k are C^k-fiber bundles. We shall abbreviate to just "Bun" in cases where a context has been establish and confusion is unlikely.

Definition D.32 *A morphism from* $\mathrm{Hom}_{Bun_k}(\xi_1, \xi_2)$*, also called a bundle map from a* C^r-*fiber bundle* $\xi_1 := (E_1, \pi_1, X_1)$ *to another fiber bundle* $\xi_2 := (E_2, \pi_2, X_2)$ *is a pair of* C^r-*maps* $(\bar{f}, f)$ *such that the following diagram commutes:*

$$\begin{array}{ccc} E_1 & \xrightarrow{\bar{f}} & E_2 \\ \downarrow & & \downarrow \\ X_1 & \xrightarrow{f} & X_2 \end{array}$$

If both maps are C^r*-isomorphisms we call the map a (*C^r*-) bundle isomorphism.*

Definition D.33 *Two fiber bundles* $\xi_1 := (E_1, \pi_1, X_1)$ *and* $\xi_2 := (E_2, \pi_2, X_2)$ *are* ***equivalent*** *in* Bun_k *or* ***isomorphic*** *if there exists a bundle isomorphism from* ξ_1 *to* ξ_2*.*

Definition D.34 *Two fiber bundles* $\xi_1 := (E_1, \pi_1, X_1)$ *and* $\xi_2 := (E_2, \pi_2, X_2)$ *are said to be* ***locally equivalent*** *if for any* $y \in E_1$ *there is an open set* U *containing* p *and a bundle equivalence* $(f, \bar{f})$ *of the restricted bundles:*

$$\begin{array}{ccc} E_1|\,U & \xrightarrow{\bar{f}} & E_2|\,f(U) \\ \downarrow & & \downarrow \\ U & \xrightarrow{f} & f(U) \end{array}$$

The Category $Bun_k(X)$

Definition D.35 *A morphism from* $\mathrm{Hom}_{Bun_k(X)}(\xi_1, \xi_2)$*, also called a bundle map over* X *from a* C^r*-fiber bundle* $\xi_1 := (E_1, \pi_1, X)$ *to another fiber bundle* $\xi_2 := (E_2, \pi_2, X)$ *is a* C^r*-map* $\bar{f}$ *such that the following diagram commutes:*

$$\begin{array}{ccccc} E_1 & & \xrightarrow{\bar{f}} & & E_2 \\ & \searrow & & \swarrow & \\ & & X & & \end{array}$$

If both maps are C^r*-isomorphisms we call the map a (*C^r*-)* ***bundle isomorphism over*** X *(also called a* ***bundle equivalence****).*

Definition D.36 *Two fiber bundles $\xi_1 := (E_1, \pi_1, X)$ and $\xi_2 := (E_2, \pi_2, X)$ are* ***equivalent*** *in $Bun_k(X)$ or* ***isomorphic*** *if there exists a (C^r-) bundle isomorphism over X from ξ_1 to ξ_2.*

By now the reader is no doubt tired of the repetitive use of the index C^r so from now on we will simple refer to space (or manifolds) and maps where the appropriate smoothness C^r will not be explicitly stated unless something only works for a specific value of r.

Definition D.37 *Two fiber bundles $\xi_1 := (E_1, \pi_1, X)$ and $\xi_2 := (E_2, \pi_2, X)$ are said to be* ***locally equivalent*** *(over X) if for any $y \in E_1$ there is an open set U containing p and a bundle equivalence $(\bar{f}, f)$ of the restricted bundles:*

$$\begin{array}{ccc} E_1|U & \xrightarrow{\bar{f}} & E_2|U \\ \downarrow & & \downarrow \\ U & \xrightarrow{\mathrm{id}} & U \end{array}$$

Now for any space X **the trivial bundle** with fiber F is the triple $(X \times F, pr_1, X)$ where pr_1 always denoted the projection onto the first factor. Any bundle over X that is bundle equivalent to $X \times F$ is referred to as a trivial bundle.

We will now add in an extra condition that we will usually need:

Definition D.38 *A (C^r-) fiber bundle $\xi := (E, \pi, X)$ is said to be* ***locally trivial*** *(with fiber F) if every for every $x \in X$ has an open neighborhood U such that $\xi_U := (E_U, \pi_U, U)$ is isomorphic to the trivial bundle $(U \times F, pr_1, U)$. Such a fiber bundle is called a* ***locally trivial fiber bundle****.*

We immediately make the following convention: **All fiber bundles in the book will be assumed to be locally trivial unless otherwise stated.** Once we have the local triviality it follows that each fiber $E_p = \pi^{-1}(p)$ is homeomorphic (in fact, C^r-diffeomorphic) to F.

Notation D.3 *We shall take the liberty of using a variety of notations when talking about bundles most of which are quite common and so the reader may as well get used to them. We sometimes write $F \hookrightarrow E \xrightarrow{\pi} X$ to refer to a fiber bundle with typical fiber F. The notation suggests that F may be embedded into E as one of the fibers. This embedding is not canonical in general.*

A **bundle chart** for a fiber bundle $F \hookrightarrow E \xrightarrow{\pi} X$ is a pair (ϕ, U) where $U \subset M$ is open and $\phi : E|_U \to U \times F$ is a map such that the following diagram commutes:

$$\begin{array}{ccccc} E|_U & & \xrightarrow{\phi} & & U \times F \\ \pi & \searrow & & \swarrow & \\ & & U & & \end{array} .$$

Such a map ϕ is also called a **local trivialization**. It follows from the definition that there is a cover of E by **bundle charts** meaning that there are a family of local trivializations $\phi_\alpha : E_{U_\alpha} \to U_\alpha \times F$ such that the open sets U_α cover M. Note that $\phi_\alpha = (\pi, \Phi_\alpha)$ for some smooth map $\Phi_\alpha : E_{U_\alpha} \to F$. It follows that so called **overlap maps** $\phi_\alpha \circ \phi_\beta^{-1} : U_\alpha \cap U_\beta \times F \to U_\alpha \cap U_\beta \times F$ must have the form $\phi_\alpha \circ \phi_\beta^{-1}(x, u) = (x, \phi_{\beta\alpha,x}(u))$ for where $\phi_{\beta\alpha,x} \in Diff^r(F)$ defined for each $x \in U_\alpha \cap U_\beta$. To be explicit, the diffeomorphism $\phi_{\beta\alpha,x}$ arises as follows;

$$y \mapsto (x, y) \mapsto \phi_\alpha \circ \phi_\beta|_{E_y}^{-1}(x, y) = (x, \phi_{\alpha\beta,x}(y)) \mapsto \phi_{\alpha\beta,x}(y).$$

The maps $U_\alpha \cap U_\beta \to Diff^r(F)$ given by $x \mapsto \phi_{\alpha\beta,x}$ are called **transition maps**.

Definition D.39 *Let $F \hookrightarrow E \xrightarrow{\pi} M$ be a (locally trivial) fiber bundle. A cover of E by bundle charts $\{\phi_\alpha, U_\alpha\}$ is called a **bundle atlas** for the bundle.*

Definition D.40 *It may be that there exists a C^r-group G (a Lie group in the smooth case) and a representation ρ of G in $Diff^r(F)$ such that for each nonempty $U_\alpha \cap U_\beta$ we have $\phi_{\beta\alpha,x} = \rho(g_{\alpha\beta}(x))$ for some C^r-map $g_{\alpha\beta} : U_\alpha \cap U_\beta \to G$. In this case we say that G serves as a **structure group** for the bundle via the representation ρ. In case the representation is a faithful one then we may as well take G to be a subgroup of $Diff^r(F)$ and then we simply have $\phi_{\beta\alpha,x} = g_{\alpha\beta}(x)$. Alternatively, we may speak in terms of group actions so that G acts on F by diffeomorphisms.*

The maps $g_{\alpha\beta} : U_\alpha \cap U_\beta \to G$ must satisfy certain consistency relations:

$$\begin{aligned} g_{\alpha\alpha}(x) &= \mathrm{id} \;\text{ for } x \in U_\alpha \\ g_{\alpha\beta}(x) g_{\beta\alpha}(x) &= \mathrm{id} \;\text{ for } x \in U_\alpha \cap U_\beta \\ g_{\alpha\beta}(x) g_{\beta\gamma}(x) g_{\gamma\alpha}(x) &= \mathrm{id} \;\text{ for } x \in U_\alpha \cap U_\beta \cap U_\gamma. \end{aligned} \tag{D.2}$$

A system of maps $g_{\alpha\beta}$ satisfying these relations is called a **cocycle** for the cover $\{U_\alpha\}$.

Definition D.41 *A fiber bundle $\xi := (F \hookrightarrow E \xrightarrow{\pi} X)$ together with a G action on F is called a G-bundle if there exists a bundle atlas $\{(U_\alpha, \phi_\alpha)\}$ for ξ such that the overlap maps have the form $\phi_\alpha \circ \phi_\beta^{-1}(x, y) = (x, g_{\alpha\beta}(x) \cdot y))$ cocycle $\{g_{\alpha\beta}\}$ for the cover $\{U_\alpha\}$.*

Theorem D.3 *Let G have C^r-action on F and suppose we are given cover $\{U_\alpha\}_{\alpha \in A}$ of a C^r-space M and cocycle $\{g_{\alpha\beta}\}$ for the cover. Then there exists a G-bundle with an atlas $\{(U_\alpha, \phi_\alpha)\}$ satisfying $\phi_\alpha \circ \phi_\beta^{-1}(x, y) = (x, g_{\alpha\beta}(x) \cdot y))$ on any nonempty overlaps $U_\alpha \cap U_\beta$.*

Proof. On the union $\Sigma := \bigcup_\alpha \{\alpha\} \times U_\alpha \times F$ define an equivalence relation such that

$$(\alpha, u, v) \in \{\alpha\} \times U_\alpha \times F$$

is equivalent to $(\beta, x, y) \in \{\beta\} \times U_\beta \times F$ if and only if $u = x$ and $v = g_{\alpha\beta}(x) \cdot y$.

The total space of our bundle is then $E := \Sigma/\sim$. The set Σ is essentially the disjoint union of the product spaces $U_\alpha \times F$ and so has an obvious topology. We then give $E := \Sigma/\sim$ the quotient topology. The bundle projection π_E is induced by $(\alpha, u, v) \mapsto u$. Notice that $\pi_E^{-1}(U_\alpha)$ To get our trivializations we define

$$\phi_\alpha(e) := (u, v) \text{ for } e \in \pi_E^{-1}(U_\alpha)$$

where (u, v) is the unique member of $U_\alpha \times F$ such that $(\alpha, u, v) \in e$. The point here is that $(\alpha, u_1, v_1) \sim (\alpha, u_2, v_2)$ only if $(u_1, v_1) = (u_2, v_2)$. Now suppose $U_\alpha \cap U_\beta \neq \emptyset$. Then for $x \in U_\alpha \cap U_\beta$ the element $\phi_\beta^{-1}(x, y)$ is in $\pi_E^{-1}(U_\alpha \cap U_\beta) = \pi_E^{-1}(U_\alpha) \cap \pi_E^{-1}(U_\beta)$ and so $\phi_\beta^{-1}(x, y) = [(\beta, x, y)] = [(\alpha, u, v)]$.

This means that $x = u$ and $v = g_{\alpha\beta}(x) \cdot y$. From this it is not hard to see that

$$\phi_\alpha \circ \phi_\beta^{-1}(x, y) = (x, g_{\alpha\beta}(x) \cdot y)).$$

We leave the question of the regularity of these maps and the C^r structure to the reader. ■

An important tool in the study of fiber bundles is the notion of a pullback bundle. We shall see that construction time and time again. Let $\xi = (F \hookrightarrow E \xrightarrow{\pi} M)$ be a -fiber bundle and suppose we have a -map $f : X \to M$. We want to define a fiber bundle $f^*\xi = (F \hookrightarrow f^*E \to X)$. As a set we have

$$f^*E = \{(x, e) \in X \times E : f(x) = \pi(e)\}.$$

The projection $f^*E \to X$ is the obvious one:$(x, e) \mapsto x \in N$.

Exercise D.7 *Exhibit fiber bundle charts for f^*E.*

Let $\{g_{\alpha\beta}\}$ be a cocycle for some cover $\{U_\alpha\}_{\alpha\in A}$ of M which determines a bundle $\xi = (F \hookrightarrow E \xrightarrow{\pi} M)$. If $f : X \to M$ as above, then $\{g_{\alpha\beta} \circ f\} = \{f^*g_{\alpha\beta}\}$ is a cocycle for the same cover and the bundle determined by this cocycle is (up to isomorphism) none other than the pullback bundle $f^*\xi$.

$$\{g_{\alpha\beta}\} \rightsquigarrow \xi$$
$$\{f^*g_{\alpha\beta}\} \rightsquigarrow f^*\xi$$

The verification of this is an exercise that is easy but constitutes important experience so the reader should not skip the next exercise:

Exercise D.8 *Verify that above claim.*

Exercise D.9 *Show that if $A \subset M$ is a subspace of the base space of a bundle $\xi = (F \hookrightarrow E \xrightarrow{\pi} M)$ and $\iota : A \hookrightarrow M$ then $\iota^{-1}(\xi)$ is naturally isomorphic to the restricted bundle $\xi_A = (F \hookrightarrow E_A \to A)$.*

An important class of fiber bundles often studied on their own is the vector bundles. Roughly, a vector bundle is a fiber bundle with fibers being vector spaces. More precisely, we make the following definition:

Definition D.42 *A real (or complex) (C^r-)* ***vector bundle*** *is a (C^r-) fiber bundle (E, π, X) such that*

(i) Each fiber $E_x := \pi^{-1}(x)$ has the structure of a real (resp. complex) vector space.

(ii) There exists a cover by bundle charts (U_α, ϕ_α) such that each restriction $\phi_\alpha|_{E_x}$ is a real (resp. complex) vector space isomorphism. We call these ***vector bundle charts*** *or VB-charts..*

Equivalently we can define a vector bundle to be a fiber bundle with typical fiber $\mathbb{F}^n$ and such that the transition maps take values in $\mathrm{GL}(n, \mathbb{F})$.

Exercise D.10 *Show that a locally trivial fiber bundle is a vector bundle if and only if the structure representation is a linear representation. D.42.*

As we indicated above, for a smooth (or C^r , $r > 0$) vector bundle we require that all the maps are smooth (or C^r , $r > 0$) and in particular we require that $f_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(k, \mathbb{F})$ are all smooth.

The tangent bundle of a manifold is a vector bundle. If the tangent bundle of a manifold is trivial we say that M is **parallelizable**.

Exercise D.11 *Show that a manifold is parallelizable if and only if there are $n - \dim M$ everywhere linearly independent vector fields $X_1, ..., X_n$ defined everywhere on M.*

The set of all vector bundles is a category **Vect**. Once again we need to specify the appropriate morphisms in this category and the correct choice should be obvious. A **vector bundle morphism** (or vector bundle map) between $\xi_1 := (E_1, \pi_1, X_1)$ and $\xi_2 := (E_2, \pi_2, X_2)$ is a bundle map $(\bar{f}, f)$:

$$\begin{array}{ccc} E_1 & \stackrel{\bar{f}}{\to} & E_2 \\ \downarrow & & \downarrow \\ X_1 & \stackrel{f}{\to} & X_2 \end{array}$$

that is linear on fibers. That is $\bar{f}\big|_{\pi_1^{-1}(x)}$ is a linear map from $\pi_1^{-1}(x)$ into the fiber $\pi_2^{-1}(x)$. We also have the category $\mathbf{Vect}(X)$ consisting of all vector bundles over the fixed space X. Here the morphisms are bundle maps of the form (F, id_X). Two vector bundles $\xi_1 := (E_1, \pi_1, X)$ and $\xi_2 := (E_2, \pi_2, X)$ over the same space X are **isomorphic** (over X) if there is a bundle isomorphism over X from ξ_1 to ξ_2 that is a linear isomorphism when restricted to each fiber. Such a map is called a vector bundle isomorphism.

In the case of vector bundles the transition maps are given by a representation of a Lie group G as a subgroup of $\mathrm{GL}(n, \mathbb{F})$ where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ depending on whether the vector bundle is real or complex. More precisely, if

$\xi = (\mathbb{F}^k \hookrightarrow E \xrightarrow{\pi} M)$ there is a Lie group homomorphism $\rho : G \to \mathrm{GL}(k, \mathbb{F})$ such that for some VB-atlas $\{U_\alpha, \phi_\alpha\}$ we have the overlap maps

$$\phi_\alpha \circ \phi_\beta^{-1} : U_\alpha \cap U_\beta \times \mathbb{F}^k \to U_\alpha \cap U_\beta \times \mathbb{F}^k$$

are given by $\phi_\alpha \circ \phi_\beta^{-1}(x, v) = (x, \rho(g_{\alpha\beta}(x))v)$ for a cocycle $\{g_{\alpha\beta}\}$. In a great many cases, the representation is faithful and we may as well assume that $G \subset \mathrm{GL}(k, \mathbb{F})$ and that the representation is the standard one given by matrix multiplication $v \mapsto gv$. On the other hand we cannot restrict ourselves to this case because of our interest in the phenomenon of spin. A simple observation that gives a hint of what we are talking about is that if $G \subset G(n, \mathbb{R})$ acts on $\mathbb{R}^k$ by matrix multiplication and $h : \widetilde{G} \to G$ is a covering group homomorphism (or any Lie group homomorphism) then $v \mapsto g \cdot v := h(g)v$ is also action. Put another way, if we define $\rho_h : \widetilde{G} \to G(n, \mathbb{R})$ by $\rho_h(g) = h(g)v$ then ρ_h is representation of $\widetilde{G}$. The reason we care about this seemingly trivial fact only becomes apparent when we try to globalize this type of lifting as well will see when we study spin structures later on.

To summarize this point we may say that whenever we have a VB-atlas $\{U_\alpha, \phi_\alpha\}$ we have the transition functions $\{\phi_{\alpha\beta}\} = \{x \mapsto \phi_{\beta\alpha,x}\}$ which are given straight from the overlap maps $\phi_\alpha \circ \phi_\beta^{-1}(x, u)$ by $\phi_\alpha \circ \phi_\beta^{-1}(x, u) = (x, \phi_{\beta\alpha,x}(u))$. Of course, the transition functions $\{\phi_{\alpha\beta}\}$ certainly form a cocycle for the cover $\{U_\alpha\}$ but there may be cases when we want a (not necessarily faithful) representation $\rho : G \to \mathrm{GL}(k, \mathbb{F}^n)$ of some group G not necessarily a subgroup of $\mathrm{GL}(k, \mathbb{F}^n)$ together with some $G-$valued cocycle $\{g_{\alpha\beta}\}$ such that $\phi_{\beta\alpha,x}(u) = \rho(g_{\alpha\beta}(x))$. Actually, we may first have to replace $\{U_\alpha, \phi_\alpha\}$ by a "refinement"; a notion we now define:

Appendix E

Useful ways to think about vector bundles

It is often useful to think about vector bundles as a family of vector spaces parameterized by some topological space. So we just have a continuously varying family $V_x : x \in X$. The fun comes when we explore the consequences of how these V_x fit together in a topological sense. We will study the idea of a tangent bundle below but for the case of a surface such as a sphere we can define the tangent bundle rather directly. For example, we may let $TS^2 = \{(x, v) : x \in S^2$ and $(x, v) = 0\}$. Thus the tangent vectors are pairs (x, v) where the x just tells us which point on the sphere we are at. The topology we put on TS^2 is the one we get by thinking of TS^2 as a subset of all pair, namely as a subset of $\mathbb{R}^3 \times \mathbb{R}^3$. The projection map is $\pi : (x, v) \mapsto x$.

Now there is another way we can associate a two dimensional vector space to each point of S^2. For this just associate a fixed copy of $\mathbb{R}^2$ to each $x \in S^2$ by taking the product space $S^2 \times \mathbb{R}^2$. This too is just a set of pairs but the topology is the product topology. Could it be that $S^2 \times \mathbb{R}^2 \to S^2$ and TS^2 are equivalent as vector bundles? The answer is no. The reason is that $S^2 \times \mathbb{R}^2 \to S^2$ is the trivial bundle and so have a nowhere zero section:

$$x \mapsto (x, v)$$

where v is any nonzero vector in $\mathbb{R}^2$. On the other hand it can be shown using the techniques of algebraic topology that TS^2 does not support a nowhere zero section. It follows that there can be no bundle isomorphism between the tangent bundle of the sphere $\pi : TS^2 \to S^2$ and the trivial bundle $pr_1 : S^2 \times \mathbb{R}^2 \to S^2$.

Recall that our original definition of a vector bundle (of rank k) did not include explicit reference to transition functions but they exist nonetheless. Accordingly, another way to picture a vector bundle is to think about it as a bunch of trivial bundles $U_\alpha \times \mathbb{R}^k$ (or $U_\alpha \times \mathbb{C}^k$ for a complex bundle)) together with "change of coordinate" maps: $U_\alpha \times \mathbb{R}^k \to U_\beta \times \mathbb{R}^k$ given by $(x, v) \mapsto (x, g_{\alpha\beta}(x)v)$. So a point (vector) in a vector bundle from $\alpha' s$ view is

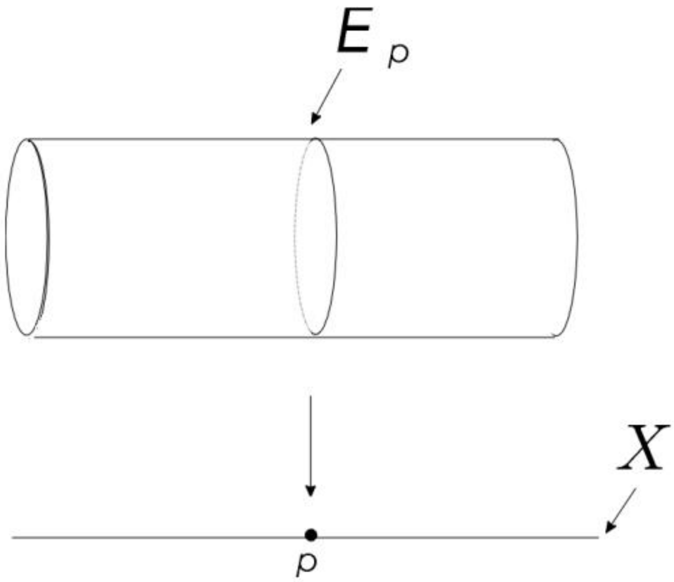
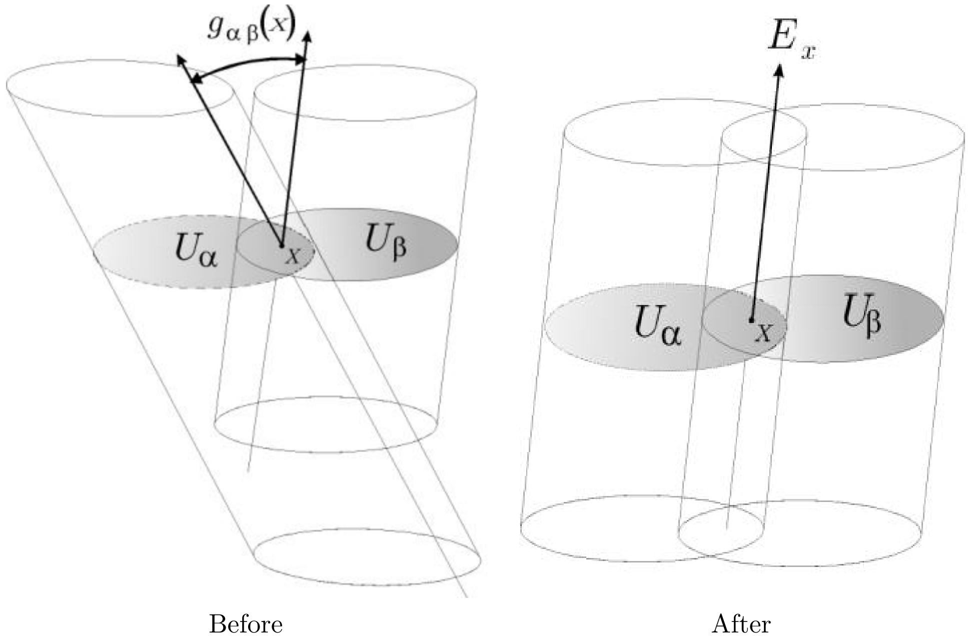


Figure E.1: Circle bundle. Schematic for fiber bundle.

a pair (x, v) while if $U_\alpha \cap U_\beta \neq \emptyset$ then from $\beta's$ viewpoint it is the pair (x, w) where $w = g_{\alpha\beta}(x)v$. Now if this really does describe a vector bundle then the maps $g_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(k, \mathbb{R})$ (or $\mathrm{GL}(k, \mathbb{C})$) must satisfy certain consistency relations:

$$
\begin{aligned}
g_{\alpha\alpha}(x) &= \mathrm{id} \ \text{ for } x \in U_\alpha \\
g_{\alpha\beta}(x)g_{\beta\alpha}(x) &= \mathrm{id} \ \text{ for } x \in U_\alpha \cap U_\beta \\
g_{\alpha\beta}(x)g_{\beta\gamma}(x)g_{\gamma\alpha}(x) &= \mathrm{id} \ \text{ for } x \in U_\alpha \cap U_\beta \cap U_\gamma.
\end{aligned}
\tag{E.1}
$$

A system of maps $g_{\alpha\beta}$ satisfying these relations is called a cocycle for the bundle atlas.



Example E.1 *Here is how the tangent bundle of the sphere looks in this picture:*
Let $U_{xy,+} = \{(x,y,z) \in S^2 : z > 0\}$
$U_{xy,+} = \{(x,y,z) \in S^2 : z > 0\}$

$$\begin{aligned} U_{z+} &= \{(x,y,z) \in S^2 : z > 0\} \\ U_{z-} &= \{(x,y,z) \in S^2 : z < 0\} \\ U_{y+} &= \{(x,y,z) \in S^2 : y > 0\} \\ U_{y+} &= \{(x,y,z) \in S^2 : y < 0\} \\ U_{x+} &= \{(x,y,z) \in S^2 : x > 0\} \\ U_{x-} &= \{(x,y,z) \in S^2 : x < 0\} \end{aligned}$$

Then for example $g_{y+,z+}(x,y,z)$ *is the matrix*

$$\begin{pmatrix} \frac{\partial}{\partial x} x & \frac{\partial}{\partial y} x \\ \frac{\partial}{\partial x}\sqrt{1-x^2-y^2} & \frac{\partial}{\partial y}\sqrt{1-x^2-y^2} \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 0 \\ -x/z & -y/z \end{pmatrix}$$

the rest are easily calculated using the Jacobian of the change of coordinate maps.

Or using just two open sets coming from stereographic projection: let U_1 *be the open subset* $\{(x,y,z) \in S^2 : z \neq -1\}$ *and let* U_2 *be the open subset*

$\{(x,y,z) \in S^2 : z \neq 1\}$. Now we take the two trivial bundles $U_1 \times \mathbb{R}^2$ and $U_2 \times \mathbb{R}^2$ and then describe $g_{12}(p)$ for $p = (x,y,z) \in U_1 \cap U_2$.

$$g_{12}(p).(v_1, v_2)^t = (w_1, w_2)^t$$

where $h_{12}(p)$ is the matrix

$$\begin{pmatrix} -(x^2-y^2) & -2xy \\ -2xy & x^2-y^2 \end{pmatrix}.$$

This last matrix is not so easy to recognize. Can you see where I got it?

Theorem E.1 *Let M be a C^r-space and suppose there is given an open cover $\{U_\alpha\}_{\alpha\in A}$ of M together with C^r-maps $g_{\alpha\beta} : U_\alpha \cap U_\beta \to GL(\mathbb{F}^k)$ satisfying the cocycle relations. Then there is a vector bundle $\pi_E : E \to M$ that has these as transition maps.*

E.1 Operations on Vector Bundles

Using the construction of Theorem E.2 we can build vector bundles from a given vector bundle in a way that globalizes the usual constructions of linear and multilinear algebra.

The first operation we consider is restriction of a vector bundle $\pi : E \to M$ to a submanifold $S \subset M$. The restriction $\pi|_S : E|_S \to S$ is the bundle whose fiber above $p \in S$ is $E_p = \pi^{-1}(p)$. Thus $E|_S = \bigsqcup_{p\in S} E_p$ and $\pi|_S$ is the restriction of π to $E|_S$.

Example E.2 *Show that $E|_S$ is a submanifold of E. Exhibit the natural VB-charts for $\pi|_S : E|_S \to S$.*

Let $E \to M$ be a vector bundle of rank n with a transition functions $\Phi_{\alpha\beta} : U_\alpha \cap U_\beta \to GL(\mathbb{R}^n)$.

E.2 Banach Vector Bundles

Fiber bundles may also be defined in case where the spaces involved are infinite dimensional manifolds. The theory is mostly the same as the finite dimensional case but there are a couple of technical considerations that we shall deal with as needed. We will only introduce vector bundles in this more general setting and only in the smooth case:

Definition E.1 *A (real or complex)* ***vector bundle*** *with typical fiber a Banach space E is a fiber bundle $\xi = (E, \pi_E, M, \mathsf{E})$ for each pair of bundle chart domains U_α and U_β with nonempty intersection, the map*

$$f_{\alpha\beta} : x \mapsto f_{\alpha\beta,x} := f_{\alpha,x} \circ f_{\beta,x}^{-1}$$

is a C^∞ morphism $g_{\alpha\beta} : U_\alpha \cap U_\beta \to GL(\mathsf{E})$.

Theorem E.2 *Let M be a smooth manifold and suppose there is given an open cover $\{U_\alpha\}_{\alpha\in A}$ of M together with maps $g_{\alpha\beta} : U_\alpha \cap U_\beta \to GL(\mathsf{E})$ satisfying the cocycle relations. Then there is a vector bundle $\pi_E : E \to M$ that has these as transition maps.*

Proof. Define an equivalence relation $\sim$ on the set

$$S = \{(x, \alpha, v) : \alpha \in A,\ x \in U_\alpha \text{ and } v \in \mathsf{E}\}$$

by requiring that $(x, \alpha, v) \sim (y, \beta, w)$ if and only if $x = y$ and $w = g_{\beta\alpha}(x)v$. Next, let $E = S/\sim$ and $\pi_E : E \to M$ be the map induced by $(x, \alpha, v) \mapsto x$. One can check that if $\{O_i, \psi_i\}$ is an atlas for M then $\{\pi_E^{-1}(O_i \cap U_\alpha), \psi_{\alpha,i}\}$ is an atlas for E where

$$\psi_{\alpha,i}[(x, \alpha, v)] = (x, \psi_i(x)) \in \mathsf{M} \times \mathsf{E}.$$

Also, the maps $\phi_\alpha : [(x, \alpha, v)] \mapsto (x, v)$ with domains $\pi_E^{-1}(U_\alpha)$ for $\alpha \in A$ give a VB-atlas on E with transition functions $\{g_{\alpha\beta}(x)\}$ as required. ■

Remark E.1 *If the manifolds involved are finite dimensional then instead of condition (4) we need only that $g_{\alpha\beta}(x) \in GL(\mathsf{E})$ the smoothness in x being automatic.*

Definition E.2 (Type II vector bundle charts) *A **(type II) vector bundle chart** on an open set $V \subset E$ is a fiber preserving diffeomorphism $\phi : V \to O \times \mathsf{E}$ that covers a diffeomorphism $\underline{\phi}:\pi_E(V) \to O$ in the sense that the following diagram commutes*

$$\begin{array}{cccc} & V & \xrightarrow{\phi} & O \times \mathsf{E} \\ \pi_E & \downarrow & & \downarrow pr_1 \\ & \pi_E(V) & \xrightarrow[\underline{\phi}]{} & O \end{array}$$

and that is a linear isomorphism on each fiber.

Example E.3 *The maps $T\psi_\alpha : TU_\alpha \to \psi_\alpha(U_\alpha) \times \mathsf{E}$ are (type II) VB-charts and so not only give TM a differentiable structure but also provide TM with a vector bundle structure. Similar remarks apply for T^*M.*

Let $\xi_1 = (E_1, \pi_1, M, \mathsf{E}_1)$ and $\xi_2 = (E_2, \pi_2, M, \mathsf{E}_2)$ be vector bundles locally isomorphic to $\mathsf{M}\times\ \mathsf{E}_1$ and $\mathsf{M}\times\ \mathsf{E}_2$ respectively. We say that the sequence of vector bundle morphisms

$$0 \to \xi_1 \xrightarrow{f} \xi_2$$

is **exact** if the following conditions hold:

1. There is an open covering of M by open sets U_α together with trivializations $\phi_{1,\alpha} : \pi_1^{-1}(U_\alpha) \to U_\alpha\times\ \mathsf{E}_1$ and $\phi_{2,\alpha} : \pi_2^{-1}(U_\alpha) \to U_\alpha\times\ \mathsf{E}_2$ such that $\mathsf{E}_2 = \mathsf{E}_1 \times \mathsf{F}$ for some Banach space F;

2. The diagram below commutes for all α :

$$\begin{array}{ccccc} & \pi_1^{-1}(U_\alpha) & \to & \pi_2^{-1}(U_\alpha) & \\ \phi_{1,\alpha} & \downarrow & & \downarrow & \phi_{2,\alpha} \\ & U_\alpha \times \mathsf{E}_1 & \to & U_\alpha \times \mathsf{E}_1 \times \mathsf{F} & \end{array}$$

Definition E.3 *A* ***subbundle*** *of a vector bundle $\xi = (E, \pi, M)$ is a vector bundle of the form $\xi = (L, \pi|_L, M)$ where $\pi|_L$ is the restriction to $L \subset E$, and where $L \subset E$ is a submanifold such that*

$$0 \to \xi|_L \to \xi$$

is exact. Here, $\xi|_L \to \xi$ is the bundle map given by inclusion: $L \hookrightarrow E$.

Equivalently, $\pi|_L : L \to M$ is a ***subbundle*** *if $L \subset E$ is a submanifold and there is a splitting $\mathsf{E} = \mathsf{E}_1 \times \mathsf{F}$ such that for each $p \in M$ there is a bundle chart $\phi : \pi^{-1}U \to U \times \mathsf{E}$ with $p \in U$ and $\phi((\pi^{-1}U) \cap L) = U \times \mathsf{E}_1 \times \{0\}$.*

Definition E.4 *The chart ϕ from the last definition is said to be adapted to the subbundle.*

Notice that if $L \subset E$ is as in the previous definition then $\pi|_L : L \to M$ is a vector bundle with VB-atlas given by the various VB-charts U, ϕ restricted to $(\pi^{-1}U) \cap S$ and composed with projection $U \times \mathsf{E}_1 \times \{0\} \to U \times \mathsf{E}_1$ so $\pi|_L$ is a bundle locally isomorphic to $M \times \mathsf{E}_1$. The fiber of $\pi|_L$ at $p \in L$ is $L_p = E_p \cap L$. Once again we remind the reader of the somewhat unfortunate fact that although the bundle includes and is indeed determined by the map $\pi|_L : L \to M$ we often refer to L itself as the subbundle.

E.3 Sections of a Vector Bundle

It is a general fact that if $\{V_x\}_{x \in X}$ is any family of $\mathbb{F}-$vector spaces (or even modules over a commutative ring) and $\mathcal{F}(X, \mathbb{F})$ is the vector space of all $\mathbb{F}$-valued functions on the index set X then the set of maps (sections) of the form $\sigma : x \mapsto \sigma(x) \in V_x$ is a module over $\mathcal{F}(X, \mathbb{F})$ the operation being the obvious one. For us the index set is some C^r-space and the family has the structure of a vector bundle. Let us restrict attention to C^∞ bundles. The space of sections $\Gamma(\xi)$ of a bundle $\xi = (\mathbb{F}^k \hookrightarrow E \longrightarrow M)$ is a vector space over $\mathbb{F}$ but also a module over the ring of smooth functions $C^\infty(M)$. The scalar multiplication is the obvious one: $(f\sigma)(x) := f(x)\sigma(x)$. Of course we also have local sections $\sigma : U \to E$ defined only on some open set. Piecing together local sections into global sections is whole topic in itself and leads to several topological constructions and in particular sheaf theory and sheaf cohomology. The first thing to observe about global sections of a vector bundle is that we always have plenty of them. But as we have seen, it is a different matter entirely if we are asking for global sections that never vanish; that is, sections σ for which $\sigma(x)$ is never the zero element of the fiber E_x. The existence or nonexistence of a global nowhere vanishing section of a bundle is an important topological fact about the bundle. In particular, since the tangent bundle of a manifold come right out of the definition of the manifold itself any fact about the tangent bundle is a fact about the manifold itself.

Appendix F

Overview of Classical Physics

F.0.1 Units of measurement

In classical mechanics we need units for measurements of length, time and mass. These are called elementary units. WE need to add a measure of electrical current to the list if we want to study electromagnetic phenomenon. Other relevant units in mechanics are derived from these alone. For example, speed has units of length$\times$time^{-1}, volume has units of length$\times$length$\times$length kinetic energy has units of mass$\times$length$\times$length$\times$length$\times$time$^{-1}\times$time^{-1} and so on. A common system, called the SI system uses meters (m), kilograms (km) and seconds (sec) for length, mass and time respectively. In this system, the unit of energy kg$\times$m^{2}sec^{-2} is called a joule. The unit of force in this system is Newtons and decomposes into elementary units as kg$\times$m$\times$sec^{-2}.

F.0.2 Newtons equations

The basic assumptions of Newtonian mechanics can be summarized by saying that the set of all mechanical events M taking place in ordinary three dimensional space is such that we can impose on this set of events a coordinate system called an inertial coordinate system. An inertial coordinate system is first of all a 1-1 correspondence between events and the vector space $\mathbb{R} \times \mathbb{R}^3$ consisting of 4-tuples (t, x, y, z). The laws of mechanics are then described by equations and expressions involving the variables (t, x, y, z) written $(t, \mathbf{x})$ where $\mathbf{x} =(x, y, z)$. There will be many correspondences between the event set and $\mathbb{R} \times \mathbb{R}^3$ but not all are inertial. An inertial coordinate system is picked out by the fact that the equations of physics take on a particularly simple form in such coordinates. Intuitively, the x, y, z variables locate an event in space while t specifies the time of an event. Also, x, y, z should be visualized as determined by measuring against a mutually perpendicular set of three axes and t is measured with respect to some sort of clock with $t = 0$ being chosen arbitrarily according to the demands of the experimental situation. Now we expect that the laws of physics should not prefer any particular such choice of mutually perpendicular axes or choice of starting time. Also, the units of measurement of length and time are conventionally determined by human beings and so the equations of the laws of physics should in some way not depend on this choice in any significant way. Careful consideration along these lines leads to a particular set of "coordinate changes" or transformations which translate among the different inertial coordinate systems. The group of transformations which is chosen for classical (non-relativistic) mechanics is the so called Galilean group Gal.

Definition F.1 *A map $g : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R} \times \mathbb{R}^3$ is called a Galilean transformation if and only if it can be decomposed as a composition of transformations of the following type:*

1. Translation of the origin:
$$(t, \mathbf{x}) \mapsto (t + t_0, \mathbf{x} + \mathbf{x}_0)$$
2. Uniform motion with velocity $\mathbf{v}$:
$$(t, \mathbf{x}) \mapsto (t, \mathbf{x}+t\mathbf{v})$$
3. Rotation of the spatial axes:
$$(t, \mathbf{x}) \mapsto (t, R\mathbf{x})$$
where $R \in O(3)$.

If $(t, \mathbf{x})$ are inertial coordinates then so will $(T, \mathbf{X})$ be inertial coordinates if and only if $(T, \mathbf{X}) = g(t, \mathbf{x})$ for some Galilean transformation. We will take this as given.

The motion of a idealized point mass moving in space is described in an inertial frame $(t, \mathbf{x})$ as a curve $t \mapsto c(t) \in \mathbb{R}^3$ with the corresponding curve $t \mapsto (t, c(t))$ in the (coordinatized) event space $\mathbb{R} \times \mathbb{R}^3$. We often write $\mathbf{x}(t)$ instead of $c(t)$. If we have a system of n particles then we may formally treat this as a single particle moving in an $3n-$dimensional space and so we have a single curve in $\mathbb{R}^{3n}$. Essentially we are concatenating the spatial part of inertial coordinates $\mathbb{R}^{3n} = \mathbb{R}^3 \times \cdots \mathbb{R}^3$ taking each factor as describing a single particle in the system so we take $\mathbf{x} = (x_1, y_1, z_1, ..., x_n, y_n, z_n)$. Thus our new inertial coordinates may be thought of as $\mathbb{R} \times \mathbb{R}^{3n}$. If we have a system of particles it will be convenient to define the momentum vector $\mathbf{p} = (m_1x_1, m_1y_1, m_1z_1, ..., m_nx_n, m_ny_n, m_nz_n) \in \mathbb{R}^{3n}$. In such coordinates, Newton's law for n particles of masses $m_1, ..., m_n$ reads
$$\frac{d^2\mathbf{p}}{dt^2} = \mathbf{F}(\mathbf{x}(t), t)$$
where $t \mapsto \mathbf{x}(t)$ describes the motion of a system of n particles in space as a smooth path in $\mathbb{R}^{3n}$ parameterized by t representing time. The equation has

units of force (Newtons in the SI system). If all bodies involved are taken into account then the force $\mathbf{F}$ cannot depend explicitly on time as can be deduced by the assumption that the form taken by $\mathbf{F}$ must be the same in any inertial coordinate system. We may not always be able to include explicitly all involved bodies and so it may be that our mathematical model will involve a changing force $\mathbf{F}$ exerted on the system from without as it were. As an example consider the effect of the tidal forces on sensitive objects on earth. Also, the example of earths gravity shows that if the earth is not taken into account as one of the particles in the system then the form of $\mathbf{F}$ will not be invariant under all spatial rotations of coordinate axes since now there is a preferred direction (up-down).

F.0.3 Classical particle motion in a conservative field

There are special systems for making measurements that can only be identified in actual practice by interaction with the physical environment. In classical mechanics, a point mass will move in a straight line unless a force is being applied to it. The coordinates in which the mathematical equations describing motion are the simplest are called inertial coordinates (x, y, z, t). If we consider a *single* particle of mass m then Newton's law simplifies to

$$m\frac{d^2\mathbf{x}}{dt^2} = \mathbf{F}(\mathbf{x}(t), t)$$

The force $\mathbf{F}$ is **conservative** if it doesn't depend on time and there is a potential function $V : \mathbb{R}^3 \to \mathbb{R}$ such that $\mathbf{F}(\mathbf{x}) = -\operatorname{grad} V(\mathbf{x})$. Assume this is the case. Then Newton's law becomes

$$m\frac{d^2}{dt^2}\mathbf{x}(t) + \operatorname{grad} V(\mathbf{x}(t)) = 0.$$

Newton's equations are often written

$$\mathbf{F}(\mathbf{x}(t)) = m\mathbf{a}(t)$$

$\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$ is the force function and we have taken it to not depend explicitly on time t. The force will be conservative so $\mathbf{F}(\mathbf{x}) = -\operatorname{grad} V(\mathbf{x})$ for some scalar function $V(\mathbf{x})$. The total energy or Hamiltonian function is a function of two vector variables $\mathbf{x}$ and $\mathbf{v}$ given (in this simple situation) by

$$H(\mathbf{x}, \mathbf{v}) = \frac{1}{2}m\left\|\mathbf{v}\right\|^2 + V(\mathbf{x})$$

so that if we plug in $\mathbf{x} = \mathbf{x}(t)$ and $\mathbf{v} = \mathbf{x}'(t)$ for the motion of a particle then we get the energy of the particle. Since this is a conservative situation $\mathbf{F}(\mathbf{x}) = -\operatorname{grad} V(\mathbf{x})$ we discover by differentiating and using equation **??** that $\frac{d}{dt}H(\mathbf{x}(t), \mathbf{x}'(t)) = 0$. This says that the total energy is conserved along any path which is a solution to equation **??** as long as $\mathbf{F}(\mathbf{x}) = -\operatorname{grad} V(\mathbf{x})$.

There is a lot of structure that can be discovered by translating the equations of motion into an arbitrary coordinate system (q^1, q^2, q^3) and then extending that to a coordinate system $(q^1, q^2, q^3, \dot{q}^1, \dot{q}^2, \dot{q}^3)$ for velocity space $\mathbb{R}^3 \times \mathbb{R}^3$. Here, $\dot{q}^1, \dot{q}^2, \dot{q}^3$ are not derivatives until we compose with a curve $\mathbb{R} \to \mathbb{R}^3$ to get functions of t. Then (and only then) we will take $\dot{q}^1(t), \dot{q}^2(t), \dot{q}^3(t)$ to be the derivatives. Sometimes $(\dot{q}^1(t), \dot{q}^2(t), \dot{q}^3(t))$ is called the **generalized velocity** vector. Its physical meaning depends on the particular form of the generalized coordinates.

In such a coordinate system we have a function $L(\mathbf{q}, \dot{\mathbf{q}})$ called the Lagrangian of the system. Now there is a variational principle that states that if $\mathbf{q}(t)$ is a path which solve the equations of motion and defined from time t_1 to time t_2 then out of all the paths which connect the same points in space at the same times t_1 and t_2 , the one that makes the following action the smallest will be the solution:

$$S(\mathbf{q}(t)) = \int_{t_1}^{t_2} L(\mathbf{q}, \dot{\mathbf{q}}) dt$$

Now this means that if we add a small variation to $\mathbf{q}$ get another path $\mathbf{q} + \delta\mathbf{q}$ then we calculate formally:

$$\begin{aligned} \delta S(\mathbf{q}(t)) &= \delta \int_{t_1}^{t_2} L(\mathbf{q}, \dot{\mathbf{q}}) dt \\ & \int_{t_1}^{t_2} \left[\delta\mathbf{q} \cdot \frac{\partial}{\partial \mathbf{q}} L(\mathbf{q}, \dot{\mathbf{q}}) + \delta\dot{\mathbf{q}} \cdot \frac{\partial}{\partial \dot{\mathbf{q}}} L(\mathbf{q}, \dot{\mathbf{q}}) \right] dt \\ &= \int_{t_1}^{t_2} \delta\mathbf{q} \cdot \left(\frac{\partial}{\partial \mathbf{q}} L(\mathbf{q}, \dot{\mathbf{q}}) - \frac{d}{dt} \frac{\partial}{\partial \dot{\mathbf{q}}} L(\mathbf{q}, \dot{\mathbf{q}}) \right) dt + \left[\delta\mathbf{q} \cdot \frac{\partial}{\partial \dot{\mathbf{q}}} L(\mathbf{q}, \dot{\mathbf{q}}) \right]_{t_1}^{t_2} \end{aligned}$$

If our variation is among those that start and end at the same space-time locations then $\delta\mathbf{q} = \mathbf{0}$ is the end points so the last term vanishes. Now if the path $\mathbf{q}(t)$ is stationary for such variations then $\delta S(\mathbf{q}(t)) = 0$ so

$$\int_{t_1}^{t_2} \delta\mathbf{q} \cdot \left(\frac{\partial}{\partial \mathbf{q}} L(\mathbf{q}, \dot{\mathbf{q}}) - \frac{d}{dt} \frac{\partial}{\partial \dot{\mathbf{q}}} L(\mathbf{q}, \dot{\mathbf{q}}) \right) dt = 0$$

and since this is true for all such paths we conclude that

$$\frac{\partial}{\partial \mathbf{q}} L(\mathbf{q}, \dot{\mathbf{q}}) - \frac{d}{dt} \frac{\partial}{\partial \dot{\mathbf{q}}} L(\mathbf{q}, \dot{\mathbf{q}}) = \mathbf{0}$$

or in indexed scalar form

$$\frac{\partial L}{\partial q^i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} = 0 \text{ for } 1 \leq i \leq 3$$

on a stationary path. This is (these are) the **Euler-Lagrange equation**(s). If $\mathbf{q}$ were just rectangular coordinates and if L were $\frac{1}{2} m \|\mathbf{v}\|^2 - V(\mathbf{x})$ this turns out to be Newton's equation. Notice, the minus sign in front of the V.

Definition F.2 *For a Lagrangian L we can associate the quantity $E = \sum \frac{\partial L}{\partial \dot{q}^i} \dot{q}^i - L(\mathbf{q}, \dot{\mathbf{q}})$.*

Let us differentiate E. We get

$$\begin{aligned}
\frac{d}{dt}E &= \frac{d}{dt}\sum \frac{\partial L}{\partial \dot{q}^i}\dot{q}^i - L(\mathbf{q},\dot{\mathbf{q}}) \\
&= \frac{\partial L}{\partial \dot{q}^i}\frac{d}{dt}\dot{q}^i - \dot{q}^i\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{d}{dt}L(\mathbf{q},\dot{\mathbf{q}}) \\
&= \frac{\partial L}{\partial \dot{q}^i}\frac{d}{dt}\dot{q}^i - \dot{q}^i\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i}\dot{q}^i - \frac{\partial L}{\partial \dot{q}^i}\frac{d}{dt}\dot{q}^i \\
&= 0 \;\; \text{by the Euler Lagrange equations.}
\end{aligned} \tag{F.1}$$

Conclusion F.1 *If L does not depend explicitly on time; $\frac{\partial L}{\partial t} = 0$, then the* ***energy** E **is conserved ;*** *$\frac{dE}{dt} = 0$ along any solution of the Euler-Lagrange equations..*

But what about spatial symmetries? Suppose that $\frac{\partial}{\partial q^i}L = 0$ for one of the coordinates q^i. Then if we define $p_i = \frac{\partial L}{\partial \dot{q}^i}$ we have

$$\frac{d}{dt}p_i = \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} = -\frac{\partial}{\partial q^i}L = 0$$

so p_i is constant along the trajectories of Euler's equations of motion. The quantity $p_i = \frac{\partial L}{\partial \dot{q}^i}$ is called a generalized momentum and we have reached the following

Conclusion F.2 *If $\frac{\partial}{\partial q^i}L = 0$ then p_i is a conserved quantity. This also applies if $\frac{\partial}{\partial \mathbf{q}}L = (\frac{\partial L}{\partial q^1}, ..., \frac{\partial L}{\partial q^n}) = 0$ with the conclusion that the vector $\mathbf{p} = \frac{\partial}{\partial \mathbf{q}}L = (\frac{\partial L}{\partial \dot{q}^1}, ..., \frac{\partial L}{\partial \dot{q}^n})$ is conserved (each component separately).*

Now let us apply this to the case a free particle. The Lagrangian in rectangular inertial coordinates are

$$L(\mathbf{x},\dot{\mathbf{x}}) = \frac{1}{2}m\,|\dot{\mathbf{x}}|^2$$

and this Lagrangian is symmetric with respect to translations $\mathbf{x} \mapsto \mathbf{x} + \mathbf{c}$

$$L(\mathbf{x}+\mathbf{c},\dot{\mathbf{x}}) = L(\mathbf{x},\dot{\mathbf{x}})$$

and so the generalized momentum vector for this is $\mathbf{p} = m\dot{\mathbf{x}}$ each component of which is conserved. This last quantity is actually the usual momentum vector.

Now let us examine the case where the Lagrangian is invariant with respect to rotations about some fixed point which we will take to be the origin of an inertial coordinate system. For instance suppose the potential function $V(\mathbf{x})$ is invariant in the sense that $V(\mathbf{x}) = V(O\mathbf{x})$ for any orthogonal matrix O. The we can take an antisymmetric matrix A and form the family of orthogonal matrices e^{sA}. The for the Lagrangian

$$L(\mathbf{x},\dot{\mathbf{x}}) = \frac{1}{2}m\,|\dot{\mathbf{x}}|^2 - V(\mathbf{x})$$

we have

$$\begin{aligned}\frac{d}{ds}L(e^{sA}\mathbf{x},e^{sA}\dot{\mathbf{x}}) &= \frac{d}{dt}(\frac{1}{2}m\left|e^{sA}\dot{\mathbf{x}}\right|^2 - V(e^{sA}\mathbf{x})) \\ &= \frac{d}{dt}(\frac{1}{2}m\left|\dot{\mathbf{x}}\right|^2 - V(\mathbf{x})) = 0\end{aligned}$$

On the other hand, recall the result of a variation $\delta\mathbf{q}$

$$\int_{t_1}^{t_2} \delta\mathbf{q}\cdot\left(\frac{\partial}{\partial\mathbf{q}}L(\mathbf{q},\dot{\mathbf{q}}) - \frac{d}{dt}\frac{\partial}{\partial\dot{\mathbf{q}}}L(\mathbf{q},\dot{\mathbf{q}})\right)dt + \left[\delta\mathbf{q}\cdot\frac{\partial}{\partial\dot{\mathbf{q}}}L(\mathbf{q},\dot{\mathbf{q}})\right]_{t_1}^{t_2}$$

what we have done is to let $\delta\mathbf{q} = A\mathbf{q}$ since to first order we have $e^{sA}\mathbf{q} = I + sA\mathbf{q}$. But if $\mathbf{q}(t)$ satisfies Euler's equation then the integral above is zero and yet the whole variation is zero too. We are led to conclude that

$$\left[\delta\mathbf{q}\cdot\frac{\partial}{\partial\dot{\mathbf{q}}}L(\mathbf{q},\dot{\mathbf{q}})\right]_{t_1}^{t_2} = 0$$

which in the present case is

$$\begin{aligned}\left[A\mathbf{x}\cdot\frac{\partial}{\partial\dot{\mathbf{x}}}(\frac{1}{2}m\left|\dot{\mathbf{x}}\right|^2 - V(\mathbf{x}))\right]_{t_1}^{t_2} &= 0 \\ \left[mA\mathbf{x}\cdot\dot{\mathbf{x}}\right]_{t_1}^{t_2} &= 0\end{aligned}$$

for all t_2 and t_1. Thus the quantity $mA\mathbf{x}\cdot\dot{\mathbf{x}}$ is conserved. Let us apply this with A equal to the following in turn

$$A = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

then we get $mA\mathbf{x}\cdot\dot{\mathbf{x}} = m(-x^2, x^1, 0)\cdot(\dot{x}^1, \dot{x}^2, \dot{x}^3) = m(x^1\dot{x}^2 - \dot{x}^1x^2)$ which is the same as $m\dot{\mathbf{x}}\times\mathbf{k} = \mathbf{p}\times\mathbf{k}$ which is called the angular momentum about the $\mathbf{k}$ axis ($\mathbf{k} = (0,0,1)$ so this is the z-axis) and is a conserved quantity. To see the point here notice that

$$e^{tA} = \begin{bmatrix} \cos t & -\sin t & 0 \\ \sin t & \cos t & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is the rotation about the z-axis. We can do the same thing for the other two coordinate axes and in fact it turns out that for any unit vector $\mathbf{u}$ the angular momentum about that axis defined by $\mathbf{p}\times\mathbf{u}$ is conserved.

Remark F.1 *We started with the assumption that L was invariant under all rotations O but if it had only been invariant under counterclockwise rotations about an axis given by a unit vector* $\mathbf{u}$ *then we could still conclude that at least* $\mathbf{p}\times\mathbf{u}$ *is conserved.*

Remark F.2 *Let begin to use the index notation (like q^i, p_i and x^i etc.) a little more since it will make the transition to fields more natural.*

Now we define the Hamiltonian function derived from a given Lagrangian via the formulas

$$H(\mathbf{q},\mathbf{p}) = \sum p_i \dot{q}^i - L(\mathbf{q},\dot{\mathbf{q}})$$
$$p_i = \frac{\partial L}{\partial \dot{q}^i}$$

where we think of $\dot{\mathbf{q}}$ as depending on $\mathbf{q}$ and $\mathbf{p}$ via the inversion of $p_i = \frac{\partial L}{\partial \dot{q}^i}$. Now it turns out that if $\mathbf{q}(t), \dot{\mathbf{q}}(t)$ satisfy the Euler Lagrange equations for L then $\mathbf{q}(t)$ and $\mathbf{p}(t)$ satisfy the Hamiltonian equations of motion

$$\frac{dq^i}{dt} = \frac{\partial H}{\partial p_i}$$
$$\frac{dp^i}{dt} = -\frac{\partial H}{\partial q^i}$$

One of the beauties of this formulation is that if $Q^i = Q^i(q^j)$ are any other coordinates on $\mathbb{R}^3$ and we define $P^i = p^j \frac{\partial Q^i}{\partial q^j}$ then taking $H(..q^i.,..p^i..) = \widetilde{H}(..Q^i..,..P_i..)$ the equations of motion have the same form in the new coordinates. More generally, if Q, P are related to q, p in such a way that the Jacobian matrix J of the coordinate change (on $\mathbb{R}^3 \times \mathbb{R}^3$) is **symplectic**

$$J^t \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$$

then the equations **??** will hold in the new coordinates. These kind of coordinate changes on the q, p space $\mathbb{R}^3 \times \mathbb{R}^3$ (momentum space) are called canonical transformations. Mechanics is, in the above sense, invariant under canonical transformations.

Next, take any smooth function $f(q,p)$ on momentum space (also called phase space). Such a function is called an observable. Then along any solution curve $(q(t), p(t))$ to Hamilton's equations we get

$$\begin{aligned} \frac{df}{dt} &= \frac{\partial f}{\partial q}\frac{dq}{dt} + \frac{\partial f}{\partial p}\frac{dp}{dt} \\ &= \frac{\partial f}{\partial q^i}\frac{\partial H}{\partial p^i} + \frac{\partial f}{\partial p^i}\frac{\partial H}{\partial q^i} \\ &= [f, H] \end{aligned}$$

where we have introduced the Poisson bracket $[f, H]$ defined by the last equality above. So we also have the equations of motion in the form $\frac{df}{dt} = [f, H]$ for any function f not just the coordinate functions q and p. Later, we shall study a geometry hiding here; **Symplectic geometry**.

Remark F.3 *For any coordinate $t, \mathbf{x}$ we will often consider the curve $(\mathbf{x}(t), \mathbf{x}'(t)) \in \mathbb{R}^{3n} \times \mathbb{R}^{3n}$ the latter product space being a simple example of a* ***velocity phase space****.*

F.0.4 Some simple mechanical systems

1. As every student of basic physics know the equations of motion for a particle falling freely through a region of space near the earths surface where the force of gravity is (nearly) constant is $\mathbf{x}''(t) = -g\mathbf{k}$ where $\mathbf{k}$ is the usual vertical unit vector corresponding to a vertical z-axis. Integrating twice gives the form of any solution $\mathbf{x}(t) = -\frac{1}{2}gt^2\mathbf{k}+t\mathbf{v}_0+\mathbf{x}_0$ for constant vectors $\mathbf{x}_0, \mathbf{v}_0 \in \mathbb{R}^3$. We get different motions depending on the initial conditions $(\mathbf{x}_0, \mathbf{v}_0)$. If the initial conditions are right, for example if $\mathbf{v}_0 = 0$ then this is reduced to the one dimensional equation $x''(t) = -g$. The path of a solution with initial conditions (x_0, v_0) is given in phase space as
$$t \mapsto (-\frac{1}{2}gt^2 + tv_0 + x_0, -gt + v_0)$$
and we have shown the phase trajectories for a few initial conditions.

2. A somewhat general 1-dimensional system is given by a Lagrangian of the form
$$L = \frac{1}{2}a(q)\dot{q}^2 - V(q) \tag{F.2}$$
and example of which is the motion of a particle of mass m along a 1-dimensional continuum and subject to a potential $V(x)$. Then the Lagrangian is $L = \frac{1}{2}m\dot{x}^2 - V(x)$. Instead of writing down the Euler-Lagrange equations we can use the fact that $E = \frac{\partial L}{\partial \dot{x}^i}\dot{x}^i - L(x,\dot{x}) = m\dot{x}^2 - (\frac{1}{2}m\dot{x}^2 - V(x)) = \frac{1}{2}m\dot{x}^2 + V(x)$ is conserved. This is the total energy which is traditionally divided into kinetic energy $\frac{1}{2}m\dot{x}^2$ and potential energy $V(x)$. We have $E = \frac{1}{2}m\dot{x}^2 + V(x)$ for some constant. Then
$$\frac{dx}{dt} = \sqrt{\frac{2E - 2V(x)}{m}}$$
and so
$$t = \sqrt{m/2}\int \frac{1}{\sqrt{E - V(x)}} + c.$$
Notice that we must always have $E - V(x) \geq 0$. This means that if $V(x)$ has a single local minimum between some points $x = a$ and $x = b$ where $E - V = 0$, then the particle must stay between $x = a$ and $x = b$ moving back and forth with some time period. What is the time period?.

3. **Central Field.** A central field is typically given by a potential of the form $V(\mathbf{x}) = -\frac{k}{|\mathbf{x}|}$. Thus the Lagrangian of a particle of mass m in this central field is
$$\frac{1}{2}m\left|\dot{\mathbf{x}}\right|^2 + \frac{k}{|\mathbf{x}|}$$

where we have centered inertial coordinates at the point where the potential has a singularity $\lim_{\mathbf{x}\to 0} V(\mathbf{x}) = \pm\infty$. In cylindrical coordinates (r, θ, z) the Lagrangian becomes

$$\frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2 + \dot{z}^2) + \frac{k}{(r^2 + z^2)^{1/2}}.$$

We are taking $q^1 = r$, $q^2 = \theta$ and $q^3 = \theta$. But if initially $z = \dot{z} = 0$ then by conservation of angular momentum discussed above the particle stays in the $z = 0$ plane. Thus we are reduced to a study of the two dimensional case:

$$\frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) + \frac{k}{r}.$$

What are Lagrange's equations? Answer:

$$\begin{aligned} 0 &= \frac{\partial L}{\partial q^1} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^1} \\ &= mr\dot{\theta}^2 - \frac{k}{r^2} - m\dot{r}\ddot{r} \end{aligned}$$

and

$$\begin{aligned} 0 &= \frac{\partial L}{\partial q^2} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^2} \\ &= -mr^2\dot{\theta}\ddot{\theta}. \end{aligned}$$

The last equation reaffirms that $\dot{\theta} = \omega_0$ is constant. Then the first equation becomes $mr\omega_0^2 - \frac{k}{r^2} - m\dot{r}\ddot{r} = 0$. On the other hand conservation of energy becomes

4.

$$\begin{aligned} \frac{1}{2}m(\dot{r}^2 + r^2\omega_0^2) + \frac{k}{r} = E_0 = \frac{1}{2}m(\dot{r}_0^2 + r_0^2\omega_0^2) + \frac{k}{r_0} \qquad &\text{or} \\ \dot{r}^2 + r^2\omega_0^2 + \frac{2k}{mr} = \frac{2E_0}{m} \end{aligned}$$

5. A simple oscillating system is given by $\frac{d^2x}{dt^2} = -x$ which has solutions of the form $x(t) = C_1 \cos t + C_2 \sin t$. This is equivalent to the system

$$\begin{aligned} x' &= v \\ v' &= -x \end{aligned}$$

6. Consider a single particle of mass m which for some reason is viewed with respect to rotating frame and an inertial frame (taken to be stationary). The rotating frame $(\mathbf{E}_1(t), \mathbf{E}_2(t), \mathbf{E}_3(t)) = \mathtt{E}$ (centered at the origin of R^3) is related to stationary frame $(e_1, e_2, e_3) = \mathtt{e}$ by an orthogonal matrix O :

$$\mathtt{E}(t) = \mathrm{O}(t)\mathtt{e}$$

and the rectangular coordinates relative to these frames are related by

$$\mathbf{x}(t) = \mathrm{O}(t)\mathbf{X}(t)$$

We then have

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathrm{O}(t)\dot{\mathbf{X}} + \dot{\mathrm{O}}(t)\mathbf{X}\\ &= \mathrm{O}(t)(\dot{\mathbf{X}} + \Omega(t)\mathbf{X})\end{aligned}$$

where $\Omega(t) = \mathrm{O}^t(t)\dot{\mathrm{O}}(t)$ is an angular velocity. The reason we have chosen to work with $\Omega(t)$ rather than directly with $\dot{\mathrm{O}}(t)$ will become clearer later in the book. Let us define the operator D_t by $D_t\mathbf{X} = \dot{\mathbf{X}} + \Omega(t)\mathbf{X}$. This is sometimes called the "total derivative". At any rate the equations of motion in the inertial frame is of the form $m\frac{d\mathbf{x}}{dt} = \mathbf{f}(\dot{\mathbf{x}}, \mathbf{x})$. In the moving frame this becomes an equation of the form

$$m\frac{d}{dt}(\mathrm{O}(t)(\dot{\mathbf{X}} + \Omega(t)\mathbf{X})) = \mathbf{f}(\mathrm{O}(t)\mathbf{X}, \mathrm{O}(t)(\dot{\mathbf{X}} + \Omega(t)\mathbf{X}))$$

and in turn

$$\mathrm{O}(t)\frac{d}{dt}(\dot{\mathbf{X}}+\Omega(t)\mathbf{X})+\dot{\mathrm{O}}(t)(\dot{\mathbf{X}}+\Omega(t)\mathbf{X}) = m^{-1}\mathbf{f}(\mathrm{O}(t)\mathbf{X}, \mathrm{O}(t)(\dot{\mathbf{X}}+\Omega(t)\mathbf{X})).$$

Now recall the definition of D_t we get

$$\mathrm{O}(t)(\frac{d}{dt}D_t\mathbf{X} + \Omega(t)D_t\mathbf{X}) = m^{-1}\mathbf{f}(\mathrm{O}(t)\mathbf{X}, \mathrm{O}(t)\mathbf{V})$$

and finally

$$mD_t^2\mathbf{X} = \mathbf{F}(\mathbf{X}, \mathbf{V}) \tag{F.3}$$

where we have defined the **relative velocity** $\mathbf{V} = \dot{\mathbf{X}}+\Omega(t)\mathbf{X}$ and $\mathbf{F}(\mathbf{X}, \mathbf{V})$ is by definition the transformed force $\mathbf{f}(\mathrm{O}(t)\mathbf{X}, \mathrm{O}(t)\mathbf{V})$. The equation we have derived would look the same in any moving frame: It is a covariant expression.

5. Rigid Body We will use this example to demonstrate how to work with the rotation group and it's Lie algebra. The advantage of this approach is that it generalizes to motions in other Lie groups and their algebra's. Let us denote the group of orthogonal matrices of determinant one by SO(3). This is the rotation group. If the Lagrangian of a particle as in the last example is invariant under actions of the orthogonal group so that $L(\mathbf{x}, \dot{\mathbf{x}}) = L(Qx, Q\dot{x})$ for $Q \in \mathrm{SO}(3)$ then the quantity $\ell = \mathbf{x} \times m\dot{\mathbf{x}}$ is constant for the motion of the particle $\mathbf{x} = \mathbf{x}(t)$ satisfying the equations of motion in the inertial frame. The matrix group SO(3) is an example of a Lie group which we study intensively in later chapters. Associated with every Lie group is its Lie algebra which in this case is the set of all anti-symmetric 3x3 matrices denoted $\mathfrak{so}(3)$. There is an interesting correspondence between and $\mathbb{R}^3$ given by

$$\begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix} \leftrightarrows (\omega_1, \omega_2, \omega_3) = \omega$$

Furthermore if we define the bracket for matrices A and B in $\mathfrak{so}(3)$ by $[A, B] = AB - BA$ then under the above correspondence $[A, B]$ corresponds to the cross product. Let us make the temporary convention that if x is an element of $\mathbb{R}^3$ then the corresponding matrix in $\mathfrak{so}(3)$ will be denoted by using the same letter but a new font while lower case refers to the inertial frame and upper to the moving frame:

$$\mathbf{x} \leftrightarrows \mathsf{x} \in \mathfrak{so}(3) \text{ and}$$
$$\mathbf{X} \leftrightarrows \mathsf{X} \in \mathfrak{so}(3) \text{ etc.}$$

	$\mathbb{R}^3$		$\mathfrak{so}(3)$
Inertial frame	$\mathbf{x}$	$\leftrightarrows$	x
Moving frame	$\mathbf{X}$	$\leftrightarrows$	X

Then we have the following chart showing how various operations match up:

$$\begin{array}{ccc}
\mathbf{x} = \mathrm{O}\mathbf{X} & \leftrightarrows & \mathsf{x} = \mathrm{O}\mathsf{X}\mathrm{O}^t \\
\mathbf{v}_1 \times \mathbf{v}_2 & \leftrightarrows & [\mathsf{v}_1, \mathsf{v}_2] \\
\mathbf{v} = \dot{\mathbf{x}} & \leftrightarrows & \mathsf{v}=\dot{\mathsf{x}} \\
\mathbf{V} = D_t\mathbf{X} = \dot{\mathbf{X}} + \Omega(t)\mathbf{X} & \leftrightarrows & \mathsf{V}=D_t\mathsf{X} = \dot{\mathsf{X}}+[\Omega(t), \mathsf{X}] \\
\ell = \mathbf{x} \times m\dot{\mathbf{x}} & \leftrightarrows & \mathsf{l} = [\mathsf{x}, m\dot{\mathsf{x}}] \\
\ell = \mathrm{O}\mathbf{L} & \leftrightarrows & \mathsf{l} = \mathrm{O}\mathsf{L}\mathrm{O}^t = [\mathsf{V}, \Omega(t)] \\
D_t\mathbf{L} = \dot{\mathbf{L}} + \Omega(t) \times \mathbf{L} & \leftrightarrows & D_t\mathsf{L} = \dot{\mathsf{L}} + [\Omega(t), \mathsf{L}]
\end{array}$$

and so on. Some of the quantities are actually defined by their position in this chart. In any case, let us differentiate $\mathbf{l} = \mathbf{x} \times m\dot{\mathbf{x}}$ and use the equations of motion to get

$$\begin{aligned}\frac{d\mathbf{l}}{dt} &= \mathbf{x} \times m\dot{\mathbf{x}} \\ &= \mathbf{x} \times m\ddot{\mathbf{x}} + \mathbf{0} \\ &= \mathbf{x} \times \mathbf{f}.\end{aligned}$$

But we have seen that if the Lagrangian (and hence the force $\mathbf{f}$) is invariant under rotations that $\frac{d\mathbf{l}}{dt} = 0$ along any solution curve. Let us examine this case. We have $\frac{d\mathbf{l}}{dt} = 0$ and in the moving frame $D_t\mathbf{L} = \dot{\mathbf{L}} + \Omega(t)\mathbf{L}$. Transferring the equations over to our $\mathfrak{so}(3)$ representation we have $D_t\mathsf{L} = \dot{\mathsf{L}} + [\Omega(t), \mathsf{L}] = 0$. Now if our particle is rigidly attached to the rotating frame, that is, if $\dot{\mathbf{x}} = 0$ then $\dot{\mathsf{X}} = 0$ and $\mathsf{V} = [\Omega(t), \mathsf{X}]$ so

$$\mathsf{L} = m[\mathsf{X}, [\Omega(t), \mathsf{X}]].$$

In Lie algebra theory the map $\mathsf{v} \mapsto [\mathsf{x}, \mathsf{v}] = -[\mathsf{v}, \mathsf{x}]$ is denoted $\mathrm{ad}(\mathsf{x})$ and is linear. With this notation the above becomes

$$\mathsf{L} = -m\,\mathrm{ad}(\mathsf{X})\Omega(t).$$

The map $I : \mathsf{X} \mapsto -m\,\mathrm{ad}(\mathsf{X})\Omega(t) = I(\mathsf{X})$ is called the momentum operator. Suppose now that we have k particles of masses $m_1, m_2, ...m_2$ each at rigidly attached to the rotating frame and each giving quantities $\mathsf{x}_i, \mathsf{X}_i$ etc. Then to total angular momentum is $\sum I(\mathsf{X}_i)$. Now if we have a continuum of mass with mass density ρ in a moving region B_t (a rigid body) then letting $\mathsf{X}_{\mathbf{u}}(t)$ denote path in $\mathfrak{so}(3)$ of the point of initially at $\mathbf{u} \in B_0 \in \mathbb{R}^3$ then we can integrate to get the total angular momentum at time t;

$$\mathsf{L}_{tot}(t) = -\int_B \mathrm{ad}(\mathsf{X}_{\mathbf{u}}(t))\Omega(t)d\rho(\mathbf{u})$$

which is a conserved quantity.

F.0.5 The Basic Ideas of Relativity

F.0.6 Variational Analysis of Classical Field Theory

In field theory we study functions $\phi : \mathbb{R}\times\mathbb{R}^3 \to \mathbb{R}^k$. We use variables $\phi(x^0, x^1, x^2, x^3) = \phi(t, x, y, z)$ A Lagrangian density is a function $\mathcal{L}(\phi, \partial\phi)$ and then the Lagrangian would be

$$L(\phi, \partial\phi) = \int_{V\subset\mathbb{R}^3} \mathcal{L}(\phi, \partial\phi)d^3x$$

and the action is

$$S = \int\int_{V\subset\mathbb{R}^3} \mathcal{L}(\phi, \partial\phi)d^3x dt = \int_{V\times I\subset\mathbb{R}^4} \mathcal{L}(\phi, \partial\phi)d^4x$$

What has happened is that the index i is replaced by the space variable $\vec{x} = (x^1, x^2, x^3)$ and we have the following translation

$$\begin{array}{lll} i & \rightarrowtail\rightarrowtail\rightarrowtail & \vec{x} \\ q & \rightarrowtail\rightarrowtail\rightarrowtail & \phi \\ q^i & \rightarrowtail\rightarrowtail\rightarrowtail & \phi(., \vec{x}) \\ q^i(t) & \rightarrowtail\rightarrowtail\rightarrowtail & \phi(t, \vec{x}) = \phi(x) \\ p^i(t) & \rightarrowtail\rightarrowtail\rightarrowtail & \partial_t \phi(t, \vec{x}) + \nabla_{\vec{x}} \phi(t, \vec{x}) = \partial\phi(x) \end{array}$$

$$\begin{array}{lll} L(q, p) & \rightarrowtail\rightarrowtail\rightarrowtail & \int_{V \subset \mathbb{R}^3} \mathcal{L}(\phi, \partial\phi) d^3x \\ S = \int L(\mathbf{q}, \dot{\mathbf{q}}) dt & \rightarrowtail\rightarrowtail\rightarrowtail & S = \int\int \mathcal{L}(\phi, \partial\phi) d^3x dt \end{array}$$

where $\partial\phi = (\partial_0\phi, \partial_1\phi, \partial_2\phi, \partial_3\phi)$. So in a way, the mechanics of classical massive particles is classical field theory on the space with three points which is the set $\{1, 2, 3\}$. Or we can view field theory as infinitely many particle systems indexed by points of space. In other words, a system with an infinite number of degrees of freedom.

Actually, we have only set up the formalism of scalar fields and have not, for instance, set things up to cover internal degrees of freedom like spin. However, we will discuss spin later in this text. Let us look at the formal variational calculus of field theory. We let $\delta\phi$ be a variation which we might later assume to vanish on the boundary of some region in space-time $U = I \times V \subset \mathbb{R} \times \mathbb{R}^3 = \mathbb{R}^4$. In general, we have

$$\begin{aligned} \delta S &= \int_U \left(\delta\phi \frac{\partial\mathcal{L}}{\partial\phi} + \partial_\mu \delta\phi \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \right) d^4x \\ &= \int_U \partial_\mu \left(\delta\phi \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \right) d^4x + \int_U \delta\phi \left(\frac{\partial\mathcal{L}}{\partial\phi} - \partial_\mu \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \right) d^4x \end{aligned}$$

Now the first term would vanish by the divergence theorem if $\delta\phi$ vanished on the boundary ∂U. If ϕ were a field that were stationary under such variations then

$$\delta S = \int_U \delta\phi \left(\frac{\partial\mathcal{L}}{\partial\phi} - \partial_\mu \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \right) d^4x = 0$$

for all $\delta\phi$ vanishing on ∂U so we can conclude that Lagrange's equation holds for ϕ stationary in this sense and visa versa:

$$\frac{\partial\mathcal{L}}{\partial\phi} - \partial_\mu \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} = 0$$

These are the field equations.

F.0.7 Symmetry and Noether's theorem for field theory

Now an interesting thing happens if the Lagrangian density is invariant under some set of transformations. Suppose that $\delta\phi$ is an infinitesimal "internal" symmetry of the Lagrangian density so that $\delta S(\delta\phi) = 0$ even though $\delta\phi$ does not vanish on the boundary. Then if ϕ is already a solution of the field equations then

$$0 = \delta S = \int_U \partial_\mu \left(\delta\phi \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) d^4x$$

for all regions U. This means that $\partial_\mu \left(\delta\phi \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) = 0$ so if we define $j^\mu = \delta\phi \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}$ we get

$$\partial_\mu j^\mu = 0$$

or

$$\frac{\partial}{\partial t} j^0 = -\nabla \cdot \overrightarrow{\mathbf{j}}$$

where $\overrightarrow{\mathbf{j}} = (j^1, j^2, j^3)$ and $\nabla \cdot \overrightarrow{\mathbf{j}} = \operatorname{div}(\overrightarrow{\mathbf{j}})$ is the spatial divergence. This looks like some sort of conservation.. Indeed, if we define the total charge at any time t by

$$Q(t) = \int j^0 d^3x$$

the assuming $\overrightarrow{\mathbf{j}}$ shrinks to zero at infinity then the divergence theorem gives

$$\begin{aligned} \frac{d}{dt} Q(t) &= \int \frac{\partial}{\partial t} j^0 d^3x \\ &= -\int \nabla \cdot \overrightarrow{\mathbf{j}} d^3x = 0 \end{aligned}$$

so the charge $Q(t)$ is a conserved quantity. Let $Q(U, t)$ denote the total charge inside a region U. The charge inside any region U can only change via a flux through the boundary:

$$\begin{aligned} \frac{d}{dt} Q(U, t) &= \int_U \frac{\partial}{\partial t} j^0 d^3x \\ &= \int_{\partial U} \overrightarrow{\mathbf{j}} \cdot \mathbf{n} dS \end{aligned}$$

which is a kind of "local conservation law". To be honest the above discussion only takes into account so called internal symmetries. An example of an internal symmetry is given by considering a curve of linear transformations of $\mathbb{R}^k$ given as matrices $C(s)$ with $C(0) = I$. Then we vary ϕ by $C(s)\phi$ so that $\delta\phi = \frac{d}{ds}\big|_0 C(s)\phi = C'(0)\phi$. Another possibility is to vary the underlying space so that $C(s, .)$ is now a curve of transformations of $\mathbb{R}^4$ so that if $\phi_s(x) = \phi(C(s, x))$ is a variation of fields then we must take into account the fact that the domain of integration is also varying:

$$L(\phi_s, \partial\phi_s) = \int_{U_s \subset \mathbb{R}^4} \mathcal{L}(\phi_s, \partial\phi_s) d^4x$$

We will make sense of this later.

F.0.8 Electricity and Magnetism

Up until now it has been mysterious how any object of matter could influence any other. It turns out that most of the forces we experience as middle sized objects pushing and pulling on each other is due to a single electromagnetic force. Without the help of special relativity there appears to be two forces; electric and magnetic. Elementary particles that carry electric charges such as electrons or protons, exert forces on each other by means of a field. In a particular Lorentz frame, the electromagnetic field is described by a skew-symmetric matrix of functions called the electromagnetic field tensor:

$$(F_{\mu\nu}) = \begin{bmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -B_z & B_y \\ -E_y & B_z & 0 & -B_x \\ -E_z & -B_y & B_x & 0 \end{bmatrix}.$$

Where we also have the forms $F_{\mu}^{\nu} = \Lambda^{s\nu} F_{\mu s}$ and $F^{\mu\nu} = \Lambda^{s\mu} F_{s}^{\nu}$. This tensor can be derived from a potential $\mathsf{A} = (A_0, A_1, A_2, A_3)$ by $F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}$. The contravariant form of the potential is $(A_0, -A_1, -A_2, -A_3)$ is a four vector often written as

$$\mathsf{A} = (\phi, \overrightarrow{\mathbf{A}}).$$

The action for a charged particle in an electromagnetic field is written in terms of A in a manifestly invariant way as

$$\int_a^b -mcd\tau - \frac{e}{c} A_\mu dx^\mu$$

so writing $\mathsf{A} = (\phi, \overrightarrow{\mathbf{A}})$ we have

$$S = \int_a^b (-mc\frac{d\tau}{dt} - e\phi(t) + \overrightarrow{\mathbf{A}} \cdot \frac{d\tilde{\mathbf{x}}}{dt})dt$$

so in a given frame the Lagrangian is

$$L(\tilde{\mathbf{x}}, \frac{d\tilde{\mathbf{x}}}{dt}, t) = -mc^2\sqrt{1-(v/c)^2} - e\phi(t) + \overrightarrow{\mathbf{A}} \cdot \frac{d\tilde{\mathbf{x}}}{dt}.$$

Remark F.4 *The system under study is that of a particle in a field and does not describe the dynamics of the field itself. For that we would need more terms in the Lagrangian.*

This is a time dependent Lagrangian because of the $\phi(t)$ term but it turns out that one can re-choose A so that the new $\phi(t)$ is zero and yet still have $F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}$. This is called change of gauge. Unfortunately, if we wish to express things in such a way that a constant field is given by a constant potential then we cannot make this choice. In any case, we have

$$L(\overrightarrow{\mathbf{x}}, \frac{d\overrightarrow{\mathbf{x}}}{dt}, t) = -mc^2\sqrt{1-(v/c)^2} - e\phi + \overrightarrow{\mathbf{A}} \cdot \frac{d\overrightarrow{\mathbf{x}}}{dt}$$

and setting $\overrightarrow{\mathbf{v}}=\frac{d\tilde{\mathbf{x}}}{dt}$ and $\left|\overrightarrow{\mathbf{v}}\right| = v$ we get the follow form for energy

$$\overrightarrow{\mathbf{v}}\cdot\frac{\partial}{\partial\overrightarrow{\mathbf{v}}}L(\tilde{\mathbf{x}},\overrightarrow{\mathbf{v}},t) - L(\tilde{\mathbf{x}},\overrightarrow{\mathbf{v}},t) = \frac{mc^2}{\sqrt{1-(v/c)^2}} + e\phi.$$

Now this is not constant with respect to time because $\frac{\partial L}{\partial t}$ is not identically zero. On the other hand, this make sense from another point of view; the particle is interacting with the field and may be picking up energy from the field.

The Euler-Lagrange equations of motion turn out to be

$$\frac{d\tilde{\mathbf{p}}}{dt} = e\tilde{\mathbf{E}} + \frac{e}{c}\overrightarrow{\mathbf{v}} \times \tilde{\mathbf{B}}$$

where $\tilde{\mathbf{E}} = -\frac{1}{c}\frac{\partial\tilde{\mathbf{A}}}{\partial t} - \operatorname{grad}\phi$ and $\tilde{\mathbf{B}} = curl\tilde{\mathbf{A}}$ are the electric and magnetic parts of the field respectively. This decomposition into electric and magnetic parts is an artifact of the choice of inertial frame and may be different in a different frame. Now the momentum $\tilde{\mathbf{p}}$ is $\frac{m\overrightarrow{\mathbf{v}}}{\sqrt{1-(v/c)^2}}$ but a speeds $v << c$ this becomes nearly equal to $m\mathbf{v}$ so the equations of motion of a charged particle reduce to

$$m\frac{d\overrightarrow{\mathbf{v}}}{dt} = e\tilde{\mathbf{E}} + \frac{e}{c}\overrightarrow{\mathbf{v}} \times \tilde{\mathbf{B}}.$$

Notice that is the particle is not moving, or if it is moving parallel the magnetic field $\tilde{\mathbf{B}}$ then the second term on the right vanishes.

The electromagnetic field equations.

We have defined the 3-vectors $\tilde{\mathbf{E}} = -\frac{1}{c}\frac{\partial\tilde{\mathbf{A}}}{\partial t} - \operatorname{grad}\phi$ and $\tilde{\mathbf{B}} = \operatorname{curl}\tilde{\mathbf{A}}$ but since the curl of a gradient is zero it is easy to see that $\operatorname{curl}\tilde{\mathbf{E}} = -\frac{1}{c}\frac{\partial\tilde{\mathbf{B}}}{\partial t}$. Also, from $\tilde{\mathbf{B}} = \operatorname{curl}\tilde{\mathbf{A}}$ we get $\operatorname{div}\tilde{\mathbf{B}} = \mathbf{0}$. This easily derived pair of equations is the first two of the four famous Maxwell's equations. Later we will see that the electromagnetic field tensor is really a differential 2-form F and these two equations reduce to the statement that the (exterior) derivative of F is zero:

$$dF = 0$$

Exercise F.1 *Apply Gauss's theorem and stokes theorem to the first two Maxwell's equations to get the integral forms. What do these equations say physically?*

One thing to notice is that these two equations do not determine $\frac{\partial}{\partial t}\tilde{\mathbf{E}}$.

Now we have not really written down a action or Lagrangian that includes terms that represent the field itself. When that part of the action is added in we get

$$S = \int_a^b (-mc - \frac{e}{c}A_\mu\frac{dx^\mu}{d\tau})d\tau + a\int_V F^{\nu\mu}F_{\nu\mu}dx^4$$

where in so called Gaussian system of units the constant a turns out to be $\frac{-1}{16\pi c}$. Now in a particular Lorentz frame and recalling 20.5 we get $= a\int_V F^{\nu\mu}F_{\nu\mu}dx^4 = \frac{1}{8\pi}\int_V \left|\tilde{\mathbf{E}}\right|^2 - \left|\tilde{\mathbf{B}}\right|^2 dtdxdydz.$

In order to get a better picture in mind let us now assume that there is a continuum of charged particle moving through space and that volume density of charge at any given moment in space-time is ρ so that if $dxdydz = dV$ then ρdV is the charge in the volume dV. Now we introduce the four vector $\rho\mathsf{u} = \rho(d\mathsf{x}/d\tau)$ where u is the velocity 4-vector of the charge at (t, x, y, z). Now recall that $\rho d\mathsf{x}/d\tau = \frac{d\tau}{dt}(\rho, \rho\overrightarrow{\mathbf{v}}) = \frac{d\tau}{dt}(\rho, \tilde{\mathbf{j}}) = \mathsf{j}$. Here $\tilde{\mathbf{j}} = \rho\overrightarrow{\mathbf{v}}$ is the charge current density as viewed in the given frame a vector field varying smoothly from point to point. Write $\mathsf{j} = (j^0, j^1, j^2, j^3)$.

Assuming now that the particle motion is determined and replacing the discrete charge e be the density we have applying the variational principle with the region $U = [a, b] \times V$ says

$$\begin{aligned} 0 &= -\delta\left(\int_V \int_a^b \frac{\rho dV}{c} dV A_\mu \frac{dx^\mu}{d\tau} d\tau + a\int_U F^{\nu\mu}F_{\nu\mu}dx^4\right) \\ &= -\delta\left(\frac{1}{c}\int_U j^\mu A_\mu + aF^{\nu\mu}F_{\nu\mu}dx^4\right) \end{aligned}$$

Now the Euler-Lagrange equations become

$$\frac{\partial\mathcal{L}}{\partial A_\nu} - \partial_\mu \frac{\partial\mathcal{L}}{\partial(\partial_\mu A_\nu)} = 0$$

where $\mathcal{L}(A_\mu, \partial_\mu A_\eta) = \frac{\rho}{c} A_\mu \frac{dx^\mu}{dt} + aF^{\nu\mu}F_{\nu\mu}$ and $F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}$. If one is careful to remember that $\partial_\mu A_\nu = \frac{\partial A_\nu}{\partial x^\mu}$ is to be treated as an independent variable one cane arrive at some complicated looking equations and then looking at the matrix 20.5 we can convert the equations into statements about the fields $\tilde{\mathbf{E}}$, $\tilde{\mathbf{B}}$,and $(\rho, \tilde{\mathbf{j}})$. We will not carry this out since we later discover a much more efficient formalism for dealing with the electromagnetic field. Namely, we will use differential forms and the Hodge star operator. At any rate the last two of Maxwell's equations read

$$\begin{aligned} \operatorname{curl} \tilde{\mathbf{B}} &= 0 \\ \operatorname{div} \tilde{\mathbf{E}} &= 4\pi\rho. \end{aligned}$$

F.0.9 Quantum Mechanics

Appendix G

Calculus on Banach Spaces

Mathematics is not only real, but it is the only reality. That is that the entire universe is made of matter is obvious. And matter is made of particles. It's made of electrons and neutrons and protons. So the entire universe is made out of particles. Now what are the particles made out of? They're not made out of anything. The only thing you can say about the reality of an electron is to cite its mathematical properties. So there's a sense in which matter has completely dissolved and what is left is just a mathematical structure.

Gardner on Gardner: JPBM Communications Award Presentation. Focus-The Newsletter of the Mathematical Association of America v. 14, no. 6, December 1994.

G.0.10 Differentiability

For simplicity and definiteness all Banach spaces in this section will be real Banach spaces. First, the reader will recall that a linear map on a normed space, say $A : \mathsf{V}_1 \to \mathsf{V}_2$, is bounded if and only it is continuous at one and therefore any point in V_1. Given two Banach spaces V_1 and V_2 with norms $\|.\|_1$ and $\|.\|_2$ we can form a Banach space from the Cartesian product $\mathsf{V}_1 \times \mathsf{V}_2$ by using the norm $\|(v,u)\| := \max\{\|v\|_1, \|u\|_2\}$. There are many equivalent norms for $\mathsf{V}_1 \times \mathsf{V}_2$ including

$$\|(v,u)\|' := \sqrt{\|v\|_1^2 + \|u\|_2^2}$$
$$\|(v,u)\|'' := \|v\|_1 + \|u\|_2 .$$

Recall that two norms on V, say $\|.\|'$and $\|.\|''$ are equivalent if there exist positive constants c and C such that

$$c\,\|x\|' \leq \|x\|'' \leq C\,\|x\|'$$

for all $x \in \mathsf{V}$. Also, if V is a Banach space and W_1 andW_2 are closed subspaces such that $\mathsf{W}_1 \cap \mathsf{W}_2 = \{0\}$ and such that every $v \in \mathsf{V}$ can be written uniquely in the form $v = w_1 + w_2$ where $w_1 \in \mathsf{W}_1$ and $w_2 \in \mathsf{W}_2$ then we write $\mathsf{V} = \mathsf{W}_1 \oplus \mathsf{W}_2$.

In this case there is the natural continuous linear isomorphism $\mathsf{W}_1 \times \mathsf{W}_2 \cong \mathsf{W}_1 \oplus \mathsf{W}_2$ given by

$$(w_1, w_2) \longleftrightarrow w_1 + w_2.$$

When it is convenient, we can identify $\mathsf{W}_1 \oplus \mathsf{W}_2$ with $\mathsf{W}_1 \times \mathsf{W}_2$ and in this case we hedge our bets, as it were, and write $w_1 + w_2$ for either (w_1, w_2) or $w_1 + w_2$ and let the context determine the precise meaning if it matters. Under the representation (w_1, w_2) we need to specify what norm we are using and there is more than one natural choice. We take $\|(w_1, w_2)\| := \max\{\|w_1\|, \|w_2\|\}$ but equivalent norms include, for example, $\|(w_1, w_2)\|_2 := \sqrt{\|w_1\|^2 + \|w_2\|^2}$ which is a good choice if the spaces happen to be Hilbert spaces.

Let E be a Banach space and $\mathsf{W} \subset \mathsf{E}$ a *closed* subspace. We say that W is **complemented**if there is a *closed* subspace W' such that $\mathsf{E} = \mathsf{W} \oplus \mathsf{W}'$. We also say that W is a **split subspace** of E.

Definition G.1 (Notation) *We will denote the set of all continuous (bounded) linear maps from a Banach space* E *to a Banach space* F *by* $L(\mathsf{E}, \mathsf{F})$*. The set of all continuous linear isomorphisms from* E *onto* F *will be denoted by* $GL(\mathsf{E}, \mathsf{F})$*. In case,* $\mathsf{E} = \mathsf{F}$ *the corresponding spaces will be denoted by* $\mathfrak{gl}(\mathsf{E})$ *and* $GL(\mathsf{E})$*. Here* $GL(\mathsf{E})$ *is a group under composition and is called the general linear group* .

Definition G.2 *Let* V_i*,* $i = 1, ..., k$ *and* W *be Banach spaces. A map* $\mu : \mathsf{V}_1 \times \cdots \times \mathsf{V}_k \to \mathsf{W}$ *is called multilinear (k-multilinear) if for each* i*,* $1 \leq i \leq k$ *and each fixed* $(w_1, ..., \widehat{w_i}, ..., w_k) \in \mathsf{V}_1 \times \cdots \times \widehat{\mathsf{V}_1} \times \cdots \times \mathsf{V}_k$ *we have that the map*

$$v \mapsto \mu(w_1, ..., \underset{i-th}{v}, ..., w_{k-1}),$$

obtained by fixing all but the i-th variable, is a bounded linear map. In other words, we require that μ *be* R*- linear in each slot separately.*

A multilinear map $\mu : \mathsf{V}_1 \times \cdots \times \mathsf{V}_k \to \mathsf{W}$ is said to be **bounded** if and only if there is a constant C such that

$$\|\mu(v_1, v_2, ..., v_k)\|_{\mathsf{W}} \leq C \|v_1\|_{\mathsf{E}_1} \|v_2\|_{\mathsf{E}_2} \cdots \|v_k\|_{\mathsf{E}_k}$$

for all $(v_1, ..., v_k) \in \mathsf{E}_1 \times \cdots \times \mathsf{E}_k$.

Notation G.1 *The set of all bounded multilinear maps* $\mathsf{E}_1 \times \cdots \times \mathsf{E}_k \to \mathsf{W}$ *will be denoted by* $L(\mathsf{E}_1, ..., \mathsf{E}_k; \mathsf{W})$*. If* $\mathsf{E}_1 = \cdots = \mathsf{E}_k = \mathsf{E}$ *then we write* $L^k(\mathsf{E}; \mathsf{W})$ *instead of* $L(\mathsf{E}, ..., \mathsf{E}; \mathsf{W})$

Definition G.3 *Let* E *be a Hilbert space with inner product denoted by* $\langle ., . \rangle$ *. Then* $O(\mathsf{E})$ *denotes the group of linear isometries from* E *onto itself. That is, the bijective linear maps* $\Phi : \mathsf{E} \to \mathsf{E}$ *such that* $\langle \Phi v, \Phi w \rangle = \langle v, w \rangle$ *for all* $v, w \in \mathsf{E}$*. The group* $O(\mathsf{E})$ *is called the* ***orthogonal group*** *(or sometimes the* ***Hilbert group*** *in the infinite dimensional case).*

Notation G.2 *For linear maps $T : \mathsf{V} \to \mathsf{W}$ we sometimes write $T \cdot v$ instead of $T(v)$ depending on the notational needs of the moment. In fact, a particularly useful notational device is the following: Suppose we have map $A : X \to L(\mathsf{V};\mathsf{W})$. Then $A(x)\cdot\ v$ makes sense but if instead of A the map needed to be indexed or something then things would get quite crowded. All in all it is sometimes better to write $\left.A\right|_x v$. In fact, if we do this then $\left.A\right|_x (v)$is also clear.*

Definition G.4 *A (bounded) multilinear map $\mu : \mathsf{V} \times \cdots \times \mathsf{V} \to \mathsf{W}$ is called* ***symmetric*** *(resp.* ***skew-symmetric*** *or* ***alternating****) if and only if for any $v_1, v_2, ..., v_k \in \mathsf{V}$ we have that*

$$\begin{aligned} \mu(v_1, v_2, ..., v_k) &= K(v_{\sigma 1}, v_{\sigma 2}, ..., v_{\sigma k}) \\ \textit{resp. } \mu(v_1, v_2, ..., v_k) &= sgn(\sigma)\mu(v_{\sigma 1}, v_{\sigma 2}, ..., v_{\sigma k}) \end{aligned}$$

for all permutations σ on the letters $\{1, 2, ..., k\}$. The set of all bounded symmetric (resp. skew-symmetric) multilinear maps $\mathsf{V} \times \cdots \times \mathsf{V} \to \mathsf{W}$ is denoted $L^k_{sym}(\mathsf{V};\mathsf{W})$ (resp. $L^k_{skew}(\mathsf{V};\mathsf{W})$ or $L^k_{alt}(\mathsf{V};\mathsf{W})$).

Now the space $L(\mathsf{V},\mathsf{W})$ is a Banach space in its own right with the norm

$$\|l\| = \sup_{v\in\mathsf{V}} \frac{\|l(v)\|_{\mathsf{W}}}{\|v\|_{\mathsf{V}}} = \sup\{\|l(v)\|_{\mathsf{W}} : \|v\|_{\mathsf{V}} = 1\}.$$

The spaces $L(\mathsf{E}_1, ..., \mathsf{E}_k; \mathsf{W})$ are also Banach spaces normed by

$$\|\mu\| := \sup\{\|\mu(v_1, v_2, ..., v_k)\|_{\mathsf{W}} : \|v_i\|_{\mathsf{E}_i} = 1 \text{ for } i = 1, .., k\}$$

Proposition G.1 *A k-multilinear map $\mu \in L(\mathsf{E}_1, ..., \mathsf{E}_k; \mathsf{W})$ is continuous if and only if it is bounded.*

Proof. ($\Leftarrow$) We shall simplify by letting $k = 2$. Let (a_1, a_2) and (v_1, v_2) be elements of $\mathsf{E}_1 \times \mathsf{E}_2$ and write

$$\begin{aligned} &\mu(v_1, v_2) - \mu(a_1, a_2) \\ &= \mu(v_1 - a_1, v_2) + \mu(a_1, v_2 - a_2). \end{aligned}$$

We then have

$$\begin{aligned} &\|\mu(v_1, v_2) - \mu(a_1, a_2)\| \\ &\leq C\|v_1 - a_1\|\|v_2\| + C\|a_1\|\|v_2 - a_2\| \end{aligned}$$

and so if $\|(v_1, v_2) - (a_1, a_2)\| \to 0$ then $\|v_i - a_i\| \to 0$ and we see that

$$\|\mu(v_1, v_2) - \mu(a_1, a_2)\| \to 0.$$

(Recall that $\|(v_1, v_2)\| := \max\{\|v_1\|, \|v_2\|\}$).

($\Rightarrow$) Start out by assuming that μ is continuous at $(0,0)$. Then for $r > 0$ sufficiently small, $(v_1, v_2) \in B((0,0), r)$ implies that $\|\mu(v_1, v_2)\| \leq 1$ so if for $i = 1, 2$ we let

$$z_i := \frac{rv_i}{\|v_1\|_i + \epsilon} \text{ for some } \epsilon > 0$$

then $(z_1, z_2) \in B((0,0), r)$ and $\|\mu(z_1, z_2)\| \leq 1$. The case $(v_1, v_2) = (0,0)$ is trivial so assume $(v_1, v_2) \neq (0,0)$. Then we have

$$\begin{aligned}\mu(z_1, z_2) &= \mu(\frac{rv_1}{\|v_1\| + \epsilon}, \frac{rv_2}{\|v_2\| + \epsilon}) \\ &= \frac{r^2}{(\|v_1\| + \epsilon)(\|v_2\| + \epsilon)}\mu(v_1, v_2) \leq 1\end{aligned}$$

and so $\mu(v_1, v_2) \leq r^{-2}(\|v_1\| + \epsilon)(\|v_2\| + \epsilon)$. Now let $\epsilon \to 0$ to get the result. ∎

We shall need to have several Banach spaces handy for examples. For the next example we need some standard notation.

Notation G.3 *In the context of $\mathbb{R}^n$, we often use the so called "multiindex notation". Let $\alpha = (\alpha_1, ..., \alpha_n)$ where the α_i are integers and $0 \leq \alpha_i \leq n$. Such an n-tuple is called a multiindex. Let $|\alpha| := \alpha_1 + ... + \alpha_n$*

$$\frac{\partial^\alpha f}{\partial x^\alpha} := \frac{\partial^{|\alpha|} f}{\partial (x^1)^{\alpha_1} \partial (x^1)^{\alpha_2} \cdots \partial (x^1)^{\alpha_n}}.$$

Example G.1 *Consider a bounded open subset Ω of $\mathbb{R}^n$. Let $L_k^p(\Omega)$ denote the Banach space obtained by taking the Banach space completion of the set $C^k(\Omega)$ of k-times continuously differentiable real valued functions on Ω with the norm given by*

$$\|f\|_{k,p} := \left(\int_\Omega \sum_{|\alpha| \leq k} \left| \frac{\partial^\alpha}{\partial x^\alpha} f(x) \right|^p \right)^{1/p}$$

Note that in particular $L_0^p(\Omega) = L^p(\Omega)$ is the usual L^p-space from real analysis.

Exercise G.1 *Show that the map $C^k(\Omega) \to C^{k-1}(\Omega)$ given by $f \mapsto \frac{\partial f}{\partial x^i}$ is bounded if we use the norms $\|f\|_{2,p}$ and $\|f\|_{2-1,p}$. Show that we may extend this to a bounded map $L_2^p(\Omega) \to L_1^p(\Omega)$.*

Proposition G.2 *There is a natural linear isomorphism $L(\mathsf{V}, L(\mathsf{V}, \mathsf{W})) \cong L^2(\mathsf{V}, \mathsf{W})$ given by*

$$l(v_1)(v_2) \longleftrightarrow l(v_1, v_2)$$

and we identify the two spaces. In fact, $L(\mathsf{V}, L(\mathsf{V}, L(\mathsf{V}, \mathsf{W})) \cong L^3(\mathsf{V}; \mathsf{W})$ and in general $L(\mathsf{V}, L(\mathsf{V}, L(\mathsf{V}, ..., L(\mathsf{V}, \mathsf{W})) \cong L^k(\mathsf{V}; \mathsf{W})$ etc.

Proof. It is easily checked that if we just define $(\iota\, T)(v_1)(v_2) = T(v_1, v_2)$ then $\iota\, T \leftrightarrow T$ does the job for the $k = 2$ case. The $k > 2$ case can be done by an inductive construction and is left as an exercise. It is also not hard to show that the isomorphism is continuous and in fact, norm preserving. ∎

Definition G.5 *A function $f : U \subset \mathsf{V} \to \mathsf{W}$ between Banach spaces and defined on an open set $U \subset \mathsf{V}$ is said to be* ***differentiable at*** *$p \in U$ if and only if there is a bounded linear map $A_p \in L(\mathsf{V}, \mathsf{W})$ such that*

$$\lim_{\|\mathsf{h}\| \to 0} \frac{f(p + \mathsf{h}) - f(p) - A_p \cdot \mathsf{h}}{\|\mathsf{h}\|} = 0$$

In anticipation of the following proposition we write $A_p = Df(p)$. We will also use the notation $Df|_p$ or sometimes $f'(p)$. The linear map $Df(p)$ is called the ***derivative*** *of f at p.*

We often write $Df|_p \cdot \mathsf{h}$. The dot in the notation just indicate a linear dependence and is not a literal "dot product". We could also write $Df(p)(\mathsf{h})$.

Exercise G.2 *Show that the map $F : L^2(\Omega) \to L^1(\Omega)$ given by $F(f) = f^2$ is differentiable at any $f_0 \in L^2(\Omega)$.*

Proposition G.3 *If A_p exists for a given function f then it is unique.*

Proof. Suppose that A_p and B_p both satisfy the requirements of the definition. That is the limit in question equals zero. For $p + \mathsf{h} \in U$ we have

$$\begin{aligned} A_p \cdot \mathsf{h} - B_p \cdot \mathsf{h} &= (f(p + \mathsf{h}) - f(p) - A_p \cdot \mathsf{h}) \\ &\quad - (f(p + \mathsf{h}) - f(p) - B_p \cdot \mathsf{h}). \end{aligned}$$

Dividing by $\|\mathsf{h}\|$ and taking the limit as $\|\mathsf{h}\| \to 0$ we get

$$\|A_p \mathsf{h} - B_p \mathsf{h}\| / \|\mathsf{h}\| \to 0$$

Now let $\mathsf{h} \neq 0$ be arbitrary and choose $\epsilon > 0$ small enough that $p + \epsilon\mathsf{h} \in U$. Then we have

$$\|A_p(\epsilon\mathsf{h}) - B_p(\epsilon\mathsf{h})\| / \|\epsilon\mathsf{h}\| \to 0.$$

But by linearity $\|A_p(\epsilon\mathsf{h}) - B_p(\epsilon\mathsf{h})\| / \|\epsilon\mathsf{h}\| = \|A_p\mathsf{h} - B_p\mathsf{h}\| / \|\mathsf{h}\|$ which doesn't even depend on ϵ so in fact $\|A_p\mathsf{h} - B_p\mathsf{h}\| = 0$. ■

If we are interested in differentiating "in one direction at a time" then we may use the natural notion of directional derivative. A map has a directional derivative $D_{\mathsf{h}}f$ at p in the direction h if the following limit exists:

$$(D_{\mathsf{h}}f)(p) := \lim_{\epsilon \to 0} \frac{f(p + \epsilon\mathsf{h}) - f(p)}{\epsilon}$$

In other words, $D_{\mathsf{h}}f(p) = \frac{d}{dt}\big|_{t=0} f(p+t\mathsf{h})$. But a function may have a directional derivative in every direction (at some fixed p), that is, for every $\mathsf{h} \in \mathsf{E}$ and yet still not be differentiable at p in the sense of definition 1.7.

Notation G.4 *The directional derivative is written as $(D_{\mathsf{h}}f)(p)$ and in case f is actually differentiable at p equal to $Df|_p \mathsf{h} = Df(p) \cdot \mathsf{h}$ (the proof is easy). Look closely; $D_{\mathsf{h}}f$ should not be confused with $Df|_{\mathsf{h}}$.*

Definition G.6 *If it happens that a function f is differentiable for all p throughout some open set U then we say that f is differentiable on U. We then have a map $Df : U \subset \mathsf{V} \to L(\mathsf{V},\mathsf{W})$ given by $p \mapsto Df(p)$. If this map is differentiable at some $p \in \mathsf{V}$ then its derivative at p is denoted $DDf(p) = D^2f(p)$ or $D^2f\big|_p$ and is an element of $L(\mathsf{V}, L(\mathsf{V},\mathsf{W})) \cong L^2(\mathsf{V};\mathsf{W})$. Similarly, we may inductively define $D^kf \in L^k(\mathsf{V};\mathsf{W})$ whenever f is sufficiently nice that the process can continue.*

Definition G.7 *We say that a map $f : U \subset \mathsf{V} \to \mathsf{W}$ is C^r-differentiable on U if $D^rf|_p \in L^r(\mathsf{V},\mathsf{W})$ exists for all $p \in U$ and if continuous D^rf as map $U \to L^r(\mathsf{V},\mathsf{W})$. If f is C^r-differentiable on U for all $r > 0$ then we say that f is C^∞ or* ***smooth*** *(on U).*

Exercise G.3 *Show directly that a bounded multilinear map is C^∞.*

Definition G.8 *A bijection f between open sets $U_\alpha \subset \mathsf{V}$ and $U_\beta \subset \mathsf{W}$ is called a C^r-**diffeomorphism** if and only if f and f^{-1} are both C^r-differentiable (on U_α and U_β respectively). If $r = \infty$ then we simply call f a diffeomorphism. Often, we will have $\mathsf{W} = \mathsf{V}$ in this situation.*

Let U be open in V. A map $f : U \to \mathsf{W}$ is called a ***local C^r diffeomorphism*** *if and only if for every $p \in U$ there is an open set $U_p \subset U$ with $p \in U_p$ such that $f|_{U_p} : U_p \to f(U_p)$ is a C^r-diffeomorphism.*

Remark G.1 *In the context of undergraduate calculus courses we are used to thinking of the derivative of a function at some $a \in \mathbb{R}$ as a number $f'(a)$ which is the slope of the tangent line on the graph at $(a, f(a))$. From the current point of view $Df(a) = Df|_a$ just gives the linear transformation $h \mapsto f'(a) \cdot h$ and the equation of the tangent line is given by $y = f(a) + f'(a)(x - a)$. This generalizes to an arbitrary differentiable map as $y = f(a) + Df(a) \cdot (x - a)$ giving a map which is the linear approximation of f at a.*

We will sometimes think of the derivative of a curve[1] $c : I \subset \mathbb{R} \to \mathsf{E}$ at $t_0 \in I$, written $\dot{c}(t_0)$, as a velocity vector and so we are identifying $\dot{c}(t_0) \in L(\mathbb{R},\mathsf{E})$ with $Dc|_{t_0} \cdot 1 \in \mathsf{E}$. Here the number 1 is playing the role of the unit vector in $\mathbb{R}$.

Let $f : U \subset \mathsf{E} \to \mathsf{F}$ be a map and suppose that we have a splitting $\mathsf{E} = \mathsf{E}_1 \times \mathsf{E}_2$. We will write $f(x, y)$ for $(x, y) \in \mathsf{E}_1 \times \mathsf{E}_2$. Now for every $(a, b) \in \mathsf{E}_1 \times \mathsf{E}_2$ the partial map $f_{a,} : y \mapsto f(a, y)$ (resp. $f_{,b} : x \mapsto f(x, b)$) is defined in some neighborhood of b (resp. a). We define the partial derivatives when they exist by $D_2f(a, b) = Df_{a,}(b)$ (resp. $D_1f(a, b) = Df_{,b}(a)$). These are, of course, linear maps.

$$D_1f(a, b) : \mathsf{E}_1 \to \mathsf{F}$$
$$D_2f(a, b) : \mathsf{E}_2 \to \mathsf{F}$$

The partial derivative can exist even in cases where f might not be differentiable in the sense we have defined. The point is that f might be differentiable only in certain directions.

[1] We will often use the letter I to denote a generic (usually open) interval in the real line.

If f has continuous partial derivatives $D_i f(x,y) : \mathsf{E}_i \to \mathsf{F}$ near $(x,y) \in \mathsf{E}_1 \times \mathsf{E}_2$ then exists and is continuous for all directions v. In this case, we have for $\mathrm{v} = (\mathrm{v}_1, \mathrm{v}_2)$,

$$\begin{aligned}&Df(x,y) \cdot (\mathrm{v}_1, \mathrm{v}_2)\\ &= D_1 f(x,y) \cdot \mathrm{v}_1 + D_2 f(x,y) \cdot \mathrm{v}_2\end{aligned}$$

G.0.11 Chain Rule, Product rule and Taylor's Theorem

Theorem G.1 (Chain Rule) *Let U_1 and U_2 be open subsets of Banach spaces E_1 and E_2 respectively. Suppose we have continuous maps composing as*

$$U_1 \xrightarrow{f} U_2 \xrightarrow{g} \mathsf{E}_3$$

where E_3 is a third Banach space. If f is differentiable at p and g is differentiable at $f(p)$ then the composition is differentiable at p and $D(g \circ f) = Dg(f(p)) \circ Dg(p)$. In other words, if $v \in \mathsf{E}_1$ then

$$D(g \circ f)|_p \cdot v = Dg|_{f(p)} \cdot (Df|_p \cdot v).$$

Furthermore, if $f \in C^r(U_1)$ and $g \in C^r(U_2)$ then $g \circ f \in C^r(U_1)$.

Proof. Let us use the notation $O_1(v)$, $O_2(v)$ etc. to mean functions such that $O_i(v) \to 0$ as $\|v\| \to 0$. Let $y = f(p)$. Since f is differentiable at p we have $f(p + \mathsf{h}) = y + Df|_p \cdot \mathsf{h} + \|\mathsf{h}\| O_1(\mathsf{h}) := y + \Delta y$ and since g is differentiable at y we have $g(y + \Delta y) = Dg|_y \cdot (\Delta y) + \|\Delta y\| O_2(\Delta y)$. Now $\Delta y \to 0$ as $\mathsf{h} \to 0$ and in turn $O_2(\Delta y) \to 0$ hence

$$\begin{aligned}g \circ f(p + \mathsf{h}) &= g(y + \Delta y)\\ &= Dg|_y \cdot (\Delta y) + \|\Delta y\| O_2(\Delta y)\\ &= Dg|_y \cdot (Df|_p \cdot \mathsf{h} + \|\mathsf{h}\| O_1(\mathsf{h})) + \|\mathsf{h}\| O_3(\mathsf{h})\\ &= Dg|_y \cdot Df|_p \cdot \mathsf{h} + \|\mathsf{h}\| Dg|_y \cdot O_1(\mathsf{h}) + \|\mathsf{h}\| O_3(\mathsf{h})\\ &= Dg|_y \cdot Df|_p \cdot \mathsf{h} + \|\mathsf{h}\| O_4(\mathsf{h})\end{aligned}$$

which implies that $g \circ f$ is differentiable at p with the derivative given by the promised formula.

Now we wish to show that $f, g \in C^r$ $r \geq 1$ implies that $g \circ f \in C^r$ also. The bilinear map defined by composition $comp : L(\mathsf{E}_1, \mathsf{E}_2) \times L(\mathsf{E}_2, \mathsf{E}_3) \to L(\mathsf{E}_1, \mathsf{E}_3)$ is bounded . Define a map

$$m_{f,g} : p \mapsto (Dg(f(p), Df(p))$$

which is defined on U_1. Consider the composition $comp \circ m_{f,g}$. Since f and g are at least C^1 this composite map is clearly continuous. Now we may proceed inductively. Consider the $r - th$ statement:

$$\text{composition of } C^r \text{ maps are } C^r$$

Suppose f and g are C^{r+1} then Df is C^r and $Dg \circ f$ is C^r by the inductive hypothesis so that $m_{f,g}$ is C^r. A bounded bilinear functional is C^∞. Thus *comp* is C^∞ and by examining $comp \circ m_{f,g}$ we see that the result follows. ∎

We will often use the following lemma without explicit mention when calculating:

Lemma G.1 *Let* $f : U \subset \mathsf{V} \to \mathsf{W}$ *be twice differentiable at* $x_0 \in U \subset \mathsf{V}$ *then the map* $D_v f : x \mapsto Df(x) \cdot v$ *is differentiable at* x_0 *and its derivative at* x_0 *is given by*

$$D(D_v f)|_{x_0} \cdot \mathsf{h} = D^2 f(x_0)(\mathsf{h}, v).$$

Proof. The map $D_v f : x \mapsto Df(x) \cdot v$ is decomposed as the composition

$$x \overset{Df}{\mapsto} Df|_x \overset{R^{\mathsf{v}}}{\mapsto} Df|_x \cdot v$$

where $R^{\mathsf{v}} : L(\mathsf{V}, \mathsf{W}) \mapsto \mathsf{W}$ is the map $(A, b) \mapsto A \cdot b$. The chain rule gives

$$\begin{aligned} D(D_{\mathsf{v}} f)(x_0) \cdot \mathsf{h} &= DR^{\mathsf{v}}(Df|_{x_0}) \cdot D(Df)|_{x_0} \cdot \mathsf{h}) \\ &= DR^{\mathsf{v}}(Df(x_0)) \cdot (D^2 f(x_0) \cdot \mathsf{h}). \end{aligned}$$

But R^{v} is linear and so $DR^{\mathsf{v}}(y) = R^v$ for all y. Thus

$$\begin{aligned} D(D_v f)|_{x_0} \cdot \mathsf{h} &= R^v(D^2 f(x_0) \cdot \mathsf{h}) \\ &= (D^2 f(x_0) \cdot \mathsf{h}) \cdot v = D^2 f(x_0)(\mathsf{h}, v). \end{aligned}$$

∎

Theorem G.2 *If* $f : U \subset \mathsf{V} \to \mathsf{W}$ *is twice differentiable on* U *such that* $D^2 f$ *is continuous, i.e. if* $f \in C^2(U)$ *then* $D^2 f$ *is symmetric:*

$$D^2 f(p)(w, v) = D^2 f(p)(v, w).$$

More generally, if $D^k f$ *exists and is continuous then* $D^k f\ (\mathbf{p}) \in \mathbf{L}^k_{sym}(\mathsf{V}; \mathsf{W})$.

Proof. Let $p \in U$ and define an affine map $A : \mathbb{R}^2 \to \mathsf{V}$ by $A(s,t) := p + sv + tw$. By the chain rule we have

$$\frac{\partial^2 (f \circ A)}{\partial s \partial t}(0) = D^2(f \circ A)(0) \cdot (\mathbf{e}_1, \mathbf{e}_2) = D^2 f(p) \cdot (v, w)$$

where $\mathbf{e}_1, \mathbf{e}_2$ is the standard basis of $\mathbb{R}^2$. Thus it suffices to prove that

$$\frac{\partial^2 (f \circ A)}{\partial s \partial t}(0) = \frac{\partial^2 (f \circ A)}{\partial t \partial s}(0).$$

In fact, for any $\ell \in \mathsf{V}^*$ we have

$$\frac{\partial^2 (\ell \circ f \circ A)}{\partial s \partial t}(0) = \ell \left(\frac{\partial^2 (f \circ A)}{\partial s \partial t} \right)(0)$$

and so by the Hahn-Banach theorem it suffices to prove that $\frac{\partial^2 (\ell \circ f \circ A)}{\partial s \partial t}(0) = \frac{\partial^2 (\ell \circ f \circ A)}{\partial t \partial s}(0)$ which is the standard 1-variable version of the theorem which we assume known. The result for $D^k f$ is proven by induction. ∎

Theorem G.3 *Let $\varrho \in L(\mathsf{F}_1, \mathsf{F}_2; \mathsf{W})$ be a bilinear map and let $f_1 : U \subset \mathsf{E} \to \mathsf{F}_1$ and $f_2 : U \subset \mathsf{E} \to \mathsf{F}_2$ be differentiable (resp. $C^r, r \geq 1$) maps. Then the composition $\varrho(f_1, f_2)$ is differentiable (resp. $C^r, r \geq 1$) on U where $\varrho(f_1, f_2) : x \mapsto \varrho(f_1(x), f_2(x))$. Furthermore,*

$$D\varrho|_x (f_1, f_2) \cdot v = \varrho(Df_1|_x \cdot v, f_2(x)) + \varrho(f_1(x), Df_2|_x \cdot v).$$

In particular, if F is an algebra with differentiable product $\star$ and $f_1 : U \subset \mathsf{E} \to \mathsf{F}$ and $f_2 : U \subset \mathsf{E} \to \mathsf{F}$ then $f_1 \star f_2$ is defined as a function and

$$D(f_1 \star f_2) \cdot v = (Df_1 \cdot v) \star (f_2) + (Df_1 \cdot v) \star (Df_2 \cdot v).$$

Proof. This is completely similar to the usual proof of the product rule and is left as an exercise. ■

The proof of this useful lemma is left as an easy exercise. It is actually quite often that this little lemma saves the day as it were.

It will be useful to define an integral for maps from an interval $[a, b]$ into a Banach space V. First we define the integral for step functions. A function f on an interval $[a, b]$ is a **step function** if there is a partition $a = t_0 < t_1 < \cdots < t_k = b$ such that f is constant, with value say f_i, on each subinterval $[t_i, t_{i+1})$. The set of step functions so defined is a vector space. We define the integral of a step function f over $[a, b]$ by

$$\int_{[a,b]} f := \sum_{i=0}^{k-1} (t_{i+1} - t_i) f_i = \sum_{i=0}^{k-1} (t_{i+1} - t_i) f(t_i).$$

One easily checks that the definition is independent of the partition chosen. Now the set of all step functions from $[a, b]$ into V is a linear subspace of the Banach space $\mathcal{B}(a, b, \mathsf{V})$ of all bounded functions of $[a, b]$ into V and the integral is a linear map on this space. Recall that the norm on $\mathcal{B}(a, b, \mathsf{V})$ is $\sup_{a \leq t < b}\{|f(t)|\}$. If we denote the closure of the space of step functions in this Banach space by $\bar{\mathcal{S}}(a, b, \mathsf{V})$ then we can extend the definition of the integral to $\bar{\mathcal{S}}(a, b, \mathsf{V})$ by continuity since on step functions we have

$$\left| \int_{[a,b]} f \right| \leq (b - a) \|f\|_\infty .$$

In the limit, this bound persists. This integral is called the **Cauchy-Bochner** integral and is a bounded linear map $\bar{\mathcal{S}}(a, b, \mathsf{V}) \to \mathsf{V}$. It is important to notice that $\bar{\mathcal{S}}(a, b, \mathsf{V})$ contains the continuous functions $C([a, b], \mathsf{V})$ because such may be uniformly approximated by elements of $\mathcal{S}(a, b, \mathsf{V})$ and so we can integrate these functions using the Cauchy-Bochner integral.

Lemma G.2 *If $\ell : \mathsf{V} \to \mathsf{W}$ is a bounded linear map of Banach spaces then for any $f \in \bar{\mathcal{S}}(a, b, \mathsf{V})$ we have*

$$\int_{[a,b]} \ell \circ f = \ell \circ \int_{[a,b]} f.$$

Proof. This is obvious for step functions. The general result follows by taking a limit for step functions converging in $\bar{\mathcal{S}}(a, b, \mathsf{V})$ to f. ■

Some facts about maps on finite dimensional spaces.

Let U be an open subset of $\mathbb{R}^n$ and let $f : U \to \mathbb{R}^m$ be a map which is differentiable at $a = (a^1, ..., a^n) \in \mathbb{R}^n$. The map f is given by m functions $f^i : U \to \mathbb{R}^m$, $1 \leq i \leq m$. Now with respect to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, the derivative is given by an $n \times m$ matrix called the **Jacobian** matrix:

$$J_a(f) := \begin{pmatrix} \frac{\partial f^1}{\partial x^1}(a) & \frac{\partial f^1}{\partial x^2}(a) & \cdots & \frac{\partial f^1}{\partial x^n}(a) \\ \frac{\partial f^2}{\partial x^1}(a) & & & \\ \vdots & & \ddots & \\ \frac{\partial f^m}{\partial x^1}(a) & & & \frac{\partial f^m}{\partial x^n}(a) \end{pmatrix}.$$

The rank of this matrix is called the rank of f at a. If $n = m$ so that $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ then the Jacobian is a square matrix and $\det(J_a(f))$ is called the **Jacobian determinant** at a. If f is differentiable near a then it follows from the inverse mapping theorem proved below that if $\det(J_a(f)) \neq 0$ then there is some open set containing a on which f has a differentiable inverse. The Jacobian of this inverse at $f(x)$ is the inverse of the Jacobian of f at x.

Notation G.5 *The Jacobian matrix was a bit tedious to write down. Of course we have the abbreviation but let us also use the suggestive notation*

$$\frac{\partial(f^1, .., f^m)}{\partial(x^1, .., x^n)}.$$

The following is the **mean value theorem**:

Theorem G.4 *Let* V *and* W *be Banach spaces. Let* $c : [a, b] \to \mathsf{V}$ *be a* C^1-*map with image contained in an open set* $U \subset \mathsf{V}$. *Also, let* $f : U \to \mathsf{W}$ *be a* C^1 *map. Then*

$$f(c(b)) - f(c(a)) = \int_0^1 Df(c(t)) \cdot c'(t)dt.$$

If $c(t) = (1 - t)x + ty$ *then*

$$f(y) - f(x) = \int_0^1 Df(c(t))dt \cdot (y - x)$$

Notice that $\int_0^1 Df(c(t))dt \in L(\mathsf{V}, \mathsf{W})$.

Proof. Use the chain rule and the 1-variable fundamental theorem of calculus for the first part. For the second use lemma 1.2. ∎

Corollary G.1 *Let* U *be a convex open set in a Banach space* V *and* $f : U \to \mathsf{W}$ *a* C^1 *map into another Banach space* W. *Then for any* $x, y \in U$ *we have*

$$\|f(y) - f(x)\| \leq C_{x,y} \|y - x\|$$

where $C_{x,y}$ *is the supremum over all values taken by* f *along the line segment which is the image of the path* $t \mapsto (1 - t)x + ty$.

Recall that for a fixed x higher derivatives $D^p f\big|_x$ are symmetric multilinear maps. For the following let $(y)^k$ denote $\underbrace{(y, y, ..., y)}_{k\text{-times}}$. With this notation we have the following version of Taylor's theorem.

Theorem G.5 (Taylor's theorem) *Given Banach spaces* V *and* W, *a* C^r *function* $f : U \to \mathsf{W}$ *and a line segment* $t \mapsto (1-t)x + ty$ *contained in* U, *we have that* $t \mapsto D^p f(x + ty) \cdot (y)^p$ *is defined and continuous for* $1 \le p \le k$ *and*

$$f(x+y) = f(x) + \frac{1}{1!}\, Df\big|_x \cdot y + \frac{1}{2!}\, D^2 f\big|_x \cdot (y)^2 + \cdots + \frac{1}{(k-1)!}\, D^{k-1} f\big|_x \cdot (y)^{k-1}$$
$$+ \int_0^1 \frac{(1-t)^{k-1}}{(k-1)!} D^k f(x+ty) \cdot (y)^k dt$$

Proof. The proof is by induction and follows the usual proof closely. The point is that we still have an integration by parts formula coming from the product rule and we still have the fundamental theorem of calculus. ■

G.0.12 Local theory of maps

Inverse Mapping Theorem

The main reason for restricting our calculus to Banach spaces is that the inverse mapping theorem holds for Banach spaces and there is no simple and general inverse mapping theory on more general topological vector spaces. The so called hard inverse mapping theorems such as that of Nash and Moser require special estimates and are constructed to apply to a very limited situation. Recently, Michor and Kriegl et. al. have promoted an approach which defines differentiability in terms of mappings of $\mathbb{R}$ into the space that makes a lot of the formal parts of calculus valid under their definition of differentiability. However, the general (and easy) inverse and implicit mapping theorems still remain limited as before to Banach spaces and more general cases have to be handled case by case.

Definition G.9 *Let* E *and* F *be Banach spaces. A map will be called a* C^r ***diffeomorphism near*** p *if there is some open set* $U \subset \operatorname{dom}(f)$ *containing* p *such that* $f\big|_U : U \to f(U)$ *is a* C^r *diffeomorphism onto an open set* $f(U)$. *The set of all maps which are diffeomorphisms near* p *will be denoted* $\operatorname{Diff}^r_p(\mathsf{E}, \mathsf{F})$. *If* f *is a* C^r *diffeomorphism near* p *for all* $p \in U = \operatorname{dom}(f)$ *then we say that* f *is a* ***local*** C^r ***diffeomorphism***.

Definition G.10 *Let* X, d_1 *and* Y, d_2 *be metric spaces. A map* $f : X \to Y$ *is said to be* ***Lipschitz continuous*** *(with constant* k*) if there is a* $k > 0$ *such that* $d(f(x_1), f(x_2)) \le kd(x_1, x_2)$ *for all* $x_1, x_2 \in X$. *If* $0 < k < 1$ *the map is called a* ***contraction mapping*** *(with constant* k*) or is said to be* k***-contractive***.

The following technical result has numerous applications and uses the idea of iterating a map. **Warning**: For this theorem f^n will denote the $n-$fold composition $f \circ f \circ \cdots \circ f$ rather than a product.

Proposition G.4 (Contraction Mapping Principle) *Let F be a closed subset of a complete metric space (M, d). Let $f : F \to F$ be a k-contractive map such that*

$$d(f(x), f(y)) \leq kd(x, y)$$

for some fixed $0 \leq k < 1$. Then

1) there is exactly one $x_0 \in F$ such that $f(x_0) = x_0$. Thus x_0 is a fixed point for f. Furthermore,

2) for any $y \in F$ the sequence $y_n := f^n(y)$ converges to the fixed point x_0 with the error estimate $d(y_n, x_0) \leq \frac{k^n}{1-k} d(y_1, x_0)$.

Proof. Let $y \in F$. By iteration

$$d(f^n(y), f^{n-1}(y)) \leq kd(f^{n-1}(y), f^{n-2}(y)) \leq \cdots \leq k^{n-1} d(f(y), y)$$

as follows:

$$\begin{aligned} d(f^{n+j+1}(y), f^n(y)) &\leq d(f^{n+j+1}(y), f^{n+j}(y)) + \cdots + d(f^{n+1}(y), f^n(y)) \\ &\leq (k^{j+1} + \cdots + k) d(f^n(y), f^{n-1}(y)) \\ &\leq \frac{k}{1-k} d(f^n(y), f^{n-1}(y)) \\ &\frac{k^n}{1-k} d(f^1(y), y)) \end{aligned}$$

¿From this, and the fact that $0 \leq k < 1$, one can conclude that the sequence $f^n(y) = x_n$ is Cauchy. Thus $f^n(y) \to x_0$ for some x_0 which is in F since F is closed. On the other hand,

$$x_0 = \lim_{n \to 0} f^n(y) = \lim_{n \to 0} f(f^{n-1}(y)) = f(x_0)$$

by continuity of f. Thus x_0 is a fixed point. If u_0 where also a fixed point then

$$d(x_0, u_0) = d(f(x_0), f(u_0)) \leq kd(x_0, u_0)$$

which forces $x_0 = u_0$. The error estimate in (2) of the statement of the theorem is left as an easy exercise. ∎

Remark G.2 *Note that a Lipschitz map f may not satisfy the hypotheses of the last theorem even if $k < 1$ since U is not a complete metric space unless $U = \mathsf{E}$.*

Definition G.11 *A continuous map $f : U \to \mathsf{E}$ such that $L_f := \mathrm{id}_U - f$ is injective has a not necessarily continuous inverse G_f and the invertible map $R_f := \mathrm{id}_{\mathsf{E}} - G_f$ will be called the* ***resolvent operator*** *for f.*

The resolvent is a term that is usually used in the context of linear maps and the definition in that context may vary slightly. Namely, what we have defined here would be the resolvent of $\pm L_f$. Be that as it may, we have the following useful result.

Theorem G.6 *Let E be a Banach space. If $f : \mathsf{E} \to \mathsf{E}$ is continuous map that is Lipschitz continuous with constant k where $0 \leq k < 1$, then the resolvent R_f exists and is Lipschitz continuous with constant $\frac{k}{1-k}$.*

Proof. Consider the equation $x - f(x) = y$. We claim that for any $y \in \mathsf{E}$ this equation has a unique solution. This follows because the map $F : \mathsf{E} \to \mathsf{E}$ defined by $F(x) = f(x) + y$ is k-contractive on the complete normed space E as a result of the hypotheses. Thus by the contraction mapping principle there is a unique x fixed by F which means a unique x such that $f(x) + y = x$. Thus the inverse G_f exists and is defined on all of E. Let $R_f := \mathrm{id}_{\mathsf{E}} - G_f$ and choose $y_1, y_2 \in \mathsf{E}$ and corresponding unique x_i , $i = 1, 2$ with $x_i - f(x_i) = y_i$. We have

$$\begin{aligned}\|R_f(y_1) - R_f(y_2)\| &= \|f(x_1) - f(x_2)\| \\ &\leq k\,\|x_1 - x_2\| \;\; \leq \\ &\leq k\,\|y_1 - R_f(y_1) - (y_2 - R_f(y_2))\| \leq \\ &\leq k\,\|y_1 - y_2\| + \|R_f(y_1) - R_f(y_2)\|\,.\end{aligned}$$

Solving this inequality we get

$$\|R_f(y_1) - R_f(y_2)\| \leq \frac{k}{1-k}\,\|y_1 - y_2\|\,.$$

■

Lemma G.3 *The space $GL(\mathsf{E}, \mathsf{F})$ of continuous linear isomorphisms is an open subset of the Banach space $L(\mathsf{E}, \mathsf{F})$. In particular, if $\|\mathrm{id} - A\| < 1$ for some $A \in GL(\mathsf{E})$ then $A^{-1} = \lim_{N\to\infty} \sum_{n=0}^{N} (\mathrm{id} - A)^n$.*

Proof. Let $A_0 \in GL(\mathsf{E}, \mathsf{F})$. The map $A \mapsto A_0^{-1} \circ A$ is continuous and maps $GL(\mathsf{E}, \mathsf{F})$ onto $GL(\mathsf{E}, \mathsf{F})$. If follows that we may assume that $\mathsf{E} = \mathsf{F}$ and that $A_0 = \mathrm{id}_{\mathsf{E}}$. Our task is to show that elements of $\mathrm{L}(\mathsf{E}, \mathsf{E})$ close enough to id_{E} are in fact elements of $GL(\mathsf{E})$. For this we show that

$$\|\mathrm{id} - A\| < 1$$

implies that $A \in GL(\mathsf{E})$. We use the fact that the norm on $\mathrm{L}(\mathsf{E}, \mathsf{E})$ is an algebra norm. Thus $\|A_1 \circ A_2\| \leq \|A_1\|\,\|A_2\|$ for all $A_1, A_2 \in \mathrm{L}(\mathsf{E}, \mathsf{E})$. We abbreviate id by "1" and denote $\mathrm{id} - A$ by Λ. Let $\Lambda^2 := \Lambda \circ \Lambda$, $\Lambda^3 := \Lambda \circ \Lambda \circ \Lambda$ and so forth. We now form a Neumann series :

$$\begin{aligned}\pi_0 &= 1 \\ \pi_1 &= 1 + \Lambda \\ \pi_2 &= 1 + \Lambda + \Lambda^2 \\ &\;\vdots \\ \pi_n &= 1 + \Lambda + \Lambda^2 + \cdots + \Lambda^n.\end{aligned}$$

By comparison with the Neumann series of real numbers formed in the same way using $\|A\|$ instead of A we see that $\{\pi_n\}$ is a Cauchy sequence since $\|\Lambda\| = \|\mathrm{id} - A\| < 1$. Thus $\{\pi_n\}$ is convergent to some element ρ. Now we have $(1-\Lambda)\pi_n = 1 - \Lambda^{n+1}$ and letting $n \to \infty$ we see that $(1-\Lambda)\rho = 1$ or in other words, $A\rho = 1$. ■

Lemma G.4 *The map* inv $:GL(\mathsf{E},\mathsf{F}) \to GL(\mathsf{E},\mathsf{F})$ *given by taking inverses is a* C^∞ *map and the derivative of* inv $:g \mapsto g^{-1}$ *at some* $g_0 \in GL(\mathsf{E},\mathsf{F})$ *is the linear map given by the formula:* $D\,\mathrm{inv}|_{g_0} : A \mapsto -g_0^{-1}Ag_0^{-1}$.

Proof. Suppose that we can show that the result is true for $g_0 = \mathrm{id}$. Then pick any $h_0 \in GL(\mathsf{E},\mathsf{F})$ and consider the isomorphisms $L_{h_0} : GL(\mathsf{E}) \to GL(\mathsf{E},\mathsf{F})$ and $R_{h_0^{-1}} : GL(\mathsf{E}) \to GL(\mathsf{E},\mathsf{F})$ given by $\phi \mapsto h_0\phi$ and $\phi \mapsto \phi h_0^{-1}$ respectively. The map $g \mapsto g^{-1}$ can be decomposed as

$$g \overset{L_{h_0^{-1}}}{\mapsto} h_0^{-1} \circ g \overset{\mathrm{inv}_{\mathsf{E}}}{\mapsto} (h_0^{-1} \circ g)^{-1} \overset{R_{h_0^{-1}}}{\mapsto} g^{-1}h_0h_0^{-1} = g^{-1}.$$

Now suppose that we have the result at $g_0 = \mathrm{id}$ in $GL(\mathsf{E})$. This means that $D\mathrm{inv}_{\mathsf{E}}|_{h_0} : A \mapsto -A$. Now by the chain rule we have

$$\begin{aligned}
(D\,\mathrm{inv}|_{h_0}) \cdot \mathrm{A} &= D(R_{h_0^{-1}} \circ \mathrm{inv}_{\mathsf{E}} \circ L_{h_0^{-1}}) \cdot \mathrm{A} \\
&= \left(R_{h_0^{-1}} \circ D\,\mathrm{inv}_{\mathsf{E}}|_{\mathrm{id}} \circ L_{h_0^{-1}}\right) \cdot \mathrm{A} \\
&= R_{h_0^{-1}} \circ (-\mathrm{A}) \circ L_{h_0^{-1}} = -h_0^{-1}\mathrm{A}h_0^{-1}
\end{aligned}$$

so the result is true for an arbitrary $h_0 \in GL(\mathsf{E},\mathsf{F})$. Thus we are reduced to showing that $D\mathrm{inv}_{\mathsf{E}}|_{\mathrm{id}} : A \mapsto -A$. The definition of derivative leads us to check that the following limit is zero.

$$\lim_{\|\mathrm{A}\| \to 0} \frac{\left\|(\mathrm{id} + \mathrm{A})^{-1} - (\mathrm{id})^{-1} - (-\mathrm{A})\right\|}{\|\mathrm{A}\|}.$$

Note that for small enough $\|\mathrm{A}\|$, the inverse $(\mathrm{id} + A)^{-1}$ exists and so the above limit makes sense. By our previous result (1.3) the above difference quotient becomes

$$\begin{aligned}
&\lim_{\|\mathrm{A}\| \to 0} \frac{\left\|(\mathrm{id} + \mathrm{A})^{-1} - \mathrm{id} + \mathrm{A}\right\|}{\|\mathrm{A}\|} \\
&= \lim_{\|\mathrm{A}\| \to 0} \frac{\|\sum_{n=0}^{\infty}(\mathrm{id} - (\mathrm{id} + \mathrm{A}))^n - \mathrm{id} + \mathrm{A}\|}{\|\mathrm{A}\|} \\
&= \lim_{\|\mathrm{A}\| \to 0} \frac{\|\sum_{n=0}^{\infty}(-\mathrm{A})^n - \mathrm{id} + \mathrm{A}\|}{\|\mathrm{A}\|} \\
&= \lim_{\|\mathrm{A}\| \to 0} \frac{\|\sum_{n=2}^{\infty}(-\mathrm{A})^n\|}{\|\mathrm{A}\|} \le \lim_{\|\mathrm{A}\| \to 0} \frac{\sum_{n=2}^{\infty}\|\mathrm{A}\|^n}{\|\mathrm{A}\|} \\
&= \lim_{\|\mathrm{A}\| \to 0} \sum_{n=1}^{\infty} \|\mathrm{A}\|^n = \lim_{\|\mathrm{A}\| \to 0} \frac{\|\mathrm{A}\|}{1 - \|\mathrm{A}\|} = 0.
\end{aligned}$$

■

Theorem G.7 (Inverse Mapping Theorem) *Let E and F be Banach spaces and $f : U \to \mathsf{F}$ be a C^r mapping defined an open set $U \subset \mathsf{E}$. Suppose that $x_0 \in U$ and that $f'(x_0) = Df|_x : \mathsf{E} \to \mathsf{F}$ is a continuous linear isomorphism. Then there exists an open set $V \subset U$ with $x_0 \in V$ such that $f : V \to f(V) \subset \mathsf{F}$ is a C^r-diffeomorphism. Furthermore the derivative of f^{-1} at y is given by $Df^{-1}\big|_y = (Df|_{f^{-1}(y)})^{-1}$.*

Proof. By considering $(Df|_x)^{-1} \circ f$ and by composing with translations we may as well just assume from the start that $f : \mathsf{E} \to \mathsf{E}$ with $x_0 = 0$, $f(0) = 0$ and $Df|_0 = \mathrm{id}_E$. Now if we let $g = x - f(x)$, then $Dg|_0 = 0$ and so if $r > 0$ is small enough then

$$\| Dg|_x \| < \frac{1}{2}$$

for $x \in B(0, 2r)$. The mean value theorem now tells us that $\|g(x_2) - g(x_1)\| \leq \frac{1}{2} \|x_2 - x_1\|$ for $x_2, x_1 \in \overline{B}(0, r)$ and that $g(\overline{B}(0, r)) \subset \overline{B}(0, r/2)$. Let $y_0 \in \overline{B}(0, r/2)$. It is not hard to show that the map $c : x \mapsto y_0 + x - f(x)$ is a contraction mapping $c : \overline{B}(0, r) \to \overline{B}(0, r)$ with constant $\frac{1}{2}$. The contraction mapping principle **??** says that c has a unique fixed point $x_0 \in \overline{B}(0, r)$. But $c(x_0) = x_0$ just translates to $y_0 + x_0 - f(x_0) = x_0$ and then $f(x_0) = y_0$. So x_0 is the unique element of $\overline{B}(0, r)$ satisfying this equation. But then since $y_0 \in \overline{B}(0, r/2)$ was an arbitrary element of $\overline{B}(0, r/2)$ it follows that the restriction $f : \overline{B}(0, r/2) \to f(\overline{B}(0, r/2))$ is invertible. But f^{-1} is also continuous since

$$\begin{aligned}
\left\| f^{-1}(y_2) - f^{-1}(y_1) \right\| &= \|x_2 - x_1\| \\
&\leq \|f(x_2) - f(x_1)\| + \|g(x_2) - g(x_1)\| \\
&\leq \|f(x_2) - f(x_1)\| + \frac{1}{2} \|x_2 - x_1\| \\
&= \|y_2 - y_1\| + \frac{1}{2} \left\| f^{-1}(y_2) - f^{-1}(y_1) \right\|
\end{aligned}$$

Thus $\left\| f^{-1}(y_2) - f^{-1}(y_1) \right\| \leq 2 \|y_2 - y_1\|$ and so f^{-1} is continuous. In fact, f^{-1} is also differentiable on $B(0, r/2)$. To see this let $f(x_2) = y_2$ and $f(x_1) = y_1$ with $x_2, x_1 \in \overline{B}(0, r)$ and $y_2, y_1 \in \overline{B}(0, r/2)$. The norm of $Df(x_1))^{-1}$ is bounded (by continuity) on $\overline{B}(0, r)$ by some number B. Setting $x_2 - x_1 = \Delta x$ and $y_2 - y_1 = \Delta y$ and using $(Df(x_1))^{-1} Df(x_1) = \mathrm{id}$ we have

$$\begin{aligned}
&\left\| f^{-1}(y_2) - f^{-1}(y_1) - (Df(x_1))^{-1} \cdot \Delta y \right\| \\
&= \left\| \Delta x - (Df(x_1))^{-1}(f(x_2) - f(x_1)) \right\| \\
&= \left\| \{(Df(x_1))^{-1} Df(x_1)\} \Delta x - \{(Df(x_1))^{-1} Df(x_1)\}(Df(x_1))^{-1}(f(x_2) - f(x_1)) \right\| \\
&\leq B \|Df(x_1)\Delta x - (f(x_2) - f(x_1))\| \leq o(\Delta x) = o(\Delta y) \text{ (by continuity).}
\end{aligned}$$

Thus $Df^{-1}(y_1)$ exists and is equal to $(Df(x_1))^{-1} = (Df(f^{-1}(y_1)))^{-1}$. A simple argument using this last equation shows that $Df^{-1}(y_1)$ depends continuously on y_1 and so f^{-1} is C^1. The fact that f^{-1} is actually C^r follows from a simple induction argument that uses the fact that Df is C^{r-1} together with lemma 1.4. This last step is left to the reader. ■

Exercise G.4 *Complete the last part of the proof of theorem*

Corollary G.2 *Let $U \subset \mathsf{E}$ be an open set and $0 \in U$. Suppose that $f : U \to \mathsf{F}$ is differentiable with $Df(p) : \mathsf{E} \to \mathsf{F}$ a (bounded) linear isomorphism for each $p \in U$. Then f is a local diffeomorphism.*

Theorem G.8 (Implicit Function Theorem I) *Let $\mathsf{E}_1, \mathsf{E}_2$ and F be Banach spaces and $U \times V \subset \mathsf{E}_1 \times \mathsf{E}_2$ open. Let $f : U \times V \to \mathsf{F}$ be a C^r mapping such that $f(x_0, y_0) = 0$. If $D_2 f_{(x_0,y_0)} : \mathsf{E}_2 \to \mathsf{F}$ is a continuous linear isomorphism then there exists a (possibly smaller) open set $U_0 \subset U$ with $x_0 \in U_0$ and unique mapping $g : U_0 \to V$ with $g(x_0) = y_0$ and such that*

$$f(x, g(x)) = 0$$

for all $x \in U_0$.

Proof. Follows from the following theorem. ■

Theorem G.9 (Implicit Function Theorem II) *Let $\mathsf{E}_1, \mathsf{E}_2$ and F be Banach spaces and $U \times V \subset \mathsf{E}_1 \times \mathsf{E}_2$ open. Let $f : U \times V \to \mathsf{F}$ be a C^r mapping such that $f(x_0, y_0) = w_0$. If $D_2 f(x_0, y_0) : \mathsf{E}_2 \to \mathsf{F}$ is a continuous linear isomorphism then there exists (possibly smaller) open sets $U_0 \subset U$ and $W_0 \subset \mathsf{F}$ with $x_0 \in U_0$ and $w_0 \in W_0$ together with a unique mapping $g : U_0 \times W_0 \to V$ such that*

$$f(x, g(x, w)) = w$$

for all $x \in U_0$. Here unique means that any other such function h defined on a neighborhood $U_0' \times W_0'$ will equal g on some neighborhood of (x_0, w_0).

Proof. Sketch: Let $\Psi : U \times V \to \mathsf{E}_1 \times \mathsf{F}$ be defined by $\Psi(x, y) = (x, f(x, y))$. Then $D\Psi(x_0, y_0)$ has the operator matrix

$$\begin{bmatrix} \mathrm{id}_{E_1} & 0 \\ D_1 f(x_0, y_0) & D_2 f(x_0, y_0) \end{bmatrix}$$

which shows that $D\Psi(x_0, y_0)$ is an isomorphism. Thus Ψ has a unique local inverse Ψ^{-1} which we may take to be defined on a product set $U_0 \times W_0$. Now Ψ^{-1} must have the form $(x, y) \mapsto (x, g(x, y))$ which means that $(x, f(x, g(x, w))) = \Psi(x, g(x, w)) = (x, w)$. Thus $f(x, g(x, w)) = w$. The fact that g is unique follows from the local uniqueness of the inverse Ψ^{-1} and is left as an exercise. ■

In the case of a map $f : U \to V$ between open subsets of Euclidean spaces (say $\mathbb{R}^n$ and $\mathbb{R}^m$) we have the notion of **rank at** $p \in U$ which is just the rank of the linear map $D_p f : \mathbb{R}^n \to \mathbb{R}^m$.

Definition G.12 *Let X, Y be topological spaces. When we write $f :: \mathsf{X} \to \mathsf{Y}$ we imply only that f is defined on some open set in X. We shall employ a similar use of the symbol "$::$" when talking about continuous maps between (open subsets* *of) topological spaces in general. If we wish to indicate that f is defined near $p \in \mathsf{X}$ and that $f(p) = q$ we will used the pointed category notation together with the symbol "$::$":*

$$f :: (\mathsf{X}, p) \to (\mathsf{Y}, q)$$

We will refer to such maps as ***local maps at*** *p. Local maps may be composed with the understanding that the domain of the composite map may become smaller: If $f :: (\mathsf{X}, p) \to (\mathsf{Y}, q)$ and $g :: (\mathsf{Y}, q) \to (\mathsf{G}, z)$ then $g \circ f :: (\mathsf{X}, p) \to (\mathsf{G}, z)$ and the domain of $g \circ f$ will be a non-empty open set. Also, we will say that two such maps $f_1 : (\mathsf{X}, p) \to (\mathsf{Y}, q)$ and $f_2 : (\mathsf{X}, p) \to (\mathsf{Y}, q)$ are* ***equal near*** *p if there is an open set O with $p \in O \subset \mathrm{dom}(f_1) \cap \mathrm{dom}(f_2)$ such that the restrictions to O are equal:*

$$f_1|_O = f_2|_O$$

in this case will simply write "$f_1 = f_2$ (near p)".

Notation G.6 *Recall that for a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ which is injective with rank r there exist linear isomorphisms $C_1 : \mathbb{R}^m \to \mathbb{R}^m$ and $C_2 : \mathbb{R}^n \to \mathbb{R}^n$ such that $C_1 \circ A \circ C_2^{-1}$ is just a projection followed by an injection:*

$$\mathbb{R}^n = \mathbb{R}^r \times \mathbb{R}^{n-r} \to \mathbb{R}^r \times 0 \to \mathbb{R}^r \times \mathbb{R}^{m-r} = \mathbb{R}^m$$

We have obvious special cases when $r = n$ or $r = m$. This fact has a local version that applies to C^∞ nonlinear maps. In order to facilitate the presentation of the following theorems we will introduce the following terminology:

Linear case.

Definition G.13 *We say that a continuous linear map $A_1 : \mathsf{E}_1 \to \mathsf{F}_1$ is* ***equivalent*** *to a map $A_2 : \mathsf{E}_2 \to \mathsf{F}_2$ if there are continuous linear isomorphisms $\alpha : \mathsf{E}_1 \to \mathsf{E}_2$ and $\beta : \mathsf{F}_1 \to \mathsf{F}_2$ such that $A_2 = \beta \circ A_1 \circ \alpha^{-1}$.*

Definition G.14 *Let $A : \mathsf{E} \to \mathsf{F}$ be an injective continuous linear map. We say that A is a* ***splitting injection*** *if there are Banach spaces F_1 and F_2 with $\mathsf{F} \cong \mathsf{F}_1 \times \mathsf{F}_2$ and if A is equivalent to the injection $\mathrm{inj}_1 : \mathsf{F}_1 \to \mathsf{F}_1 \times \mathsf{F}_2$.*

Lemma G.5 *If $A : \mathsf{E} \to \mathsf{F}$ is a splitting injection as above then there exists a linear isomorphism $\delta : \mathsf{F} \to \mathsf{E} \times \mathsf{F}_2$ such that $\delta \circ A : \mathsf{E} \to \mathsf{E} \times \mathsf{F}_2$ is the injection $x \mapsto (x, 0)$.*

Proof. By definition there are isomorphisms $\alpha : \mathsf{E} \to \mathsf{F}_1$ and $\beta : \mathsf{F} \to \mathsf{F}_1 \times \mathsf{F}_2$ such that $\beta \circ A \circ \alpha^{-1}$ is the injection $\mathsf{F}_1 \to \mathsf{F}_1 \times \mathsf{F}_2$. Since α is an isomorphism we may compose as follows

$$\begin{aligned}
&(\alpha^{-1} \times \mathrm{id}_{\mathsf{E}}) \circ \beta \circ A \circ \alpha^{-1} \circ \alpha \\
&= (\alpha^{-1} \times \mathrm{id}_{\mathsf{E}}) \circ \beta \circ A \\
&= \delta \circ A
\end{aligned}$$

to get a map which is easily seen to have the correct form. ■

If A is a splitting injection as above it easy to see that there are closed subspaces F_1 and F_2 of F such that $\mathsf{F} = \mathsf{F}_1 \oplus \mathsf{F}_2$ and such that A maps E isomorphically onto F_1 .

Definition G.15 *Let $A : \mathsf{E} \to \mathsf{F}$ be an surjective continuous linear map. We say that A is a* ***splitting surjection*** *if there are Banach spaces E_1 and E_2 with $\mathsf{E} \cong \mathsf{E}_1 \times \mathsf{E}_2$ and if A is equivalent to the projection $pr_1 : \mathsf{E}_1 \times \mathsf{E}_2 \to \mathsf{E}_1$.*

Lemma G.6 *If Let $A : \mathsf{E} \to \mathsf{F}$ is a splitting surjection then there is a linear isomorphism $\delta : \mathsf{F} \times \mathsf{E}_2 \to \mathsf{E}$ such that $A \circ \delta : \mathsf{F} \times \mathsf{E}_2 \to \mathsf{F}$ is the projection $(x, y) \mapsto x$.*

Proof. By definition there exist isomorphisms $\alpha : \mathsf{E} \to \mathsf{E}_1 \times \mathsf{E}_2$ and $\beta : \mathsf{F} \to \mathsf{E}_1$ such that $\beta \circ A \circ \alpha^{-1}$ is the projection $pr_1 : \mathsf{E}_1 \times \mathsf{E}_2 \to \mathsf{E}_1$. We form another map by composition by isomorphisms;

$$\begin{aligned} &\beta^{-1} \circ \beta \circ A \circ \alpha^{-1} \circ (\beta, \mathrm{id}_{\mathsf{E}_2}) \\ &= A \circ \alpha^{-1} \circ (\beta, \mathrm{id}_{\mathsf{E}_2}) := A \circ \delta \end{aligned}$$

and check that this does the job. ■

If A is a splitting surjection as above it easy to see that there are closed subspaces E_1 and E_2 of E such that $\mathsf{E} = \mathsf{E}_1 \oplus \mathsf{E}_2$ and such that A maps E onto E_1 as a projection $x + y \mapsto x$.

Local (nonlinear) case.

Definition G.16 *Let $f_1 : (\mathsf{E}_1, p_1) \to (\mathsf{F}_1, q_1)$ be a local map. We say that f_1 is* ***locally equivalent*** *near $p_1$to $f_2 : (\mathsf{E}_2, p_2) \to (\mathsf{F}_2, q_2)$ if there exist local diffeomorphisms $\alpha : \mathsf{E}_1 \to \mathsf{E}_2$ and $\beta : \mathsf{F}_1 \to \mathsf{F}_2$ such that $f_1 = \alpha \circ f_2 \circ \beta^{-1}$ (near p) or equivalently if $f_2 = \beta \circ f_1 \circ \alpha^{-1}$ (near p_2).*

Definition G.17 *Let $f :: \mathsf{E}, p \to \mathsf{F}, q$ be a local map. We say that f is a* ***locally splitting injection*** *or* ***local immersion*** *if there are Banach spaces F_1 and F_2 with $\mathsf{F} \cong \mathsf{F}_1 \times \mathsf{F}_2$ and if f is locally equivalent near p to the injection $\mathrm{inj}_1 :: (\mathsf{F}_1, 0) \to (\mathsf{F}_1 \times \mathsf{F}_2, 0)$.*

By restricting the maps to possibly smaller open sets we can arrange that $\beta \circ f \circ \alpha^{-1}$ is given by $U' \overset{\alpha^{-1}}{\to} U \overset{f}{\to} V \overset{\beta}{\to} U' \times V'$ which we will call a **nice** local injection.

Lemma G.7 *If f is a locally splitting injection as above there is an open set U_1 containing p and local diffeomorphism $\varphi : U_1 \subset \mathsf{F} \to U_2 \subset \mathsf{E} \times \mathsf{F}_2$ and such that $\varphi \circ f(x) = (x, 0)$ for all $x \in U_1$.*

Proof. This is done using the same idea as in the proof of lemma 1.5.

$$\begin{array}{cccc} & (\mathsf{E},p) & \to & (\mathsf{F},q) \\ \alpha & \updownarrow & & \updownarrow \beta \\ & (\mathsf{F}_1,0) & \to & (\mathsf{F}_1\times\mathsf{F}_2, \quad (0,0)) \\ \alpha & \updownarrow & \mathrm{inj}_1 & \updownarrow \alpha^{-1}\times\mathrm{id} \\ & (\mathsf{E},p) & \to & (\mathsf{E}\times\mathsf{F}_2, \quad (p,0)) \end{array}$$

■

Definition G.18 *Let $f :: (\mathsf{E},p) \to (\mathsf{F},q)$ be a local map. We say that f is a* ***locally splitting surjection*** *or* ***local submersion*** *if there are Banach spaces E_1 and E_2 with $\mathsf{E} \cong \mathsf{E}_1\times\mathsf{E}_2$ and if f is locally equivalent (at p) to the projection $pr_1 : \mathsf{E}_1\times\mathsf{E}_2 \to \mathsf{E}_1$.*

Again, by restriction of the domains to smaller open sets we can arrange that projection $\beta\circ f\circ\alpha^{-1}$ is given by $U'\times V' \overset{\alpha^{-1}}{\to} U \overset{f}{\to} V \overset{\beta}{\to} U'$ which we will call a **nice** local projection.

Lemma G.8 *If f is a locally splitting surjection as above there are open sets $U_1\times U_2 \subset \mathsf{F}\times\mathsf{E}_2$ and $V\subset\mathsf{F}$ together with a local diffeomorphism $\varphi : U_1\times U_2 \subset \mathsf{F}\times\mathsf{E}_2 \to V\subset\mathsf{E}$ such that $f\circ\varphi(u,v) = u$ for all $(u,v)\in U_1\times U_2$.*

Proof. *This is the local (nonlinear) version of lemma 1.6 and is proved just as easily. Examine the diagram for guidance if you get lost:*

$$\begin{array}{ccc} (\mathsf{E},p) & \to & (\mathsf{F},q) \\ \updownarrow & & \updownarrow \\ (\mathsf{E}_1\times\mathsf{E}_2,(0,0)) & \to & (\mathsf{E}_1,0) \\ \updownarrow & pr_1 & \updownarrow \\ (\mathsf{F}\times\mathsf{E}_2,(q,0)) & \to & (\mathsf{F},q) \end{array}$$

■

Theorem G.10 (local immersion) *Let $f :: (\mathsf{E},p)\to(\mathsf{F},q)$ be a local map. If $Df|_p : (\mathsf{E},p)\to(\mathsf{F},q)$ is a splitting injection then $f :: (\mathsf{E},p)\to(\mathsf{F},q)$ is a local immersion.*

Theorem G.11 (local immersion- finite dimensional case) *Let $f :: \mathbb{R}^n \to \mathbb{R}^n\times\mathbb{R}^k)$ be a map of constant rank n in some neighborhood of $0\in\mathbb{R}^n$. Then there is $g_1 :: \mathbb{R}^n \to \mathbb{R}^n$ with $g_1(0) = 0$, and a $g_2 :: \mathbb{R}^n \to \mathbb{R}^n\times\mathbb{R}^k$ such that $g_2\circ f\circ g_1^{-1}$ is just given by $x\mapsto(x,0)\in\mathbb{R}^n\times\mathbb{R}^k$.*

We have a similar but complementary theorem which we state in a slightly more informal manner.

Theorem G.12 (local submersion) *Let $f :: (\mathsf{E},p) \to (\mathsf{F},q)$ be a local map. If $Df|_p : (\mathsf{E},p) \to (\mathsf{F},q)$ is a splitting surjection then $f :: (\mathsf{E},p) \to (\mathsf{F},q)$ is a local submersion.*

Theorem G.13 (local submersion -finite dimensional case) *Let* $f :: (\mathbb{R}^n \times \mathbb{R}^k, 0) \to (\mathbb{R}^n, 0)$ *be a local map with constant rank* n *near* 0*. Then there are diffeomorphisms* $g_1 :: (\mathbb{R}^n \times \mathbb{R}^k, 0) \to (\mathbb{R}^n \times \mathbb{R}^k, 0)$ *and* $g_2 :: (\mathbb{R}^n, 0) \to (\mathbb{R}^n, 0)$ *such that near* 0 *the map*

$$g_2 \circ f \circ g_1^{-1} :: (\mathbb{R}^n \times \mathbb{R}^k, 0) \to (\mathbb{R}^n, 0)$$

is just the projection $(x, y) \mapsto x$*.*

If the reader thinks about what is meant by local immersion and local submersion he/she will realize that in each case the derivative map Df_p has full rank. That is, the rank of the Jacobian matrix in either case is a big as the dimensions of the spaces involved allow. Now rank is only a semicontinuous and this is what makes full rank extend from points out onto neighborhoods so to speak. On the other hand, we can get more general maps into the picture if we explicitly assume that the rank is locally constant. We will state and prove the following theorem only for the finite dimensional case. There is a Banach version of this theorem but

Theorem G.14 (The Rank Theorem) *Let* $f : (\mathbb{R}^n, p) \to (\mathbb{R}^m, q)$ *be a local map such that* Df *has constant rank* r *in an open set containing* p*. Then there are local diffeomorphisms* $g_1 : (\mathbb{R}^n, p) \to (\mathbb{R}^n, 0)$ *and* $g_2 : (\mathbb{R}^m, q) \to (\mathbb{R}^m, 0)$ *such that* $g_2 \circ f \circ g_1^{-1}$ *is a local diffeomorphism near* 0 *with the form*

$$(x^1, ...x^n) \mapsto (x^1, ...x^r, 0, ..., 0).$$

Proof. Without loss of generality we may assume that $f : (\mathbb{R}^n, 0) \to (\mathbb{R}^m, 0)$ and that (reindexing) the $r \times r$ matrix

$$\left(\frac{\partial f^j}{\partial x^j} \right)_{1 \le i,j \le r}$$

is nonsingular in an open ball centered at the origin of $\mathbb{R}^n$. Now form a map $g_1(x^1, ...x^n) = (f^1(x), ..., f^r(x), x^{r+1}, ..., x^n)$. The Jacobian matrix of g_1 has the block matrix form

$$\begin{bmatrix} \left(\frac{\partial f^i}{\partial x^j} \right) & \\ 0 & I_{n-r} \end{bmatrix}$$

which has nonzero determinant at 0 and so by the inverse mapping theorem g_1 must be a local diffeomorphism near 0. Restrict the domain of g_1 to this possibly smaller open set. It is not hard to see that the map $f \circ g_1^{-1}$ is of the form $(z^1, ..., z^n) \mapsto (z^1, ..., z^r, \gamma^{r+1}(z), ..., \gamma^m(z))$ and so has Jacobian matrix of the form

$$\begin{bmatrix} I_r & 0 \\ * & \left(\frac{\partial \gamma^i}{\partial x^j} \right) \end{bmatrix}.$$

Now the rank of $\left(\frac{\partial \gamma^i}{\partial x^j} \right)_{r+1 \le i \le m,\ r+1 \le j \le n}$ must be zero near 0 since the $\text{rank}(f) = \text{rank}(f \circ h^{-1}) = r$ near 0. On the said (possibly smaller) neighborhood we now define the map $g_2 : (\mathbb{R}^m, q) \to (\mathbb{R}^m, 0)$ by

$$(y^1, ..., y^m) \mapsto (y^1, ..., y^r, y^{r+1} - \gamma^{r+1}(y_*, 0), ..., y^m - \gamma^m(y_*, 0))$$

where $(y_*, 0) = (y^1, ..., y^r, 0, ..., 0)$. The Jacobian matrix of g_2 has the form

$$\begin{bmatrix} I_r & 0 \\ * & I \end{bmatrix}$$

and so is invertible and the composition $g_2 \circ f \circ g_1^{-1}$ has the form

$$\begin{aligned} z &\overset{f\circ g_1^{-1}}{\mapsto} (z_*, \gamma_{r+1}(z), ..., \gamma_m(z)) \\ &\overset{g_2}{\mapsto} (z_*, \gamma_{r+1}(z) - \gamma_{r+1}(z_*, 0), ..., \gamma_m(z) - \gamma_m(z_*, 0)) \end{aligned}$$

where $(z_*, 0) = (z^1, ..., z^r, 0, ..., 0)$. It is not difficult to check that $g_2 \circ f \circ g_1^{-1}$ has the required form near 0. ■

Remark G.3 *In this section we have defined several things in terms of the canonical projections onto, or injections into, Cartesian products. The fact that we projected onto the first factor and injected into the first factor is irrelevant and it could have been either factor. Thus we will freely use the theorem as if we had made any other choice of factor. This is the case in our definitions of submanifolds, immersions, and submersions.*

G.0.13 The Tangent Bundle of an Open Subset of a Banach Space

Later on we will define the notion of a tangent space and tangent bundle for a differentiable manifold which locally looks like a Banach space. Here we give a definition that applies to the case of an open set U in a Banach space.

Definition G.19 *Let* E *be a Banach space and* $U \subset \mathsf{E}$ *an open subset. A* ***tangent vector*** *at* $x \in U$ *is a pair* (x, v) *where* $v \in \mathsf{E}$*. The* ***tangent space*** *at* $x \in U$ *is defined to be* $T_xU := T_x\mathsf{E} := \{x\} \times \mathsf{E}$ *and the tangent bundle* TU *over* U *is the union of the tangent spaces and so is just* $TU = U \times \mathsf{E}$*. Similarly the cotangent bundle over* U *is defined to be* $T^*U = U \times \mathsf{E}^*$*. A tangent space* $T_x\mathsf{E}$ *is also sometimes called the fiber at* x*.*

We give this definition in anticipation of our study of the tangent space at a point of a differentiable manifold. In this case however, it is often not necessary to distinguish between T_xU and E since we can often tell from context that an element $v \in \mathsf{E}$ is to be interpreted as based at some point $x \in U$. For instance a vector field in this setting is just a map $X : U \to \mathsf{E}$ but where $X(x)$ should be thought of as based at x.

Definition G.20 *If* $f : U \to \mathsf{F}$ *is a* C^r *map into a Banach space* F *then the* ***tangent map*** $Tf : TU \to T\mathsf{F}$ *is defined by*

$$Tf \cdot (x, v) = (f(x), Df(x) \cdot v).$$

The map takes the tangent space $T_xU = T_x\mathsf{E}$ *linearly into the tangent space* $T_{f(x)}\mathsf{F}$ *for each* $x \in U$*. The projection onto the first factor is written* τ_U :

$TU = U \times \mathsf{E} \to U$ and given by $\tau_U(x,v) = x$. We also have a projection $\pi_U : T^*U = U \times \mathsf{E}^* \to U$ defined similarly.

If $f : U \to V$ is a diffeomorphism of open sets U and V in E and F respectively then Tf is a diffeomorphism that is linear on the fibers and such that we have a commutative diagram:

$$\begin{array}{llcl} TU = & U \times \mathsf{E} & \overset{Tf}{\to} & V \times \mathsf{F} = TV \\ pr_1 & \downarrow & & \downarrow pr_1 \\ & U & \underset{f}{\to} & V \end{array}$$

The pair is an example of what is called a **local bundle map**. In this context we will denote the projection map $TU = U \times \mathsf{E} \to U$ by τ_U.

The chain rule looks much better if we use the tangent map:

Theorem G.15 *Let U_1 and U_2 be open subsets of Banach spaces E_1 and E_2 respectively. Suppose we have differentiable (resp. $C^r, r \geq 1$) maps composing as*

$$U_1 \xrightarrow{f} U_2 \xrightarrow{g} \mathsf{E}_3$$

where E_3 is a third Banach space. Then the composition is $g \circ f$ differentiable (resp. $C^r, r \geq 1$) and $T(g \circ f) = Tg \circ Tf$

$$\begin{array}{ccccc} TU_1 & \overset{Tf}{\to} & TU_2 & \overset{Tg}{\to} & T\mathsf{E}_3 \\ \tau_{U_1} \downarrow & & \tau_{U_2} \downarrow & & \downarrow \tau_{E_3} \\ U_1 & \overset{f}{\to} & U_2 & \overset{g}{\to} & \mathsf{E}_3 \end{array}$$

Notation G.7 *(and convention) There are three ways to express the "differential/derivative" of a differentiable map $f : U \subset \mathsf{E} \to \mathsf{F}$.*

1. *The first is just $Df : \mathsf{E} \to \mathsf{F}$ or more precisely $Df|_x : \mathsf{E} \to \mathsf{F}$ for any point $x \in U$.*

2. *This one is new for us. It is common but not completely standard :*

   $$dF : TU \to \mathsf{F}$$

   This is just the map $(x,v) \to Df|_x v$. We will use this notation also in the setting of maps from manifolds into vector spaces where there is a canonical trivialization of the tangent bundle of the target manifold (all of these terms will be defined). The most overused symbol for various "differentials" is d. We will use this in connection with Lie group also.

3. *Lastly the tangent map $Tf : TU \to T\mathsf{F}$ which we defined above. This is the one that generalizes to manifolds without problems.*

   In the local setting that we are studying now these three all contain essentially the same information so the choice to use one over the other is merely aesthetic.

G.1 Localization and Smooth Model Spaces

G.2 Exterior Derivative

Let $\omega_U : U \to L^k_{alt}(\mathsf{M};\mathsf{M})$. In the following calculation we will identify $L^k_{alt}(\mathsf{M};\mathsf{M})$ with the $L(\wedge^k\mathsf{M},\mathsf{M})$. For $\xi_0,...,\xi_k$ maps $\xi_i : U \to \mathsf{M}$ we have

$$\begin{aligned}
&D\langle\omega_U,\xi_0,...,\xi_k\rangle(x)\cdot\xi_i \\
&= \left.\frac{d}{dt}\right|_0 \langle\omega_U(x+t\xi_i),\xi_0(x+t\xi_i)\wedge...\wedge\widehat{\xi_i}\wedge...\wedge\xi_k(x+t\xi_i)\rangle \\
&= \langle\omega_U(x),\left.\frac{d}{dt}\right|_0[\xi_0(x+t\xi_i)\wedge...\wedge\widehat{\xi_i}\wedge...\wedge\xi_k(x+t\xi_i)]\rangle \\
&+\langle\left.\frac{d}{dt}\right|_0\omega_U(x),\xi_0(x+t\xi_i)\wedge...\wedge\widehat{\xi_i}\wedge...\wedge\xi_k(x+t\xi_i)\rangle \\
&= \langle\omega_U(x),\sum_{j=0}^{i-1}(-1)^j\xi_j^{'}(x)\xi_i\wedge[\xi_0(x)\wedge...\wedge\widehat{\xi_j}\wedge...\wedge\widehat{\xi_i}\wedge...\wedge\xi_k(x)]\rangle \\
&+\langle\omega_U(x),\sum_{j=i+1}^{k}(-1)^{j-1}\xi_j^{'}(x)\xi_i\wedge[\xi_0(x)\wedge...\wedge\widehat{\xi_i}\wedge...\wedge\widehat{\xi_j}\wedge...\wedge\xi_k(x)]\rangle \\
&+\langle\omega_U'(x)\xi_i,\xi_0(x)\wedge...\wedge\widehat{\xi_i}\wedge...\wedge\xi_k(x)\rangle
\end{aligned}$$

Theorem G.16 *There is a unique graded (sheaf) map $d:\Omega_M\to\Omega_M$, called the **exterior derivative**, such that*

1) $d\circ d=0$

2) d is a graded derivation of degree one, that is

$$d(\alpha\wedge\beta)=(d\alpha)\wedge\beta+(-1)^k\alpha\wedge(d\beta) \tag{G.1}$$

for $\alpha\in\Omega^k_M$.

Furthermore, if $\omega\in\Omega^k(U)$ and $X_0,X_1,...,X_k\in\mathfrak{X}_M(U)$ then

$$\begin{aligned}
d\omega &= \sum_{0\le i\le k}(-1)^iX_i(\omega(X_0,...,\widehat{X_i},...,X_k)) \\
&+\sum_{0\le i<j\le k}(-1)^{i+j}\omega([X_i,X_j],X_0,...,\widehat{X_i},...,\widehat{X_j},...,X_k).
\end{aligned}$$

In particular, we have the following useful formula for $\omega\in\Omega^1_M$ and $X,Y\in\mathfrak{X}_M(U)$:

$$d\omega(X,Y)=X(\omega(Y))-Y\omega(X)-\omega([X,Y])$$

Proof. First we give a local definition in terms of coordinates and then show that the global formula **??** agree with the local formula. Let U, ψ be a local chart on M. We will first define the exterior derivative on the open set $V = \psi(U) \subset \mathsf{M}$. Let $\xi_0, ..., \xi_k$ be local vector fields. The local representation of a form ω is a map $\omega_U : V \to L^k_{skew}(\mathsf{M}; \mathsf{M})$ and so has at some $x \in V$ has a derivative $D\omega_U(x) \in L(\mathsf{M}, L^k_{skew}(\mathsf{M}; \mathsf{M}))$. We define

$$d\omega_U(x)(\xi_0, ..., \xi_k) := \sum_{i=0}^{k} (-1)^i (D\omega_U(x)\xi_i(x))(\xi_0(x), ..., \widehat{\xi_i}, ..., \xi_k(x))$$

where $D\omega_U(x)\xi_i(x) \in L^k_{skew}(\mathsf{M}; \mathsf{M})$. This certainly defines a differential form in $\Omega^{k+1}(U)$. Let us call the right hand side of this local formula LOC. We wish to show that the global formula in local coordinates reduces to this local formula. Let us denote the first and second term of the global when expressed in local coordinates $L1$ and $L2$. Using, our calculation G.2 we have

$$L1 = \sum_{i=0}^{k} (-1)^i \xi_i(\omega(\xi_0, ..., \widehat{\xi_i}, ..., \xi_k)) = \sum_{i=0}^{k} (-1)^i D(\omega(\xi_0, ..., \widehat{\xi_i}, ..., \xi_k))(x)\xi_i(x)$$

$$\begin{aligned}
=&\langle \omega_U(x), \sum_{i=0}^{k} (-1)^i \sum_{j=0}^{i-1} (-1)^j \xi_j^{'}(x)\xi_i \wedge [\xi_0(x) \wedge ... \wedge \widehat{\xi_j} \wedge ... \wedge \widehat{\xi_i} \wedge ... \wedge \xi_k(x)] \rangle \\
&+ \langle \omega_U(x), \sum_{i=0}^{k} (-1)^i \sum_{j=i+1}^{k} (-1)^{j-1} \xi_j^{'}(x)\xi_i \wedge [\xi_0(x) \wedge ... \wedge \widehat{\xi_i} \wedge ... \\
&\qquad ... \wedge \widehat{\xi_j} \wedge ... \wedge \xi_k(x)] \rangle \\
&+ \sum_{i=0}^{k} (-1)^i \langle \omega_U'(x)\xi_i, \xi_0(x) \wedge ... \wedge \widehat{\xi_i} \wedge ... \wedge \xi_k(x) \rangle \\
=& \langle \omega_U(x), \sum_{i=0}^{k} \sum_{i<j}^{k} (-1)^{i+j} (\xi_j^{'}(x)\xi_i - \xi_i^{'}(x)\xi_j) \wedge \xi_0(x) \wedge ... \\
&\qquad .... \wedge \widehat{\xi_j} \wedge ... \wedge \widehat{\xi_i} \wedge ... \wedge \xi_k(x) \rangle \\
=& LOC + L2.
\end{aligned}$$

So our global formula reduces to the local one when expressed in local coordinates. ■

Remark G.4 *d is a local operator and so commutes with restrictions to open sets. In other words, if U is an open subset of M and d_U denotes the analogous operator on the manifold U then $d_U \alpha\big|_U = (d\alpha)\big|_U$. This operator can thus be expressed locally. In order to save on notation we will use d to denote the exterior derivative on any manifold, forms of any degree and for the restrictions d_U for any open set. It is exactly because d is a natural operator that this will cause no harm.*

G.3 Topological Vector Spaces

Definition G.21 *A **topological vector space (TVS)** is a vector space* V *with a Hausdorff topology such that the addition and scalar multiplication operations are (jointly) continuous.*

Definition G.22 *A sequence (or net)* x_n *in a TVS is call a Cauchy sequence if and only if for every neighborhood* U *of* 0 *there is a number* N_U *such that* $x_l - x_k \in U$ *for all* $k, l \geq N_U$. *A TVS is called complete if every Cauchy sequence (or net) is convergent.*

A relatively nice situation is when V has a norm which induces the topology. Recall that a **norm** is a function $\|\| : v \mapsto \|v\| \in \mathbb{R}$ defined on V such that for all $v, w \in \mathsf{V}$ we have

1. $\|v\| \geq 0$ and $\|v\| = 0$ if and only if $v = 0$,
2. $\|v + w\| \leq \|v\| + \|w\|$,
3. $\|\alpha v\| = |\alpha| \, \|v\|$ for all $\alpha \in \mathbb{R}$.

In this case we have a metric on V given by $\mathrm{dist}(v, w) := \|v - w\|$.

Definition G.23 *A **seminorm** is a function* $\|.\| : v \mapsto \|v\| \in \mathbb{R}$ *such that 2) and 3) above hold but instead of 1) we require only that* $\|v\| \geq 0$.

Definition G.24 *A **locally convex topological vector space*** V *is a TVS such that it's topology is generated by a family of seminorms* $\{\|.\|_\alpha\}_\alpha$. *This means that we give* V *the weakest topology such that all* $\|.\|_\alpha$ *are continuous. Since we have taken a TVS to be Hausdorff we require that the family of seminorms is **sufficient** in the sense that for each* $x \in \mathsf{V}$ *we have* $\bigcap\{x : \|x\|_\alpha = 0\} = \{0\}$. *A locally convex topological vector space is sometimes called a locally convex space and so we abbreviate the latter to **LCS**.*

Example G.2 *Let* Ω *be an open set in* $\mathbb{R}^n$ *or any manifold. For each* $x \in \Omega$ *define a seminorm* ρ_x *on* $C(\Omega)$ *by* $\rho_x(f) = f(x)$. *This family of seminorms makes* $C(\Omega)$ *a topological vector space. In this topology convergence is pointwise convergence. Also,* $C(\Omega)$ *is not complete with this TVS structure.*

Definition G.25 *An LCS which is complete (every Cauchy sequence converges) and metrizable is called a **Frechet space**.*

Definition G.26 *A curve* $c : \mathbb{R} \to \mathsf{V}$ *where* V *is a LCS is differentiable if the limit*

$$\dot{c}(t) := \lim_{\epsilon \to 0} \frac{c(t + \epsilon) - c(t)}{\epsilon}$$

exists for all t *thus defining a new curve* $\frac{dc}{dt} : \mathbb{R} \to \mathsf{V}$. *This curve may also have a derivative and so on. It all the iterated derivatives* $\frac{d^k c}{dt^k}$ *exist then we say the curve is **smooth**.*

Definition G.27 *Map* $f : \mathsf{V} \rightarrow \mathsf{W}$ *between locally convex spaces is called smooth if* $f \circ c$ *is smooth for every smooth curve* $c : \mathbb{R} \rightarrow \mathsf{V}$.

Notice that in this general setting we have so far only defined smoothness for curves and not for maps between open subsets of a LCS. We will however, make a new very flexible definition of smoothness that is appropriate for the locally convex spaces which are not Banach spaces.

Appendix H

Existence and uniqueness for differential equations

Theorem H.1 *Let E be a Banach space and let $X : U \subset E \to E$ be a smooth map. Given any $x_0 \in U$ there is a smooth curve $c : (-\epsilon, \epsilon) \to U$ with $c(0) = x_0$ such that $c'(t) = X(c(t))$ for all $t \in (-\epsilon, \epsilon)$. If $c_1 : (-\epsilon_1, \epsilon_1) \to U$ is another such curve with $c_1(0) = x_0$ and $c_1'(t) = X(c(t))$ for all $t \in (-\epsilon_1, \epsilon_1)$ then $c = c_1$ on the intersection $(-\epsilon_1, \epsilon_1) \cap (-\epsilon, \epsilon)$. Furthermore, there is an open set V with $x_0 \in V \subset U$ and a smooth map $\Phi : V \times (-a, a) \to U$ such that $t \mapsto c_x(t) := \Phi(x, t)$ is a curve satisfying $c'(t) = X(c(t))$ for all $t \in (-a, a)$.*

H.0.1 Differential equations depending on a parameter.

Theorem H.2 *Let J be an open interval on the real line containing 0 and suppose that for some Banach spaces E and F we have a smooth map $F : J \times U \times V \to \mathsf{F}$ where $U \subset \mathsf{E}$ and $V \subset \mathsf{F}$. Given any fixed point $(x_0, y_0) \in U \times V$ there exist a subinterval $J_0 \subset J$ containing 0 and open balls $B_1 \subset U$ and $B_2 \subset V$ with $(x_0, y_0) \in B_1 \times B_2$ and a unique smooth map*

$$\beta : J_0 \times B_1 \times B_2 \to V$$

such that

1) $\frac{d}{dt}\beta(t, x, y) = F(t, x, \beta(t, x, y))$ for all $(t, x, y) \in J_0 \times B_1 \times B_2$ and

2) $\beta(0, x, y) = y$.

Furthermore,

3) if we let $\beta(t, x) := \beta(t, x, y)$ for fixed y then

$$\begin{aligned}\frac{d}{dt} D_2\beta(t, x) \cdot v &= D_2F(t, x, \beta(t, x)) \cdot v \\ &\quad + D_3F(t, x, \beta(t, x)) \cdot D_2\beta(t, x) \cdot v\end{aligned}$$

for all $v \in \mathsf{E}$.

H.0.2 Smooth Banach Vector Bundles

The tangent bundle and cotangent bundle are examples of a general object called a (smooth) vector bundle which we have previously defined in the finite dimensional case. As a sort of review and also to introduce the ideas in the case of infinite dimensional manifolds we will define again the notion of a smooth vector bundle. For simplicity we will consider only C^∞ manifold and maps in this section. Let E be a Banach space. The most important case is when E is a finite dimensional vector space and in that case we might as well take $\mathsf{E} = \mathbb{R}^n$. It will be convenient to introduce the concept of a general fiber bundle and then specialize to vector bundles. The following definition is not the most efficient logically since there is some redundancy built in but is presented in this form for pedagogical reasons.

Definition H.1 *Let F be a smooth manifold modelled on F. A smooth* ***fiber bundle*** *$\xi = (E, \pi_E, M, F)$ with typical fiber F consists of*

1) smooth manifolds E and M referred to as the ***total space*** *and the* ***base space*** *respectively and modeled on Banach spaces $\mathsf{M} \times \mathsf{F}$ and M respectively;*

2) a smooth surjection $\pi_E : E \to M$ such that each ***fiber*** *$E_x = \pi^{-1}\{x\}$ is diffeomorphic to F;*

3) a cover of the ***base space*** *M by domains of maps $\phi_\alpha : E_{U_\alpha} := \pi_E^{-1}(U_\alpha) \to U_\alpha \times F$, called* ***bundle charts,*** *which are such that the following diagram commutes:*

$$\begin{array}{ccc} E_{U_\alpha} & \overset{(\pi_{U_\alpha}, \Phi_\alpha)}{\to} & U_\alpha \times F \\ \pi_E \quad \searrow & & \swarrow \\ & U_\alpha & \end{array} .$$

Thus each ϕ_α is of the form $(\pi_{U_\alpha}, \Phi_\alpha)$ where $\pi_{U_\alpha} := \pi_E | U_\alpha$ and $\Phi_\alpha : E_{U_\alpha} \to F$ is a smooth submersion.

Definition H.2 *The family of bundle charts whose domains cover the base space of a fiber bundle as in the above definition is called a* ***bundle atlas****.*

For all $x \in U_\alpha$, each restriction $\Phi_{\alpha,x} := \Phi_\alpha|_{E_x}$ is a diffeomorphism onto $\mathbf{F}$. Whenever we have two bundle charts $\phi_\alpha = (\pi_E, \Phi_\alpha)$ and $\phi_\beta = (\pi_E, \Phi_\beta)$ such that $U_\alpha \cap U_\beta \neq \emptyset$ then for every $x \in U_\alpha \cap U_\beta$ we have the diffeomorphism $\Phi_{\alpha\beta,x} = \Phi_{\alpha,x} \circ \Phi_{\beta,x}^{-1} : F \to F$. Thus we have map $g_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{Diff}(F)$ given by $g_{\alpha\beta}(x) := \Phi_{\alpha\beta,x}$. Notice that $g_{\beta\alpha} \circ g_{\alpha\beta}^{-1} = \mathrm{id}$. The maps so formed satisfy the following cocycle conditions:

$$g_{\gamma\beta} \circ g_{\alpha\gamma} = g_{\alpha\beta} \text{ whenever } U_\alpha \cap U_\beta \cap U_\gamma \neq \emptyset.$$

Let ξ be as above and let U be open in M. Suppose we have a smooth map $\phi : E_U \to U \times F$ such that the diagram

$$\begin{array}{ccc} E_U & \overset{\phi}{\to} & U \times F \\ \pi_U \quad \searrow & & \swarrow \\ & U & \end{array}$$

where $E_U := \pi_E^{-1}(U)$ as before. We call ϕ a trivialization and even if ϕ was not one of the bundle charts of a given bundle atlas, it must have the form (π_E, Φ) and we may enlarge the atlas by including this map. We do not wish to consider the new atlas as determining a new bundle so instead we say that the new atlas is equivalent. There is a unique maximal atlas for the bundle which just contains every possible trivialization.

Now given any open set U in the base space M of a fiber bundle $\xi = (E, \pi_E, M, F)$ we may form the restriction $\xi|U$ which is the fiber bundle $(\pi_E^{-1}(U), \pi_E|U, U, F)$. To simplify notation we write $E_U := \pi_E^{-1}(U)$ and $\pi_E|U := \pi_U$. This is a special case of the notion of a pull-back bundle.

One way in which vector bundles differ from general fiber bundles is with regard to the existence on global sections. A vector bundle always has at least one global section. Namely, the **zero section** $0_E : M \to E$ which is given by $x \mapsto 0_x \in E_x$. Our main interest at this point is the notion of a vector bundle. Before we proceed with our study of vector bundles we include one example of fiber bundle that is not a vector bundle.

Example H.1 *Let M and F be smooth manifolds and consider the projection map $pr_1 : M \times F \to M$. This is a smooth fiber bundle with typical fiber F and is called a* ***product bundle*** *or a* ***trivial bundle****.*

Example H.2 *Consider the tangent bundle $\tau_M : TM \to M$ of a smooth manifold modeled on $\mathbb{R}^n$. This is certainly a fiber bundle (in fact, a vector bundle) with typical fiber $\mathbb{R}^n$ but we also have the* ***bundle of nonzero vectors*** *$\pi : TM^\times \to M$ defined by letting $TM^\times := \{v \in TM : v \neq 0\}$ and $\pi := \tau_M|_{TM^\times}$. This bundle may have no global sections.*

Remark H.1 *A "structure" on a fiber bundle is determined by requiring that the atlas be paired down so that the transition maps all have values in some subset G (usually a subgroup) of $\mathrm{Diff}(F)$. Thus we speak of a $G-$**atlas** for $\xi = (E, \pi_E, M, F)$. In this case, a trivialization $(\pi_E, \Phi) : E_U \to U \times F$ is compatible with a given $G-$atlas $\mathcal{A}(G)$ if $\Phi_{\alpha,x} \circ \Phi_x^{-1} \in G$ and $\Phi_x \circ \Phi_{\alpha,x}^{-1} \in G$ for all $(\pi_E, \Phi_\alpha) \in \mathcal{A}(G)$. The set of all trivializations (bundle charts) compatible which a given $G-$atlas is a maximal $G-$atlas and is called a G**-structure.** Clearly, any $G-$atlas determines a G-structure. A fiber bundle $\xi = (E, \pi_E, M, F)$ with a G-atlas is called a G-bundle and any two G-atlases contained in the same maximal G-atlas are considered equivalent and determine the same G-bundle. We will study this idea in detail after we have introduced the notion of a Lie group.*

We now introduce our current object of interest.

Definition H.3 *A (real)* ***vector bundle*** *is a fiber bundle $\xi = (E, \pi_E, M, \mathsf{E})$ with typical fiber a (real) Banach space E such that for each pair of bundle chart domains U_α and U_β with nonempty intersection, the map*

$$g_{\alpha\beta} : x \mapsto \Phi_{\alpha\beta,x} := \Phi_{\alpha,x} \circ \Phi_{\beta,x}^{-1}$$

is a C^∞ morphism $g_{\alpha\beta} : U_\alpha \cap U_\beta \to GL(\mathsf{E})$. If E is finite dimensional, say $\mathsf{E} = \mathbb{R}^n$, then we say that $\xi = (E, \pi_E, M, \mathbb{R}^n)$ has rank n.

So if $v_x \in \pi_E^{-1}(x) \subset E_{U_\alpha}$ then $\phi_\alpha(v_x) = (x, \Phi_{\alpha,x}(v_x))$ for $\Phi_{\alpha,x} : E_x \to \mathsf{E}$ a diffeomorphism. Thus we can transfer the vector space structure of V to each fiber E_x in a well defined way since $\Phi_{\alpha,x} \circ \Phi_{\beta,x}^{-1} \in GL(\mathsf{E})$ for any x in the intersection of two VB-chart domains U_α and U_β. Notice that we have $\phi_\alpha \circ \phi_\beta^{-1}(x, v) = (x, g_{\alpha\beta}(x) \cdot v)$ where $g_{\alpha\beta} : U_\alpha \cap U_\beta \to GL(\mathsf{E})$ is differentiable and is given by $g_{\alpha\beta}(x) = \Phi_{\alpha,x} \circ \Phi_{\beta,x}^{-1}$. Notice that $g_{\alpha\beta}(x) \in GL(\mathsf{E})$.

A **complex vector bundle** is defined in an analogous way. For a complex vector bundle the typical fiber is a complex vector space (Banach space) and the transition maps have values in $GL(\mathsf{E}; \mathbb{C})$.

The set of all sections of real (resp. complex) vector bundle is a vector space over $\mathbb{R}$ (resp. $\mathbb{C}$) and a module over the ring of smooth real valued (resp. complex valued) functions.

Remark H.2 *If E is of finite dimension then the smoothness of the maps $g_{\alpha\beta} : U_\alpha \cap U_\beta \to GL(\mathsf{E})$ is automatic.*

Definition H.4 *The maps $g_{\alpha\beta}(x) := \Phi_{\alpha,x} \circ \Phi_{\beta,x}^{-1}$ are called **transition maps**.*

The transition maps always satisfy the following **cocycle condition**:

$$g_{\gamma\beta}(x) \circ g_{\beta\alpha}(x) = g_{\gamma\alpha}(x)$$

In fact, these maps encode the entire vector bundle up to isomorphism:

Remark H.3 *The following definition assumes the reader knows the definition of a Lie group and has a basic familiarity with Lie groups and Lie group homomorphisms. We shall study Lie groups in Chapter* **??**. *The reader may skip this definition.*

Definition H.5 *Let G be a Lie subgroup of $GL(\mathsf{E})$. We say that $\pi_E : E \to M$ has a **structure group** G if there is a cover by trivializations (vector bundle charts) $\phi_\alpha : E_{U_\alpha} \to U_\alpha \times \mathsf{E}$ such that for every non-empty intersection $U_\alpha \cap U_\beta$, the transition maps $g_{\alpha\beta}$ take values in G.*

Remark H.4 *Sometimes it is convenient to define the notion of vector bundle chart in a slightly different way. Notice that U_α is an open set in M and so ϕ_α is not quite a chart for the total space manifold E. But by choosing a possibly smaller open set inside U_α we may assume that U_α is the domain of an admissible chart U_α, ψ_α for M. Then we can compose to get a map $\widetilde{\phi}_\alpha : E_{U_\alpha} \to \psi_\alpha(U_\alpha) \times \mathsf{E}$. The maps now can serve as admissible charts for the differentiable manifold E. This leads to an alternative definition of VB-chart which fits better with what we did for the tangent bundle and cotangent bundle:*

Definition H.6 (Type II vector bundle charts) *A* ***(type II) vector bundle chart*** *on an open set $V \subset E$ is a fiber preserving diffeomorphism $\phi : V \to O \times \mathsf{E}$ which covers a diffeomorphism $\underline{\phi}{:}\pi_E(V) \to O$ in the sense that the following diagram commutes*

$$\begin{array}{cccc} & V & \overset{\phi}{\to} & O \times \mathsf{E} \\ \pi_E & \downarrow & & \downarrow pr_1 \\ & \pi_E(V) & \to & O \\ & & \underline{\phi} & \end{array}$$

and which is a linear isomorphism on each fiber.

Example H.3 *The maps $T\psi_\alpha : TU_\alpha \to \psi_\alpha(U_\alpha) \times \mathsf{E}$ are (type II) VB-charts and so not only give TM a differentiable structure but also provide TM with a vector bundle structure. Similar remarks apply for T^*M.*

Example H.4 *Let E be a vector space and let $E = M \times \mathsf{E}$. The using the projection $pr_1 : M \times \mathsf{E} \to M$ we obtain a vector bundle. A vector bundle of this simple form is called a trivial vector bundle.*

Define the sum of two section s_1 and s_2 by $(s_1 + s_2)(p) := s_1(p) + s_2(p)$. For any $f \in C^\infty(U)$ and $s \in \Gamma(U, E)$ define a section fs by $(fs)(p) = f(p)s(p)$. Under these obvious definitions $\Gamma(U, E)$ becomes a $C^\infty(U)$-module.

The the appropriate morphism in our current context is the vector bundle morphism:

Definition H.7 ***Definition H.8*** *Let (E, π_E, M) and $(F, \pi_F, N\ )$ be vector bundles. A* ***vector bundle morphism*** *$(E, \pi_E, M) \to (F, \pi_F, N)$ is a pair of maps $f : E \to F$ and $f_0 : M \to N$ such that*

1. ***Definition H.9*** *1) The following diagram commutes:*

$$\begin{array}{ccccc} & & f & & \\ & E & \to & F & \\ \pi_E & \downarrow & & \downarrow & \pi_F \\ & M & \to & N & \\ & & f_0 & & \end{array}$$

and $f|_{E_p}$ is a continuous linear map from E_p into $F_{f_0(p)}$ for each $p \in M$.

2) For each $x_0 \in M$ there exist VB-charts $(\pi_E, \Phi) : E_U \to U \times \mathsf{E}$ and $(\pi_E, \Phi'_\alpha) : F_{U'} \to U' \times \mathsf{E}'$ with $x_0 \in U$ and $f_0(U) \subset V$ such that

$$x \mapsto \Phi'|_{F_{f(x)}} \circ f_0 \circ \Phi|_{E_x}$$

is a smooth map from U into $GL(\mathsf{E}, \mathsf{E}')$.

Notation H.1 *Each of the following is a valid way to refer to a vector bundle morphism:*

1) $(f, f_0) : (E, \pi_E, M, \mathsf{E}) \to (F, \pi_F, N, \mathsf{F})$

2) $f : (E, \pi_E, M) \to (F, \pi_F, N)$ *(the map* f_0 *is induced and hence understood)*

3) $f : \xi_1 \to \xi_2$ *(this one is concise and fairly exact once it is set down that* $\xi_1 = (E_1, \pi_1, M)$ *and* $\xi_2 = (E_2, \pi_2, M)$*)*

4) $f : \pi_E \to \pi_F$

5) $E \xrightarrow{f} F$

Remark H.5 *There are many variations of these notations in use and the reader would do well to get used to this kind of variety. Actually, there will be much more serious notational difficulties in store for the novice. It has been said that notation is one of the most difficult aspects of differential geometry. On the other hand, once the underlying geometric picture has been properly understood , one may "see through" the notation. Drawing diagrams while interpreting equations is often a good idea.*

Definition H.10 ***Definition H.11*** *If* f *is an (linear) isomorphism on each fiber* E_p *then we say that* f *is a* ***vector bundle isomorphism*** *and the two bundles are considered equivalent.*

Notation H.2 *If* $\widetilde{f}$ *is a vector bundle morphism from a vector bundle* $\pi_E : E \to M$ *to a vector bundle* $\pi_F : F \to M$ *we will sometimes write this as* $\widetilde{f} : \pi_E \to \pi_F$ *or* $\pi_E \xrightarrow{\widetilde{f}} \pi_F$.

Definition H.12 *A vector bundle is called* ***trivial*** *if there is a there is a vector bundle isomorphism onto a trivial bundle:*

$$\begin{array}{ccc} E & \overset{\cong}{\to} & M \times \mathsf{E} \\ \pi_E \searrow & & \swarrow pr_1 \\ & M & \end{array}$$

Now we make the observation that a section of a trivial bundle is in a sense, nothing more than a vector-valued function since all sections $s \in \Gamma(M, M \times \mathsf{E})$ are of the form $p \to (p, f(p))$ for a unique function $f \in C^\infty(M, \mathsf{E})$. It is common to identify the function with the section.

Now there is an important observation to be made here; a trivial bundle always has a section which never takes on the value zero. There reason is that we may always take a trivialization $\phi : E \to M \times \mathsf{E}$ and then transfer the obvious nowhere-zero section $p \mapsto (p, 1)$ over to E. In other words, we let $s_1 : M \to E$ be defined by $s_1(p) = \phi^{-1}(p, 1)$. We now use this to exhibit a very simple example of a non-trivial vector bundle:

Example H.5 (Möbius bundle) *Let E be the quotient space of $[0,1]\times\mathbb{R}$ under the identification of $(0,t)$ with $(1,-t)$. The projection $[0,1]\times\mathbb{R}\to[0,1]$ becomes a map $E\to S^1$ after composition with the quotient map:*

$$\begin{array}{ccc} [0,1]\times\mathbb{R} & \to & [0,1] \\ \downarrow & & \downarrow \\ E & \to & S^1 \end{array}$$

Here the circle arises as $[0,1]/\sim$ where we have the induced equivalence relation given by taking $0\sim 1$ in $[0,1]$. The familiar Mobius band has an interior which is diffeomorphic to the Mobius bundle.

Now we ask if it is possible to have a nowhere vanishing section of E. It is easy to see that sections of E correspond to continuous functions $f:[0,1]\to\mathbb{R}$ such that $f(0)=-f(1)$. But then continuity forces such a function to take on the value zero which means that the corresponding section of E must vanish somewhere on $S^1=[0,1]/\sim$. Of course, examining a model of a Mobius band is even more convincing; any nonzero section of E could be, if such existed, normalized to give a map from S^1 to the boundary of a Möbius band which only went around once, so to speak, and inspection of a model would convince the reader that this is impossible.

Let $\xi_1=(E_1,\pi_1,M,\mathsf{E}_1)$ and $\xi_2=(E_2,\pi_2,M,\mathsf{E}_2)$ be vector bundles locally isomorphic to $\mathsf{M}\times\ \mathsf{E}_1$ and $\mathsf{M}\times\ \mathsf{E}_2$ respectively. We say that the sequence of vector bundle morphisms

$$0\to\xi_1\xrightarrow{f}\xi_2$$

is **exact** if the following conditions hold:

1. There is an open covering of M by open sets U_α together with trivializations $\phi_{1,\alpha}:\pi_1^{-1}(U_\alpha)\to U_\alpha\times\ \mathsf{E}_1$ and $\phi_{2,\alpha}:\pi_2^{-1}(U_\alpha)\to U_\alpha\times\ \mathsf{E}_2$ such that $\mathsf{E}_2=\mathsf{E}_1\times\mathsf{F}$ for some Banach space F;

2. the diagram below commutes for all α :

$$\begin{array}{ccccc} & \pi_1^{-1}(U_\alpha) & \to & \pi_2^{-1}(U_\alpha) & \\ \phi_{1,\alpha} & \downarrow & & \downarrow & \phi_{2,\alpha} \\ & U_\alpha\times\mathsf{E}_1 & \to & U_\alpha\times\mathsf{E}_1\times\mathsf{F} & \end{array}$$

Definition H.13 *A* ***subbundle*** *of a vector bundle $\xi=(E,\pi,M)$ is a vector bundle of the form $\xi=(L,\pi|_L,M)$ where $\pi|_L$ is the restriction to $L\subset E$, and where $L\subset E$ is a submanifold such that*

$$0\to\xi|_L\to\xi$$

is exact. Here, $\xi|_L\to\xi$ is the bundle map given by inclusion: $L\hookrightarrow E$.

Equivalently, $\pi|_L:L\to M$ is a ***subbundle*** *if $L\subset E$ is a submanifold and there is a splitting $\mathsf{E}=\mathsf{E}_1\times\mathsf{F}$ such that for each $p\in M$ there is a bundle chart $\phi:\pi^{-1}U\to U\times\mathsf{E}$ with $p\in U$ and $\phi((\pi^{-1}U)\cap L)=U\times\mathsf{E}_1\times\{0\}$.*

Definition H.14 *The chart ϕ from the last definition is said to be adapted to the subbundle.*

Notice that if $L \subset E$ is as in the previous definition then $\pi|_L : L \to M$ is a vector bundle with VB-atlas given by the various VB-charts U, ϕ restricted to $(\pi^{-1}U) \cap S$ and composed with projection $U \times \mathsf{E}_1 \times \{0\} \to U \times \mathsf{E}_1$ so $\pi|_L$ is a bundle locally isomorphic to $M \times \mathsf{E}_1$. The fiber of $\pi|_L$ at $p \in L$ is $L_p = E_p \cap L$. Once again we remind the reader of the somewhat unfortunate fact that although the bundle includes and is indeed determined by the map $\pi|_L : L \to M$ we often refer to L itself as the subbundle. In order to help the reader see what is going on here lets us look at how the definition of subbundle looks if we are in the finite dimensional case. We take $\mathsf{M} = \mathbb{R}^n$, $\mathsf{E} = \mathbb{R}^m$ and $\mathsf{E}_1 \times \mathsf{F}$ is the decomposition $\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^{m-k}$. Thus the bundle $\pi : E \to M$ has rank m (i.e. the typical fiber is $\mathbb{R}^m$) while the subbundle $\pi|_L : L \to M$ has rank k. The condition described in the definition of subbundle translates into there being a VB-chart $\phi : \pi^{-1}U \to U \times \mathbb{R}^k \times \mathbb{R}^{m-k}$ with $\phi((\pi^{-1}U) \cap L) = U \times \mathbb{R}^k \times \{0\}$. What if our original bundle was the trivial bundle $pr_1 : U \times \mathbb{R}^m \to U$? Then the our adapted chart must be a map $U \times \mathbb{R}^m \to U \times \mathbb{R}^k \times \mathbb{R}^{m-k}$ which must have the form $(x, v) \mapsto (x, f(x)v, 0)$ where for each x the $f(x)$ is a linear map $\mathbb{R}^m \to \mathbb{R}^k$.

H.0.3 Formulary

We now define the pseudogroup(s) relevant to the study of foliations. Let $\mathsf{M} = \mathsf{E} \times \mathsf{F}$ be a (split) Banach space. Define $\mathcal{G}_{\mathsf{M},\mathsf{F}}$ to be the set of all diffeomorphisms from open subsets of $\mathsf{E} \times \mathsf{F}$ to open subsets of $\mathsf{E} \times \mathsf{F}$ of the form

$$\Phi(x, y) = (f(x, y), g(y)).$$

In case M is n dimensional and $\mathsf{M} = \mathbb{R}^n$ is decomposed as $\mathbb{R}^k \times \mathbb{R}^q$ we write $\mathcal{G}_{\mathsf{M},\mathsf{F}} = \mathcal{G}_{n,q}$. We can then the following definition:

Definition H.15 *A $\mathcal{G}_{\mathsf{M},\mathsf{F}}$ structure on a manifold M modeled on $\mathsf{M} = \mathsf{E} \times \mathsf{F}$ is a maximal atlas of charts satisfying the condition that the overlap maps are all members of $\mathcal{G}_{\mathsf{M},\mathsf{F}}$.*

1) $\iota_{[X,Y]} = \mathcal{L}_X \circ \iota_Y + \iota_Y \circ \mathcal{L}_X$
2) $\mathcal{L}_{fX}\omega = f\mathcal{L}_X\omega + df \wedge \iota_X\omega$ for all $\omega \in \Omega(M)$
3) $d(\alpha \wedge \beta) = (d\alpha) \wedge \beta + (-1)^k \alpha \wedge (d\beta)$
4) $\frac{d}{dt}\mathrm{Fl}_t^{X*}Y = \mathrm{Fl}_t^{X*}(L_X Y)$
5) $[X, Y] = \sum_{i,j=1}^m \left(X^j \frac{\partial Y^i}{\partial x^j} - Y^j \frac{\partial X^i}{\partial x^j} \right) \frac{\partial}{\partial x^i}$

Example H.6 (Frame bundle) *Let M be a smooth manifold of dimension n. Let $F_x(M)$ denote the set of all bases (frames) for the vector space T_xM. Now let $F(M) := \bigcup_{x \in M} F_x(M)$. Define the natural projection $\pi : F(M) \to M$ by $\pi(\mathbf{f}) = x$ for all frames $\mathbf{f} = (f_i)$ for the space T_xM. It can be shown that $F(M)$ has a natural smooth structure. It is also a $GL(n, \mathbb{R})$-bundle whose typical fiber* *is also $GL(n,\mathbb{R})$. The bundle charts are built using the charts for M in the following way: Let U_α, ψ_α be a chart for M. Any frame $\mathbf{f} = (f_i)$ at some point $x \in U_\alpha$ may be written as*

$$f_i = \sum c_i^j \left. \frac{\partial}{\partial x^j} \right|_x .$$

We then map $\mathbf{f}$ to $(x, (c_i^j)) \in U_\alpha \times GL(n,\mathbb{R})$. This recipe gives a map $\pi^{-1}(U_\alpha) \to U_\alpha \times GL(n,\mathbb{R})$ which is a bundle chart.

Definition H.16 *A bundle morphism $(f, f_0) : \xi_1 \to \xi_2$ from one fiber bundle $\xi_1 = (E_1, \pi_{E_1}, M_1, F_1)$ to another $\xi_2 = (E_2, \pi_{E_2}, M_2, F_2)$ is a pair of maps (f, f_0) such that the following diagram commutates*

$$\begin{array}{ccc} E_1 & \overset{f}{\to} & E_2 \\ \pi_{E_1} \downarrow & & \pi_{E_2} \downarrow \\ M_1 & \overset{f_0}{\to} & M_2 \end{array}$$

In case $M_1 = M_2$ and $f_0 = \mathrm{id}_M$ we call f a strong bundle morphism. In the latter case if $f : E_1 \to E_2$ is also a diffeomorphism then we call it a ***bundle isomorphism.***

Definition H.17 *Let ξ_1 and ξ_2 be fiber bundles with the same base space M. If there exists a bundle isomorphism $(f, \mathrm{id}_M) : \xi_1 \to \xi_2$ we say that ξ_1 and ξ_2 are* ***isomorphic as fiber bundles over*** *M and write $\xi_1 \overset{fib}{\cong} \xi_2$.*

Now given any open set U in the base space M of a fiber bundle $\xi = (E, \pi_E, M, F)$ we may form the restriction $\xi|\,U$ which is the fiber bundle $(\pi_E^{-1}(U), \pi_E|\,U, U, F)$. To simplify notation we write $E_U := \pi_E^{-1}(U)$ and $\pi_E|\,U := \pi_U$.

Example H.7 *Let M and F be smooth manifolds and consider the projection map $pr_1 : M \times F \to M$. This is a smooth fiber bundle with typical fiber F and is called a* ***product bundle*** *or a* ***trivial bundle****.*

A fiber bundle which is isomorphic to a product bundle is also called a trivial bundle. The definition of a fiber bundle ξ with typical fiber F includes the existence of a cover of the base space by a family of open sets $\{U_\alpha\}_{\alpha \in A}$ such that $\xi|\,U_\alpha \overset{fib}{\cong} U \times F$ for all $\alpha \in A$. Thus, fiber bundles as we have defined them, are all locally trivial.

Misc

$1-$form $\theta = \sum e_j \theta^i$ which takes any vector to itself:

$$\begin{aligned} \theta(v_p) &= \sum e_j(p)\theta^i(v_p) \\ &= \sum v^i e_j(p) = v_p \end{aligned}$$

Let us write $d^{\nabla}\theta = \frac{1}{2}\sum e_k \otimes T^k_{ij}\theta^i \wedge \theta^j = \frac{1}{2}\sum e_k \otimes \tau^k$. If ∇ is the Levi-Civita connection on M then consider the projection $P^\wedge : E \otimes TM \otimes T^*M$ given by $P^\wedge T(\xi, v) = T(\xi, v) - T(v, \xi)$. We have

$$\begin{aligned}\nabla e_j &= \omega^k_j e_k = e\omega \\ \nabla\theta^j &= -\omega^j_k \theta^k\end{aligned}$$

$\nabla_\xi(e_j \otimes \theta^j)$
$P^\wedge(\nabla_\xi \theta^j)(v) = -\omega^j_k(\xi)\theta^k(v) + \omega^j_k(v)\theta^k(\xi) = -\omega^j_k \wedge \theta^k$
Let $T(\xi, v) = \nabla_\xi(e_i \otimes \theta^j)(v)$
$= (\nabla_\xi e_i) \otimes \theta^j(v) + e_i \otimes (\nabla_\xi \theta^j)(v) = \omega^k_i(\xi) e_k \otimes \theta^j(v) + e_i \otimes (-\omega^j_k(\xi)\theta^k(v))$
$= \omega^k_i(\xi) e_k \otimes \theta^j(v) + e_i \otimes (-\omega^k_j(\xi)\theta^j(v)) = e_k \otimes (\omega^k_i(\xi) - \omega^k_j(\xi))\theta^j(v)$
Then

$$\begin{aligned}(P^\wedge T)(\xi, v) &= T(\xi, v) - T(v, \xi) \\ &= (\nabla e_j) \wedge \theta^j + e_j \otimes d\theta^j \\ &= d^{\nabla}(e_j \otimes \theta^j)\end{aligned}$$

$$\begin{aligned}d^{\nabla}\theta &= d^{\nabla}\sum e_j\theta^j \\ &= \sum(\nabla e_j) \wedge \theta^j + \sum e_j \otimes d\theta^j \\ &= \sum(\sum_k e_k \otimes \omega^k_j) \wedge \theta^j + \sum e_k \otimes d\theta^k \\ &= \sum_k e_k \otimes (\sum_j \omega^k_j \wedge \theta^j + d\theta^k)\end{aligned} \tag{H.1}$$

So that $\sum_j \omega^k_j \wedge \theta^j + d\theta^k = \frac{1}{2}\tau^k$. Now let $\sigma = \sum f^j e_j$ be a vector field

$$\begin{aligned}d^{\nabla}d^{\nabla}\sigma = d^{\nabla}\left(d^{\nabla}\sum e_j f^j\right) &= d^{\nabla}\left(\sum(\nabla e_j)f^j + \sum e_j \otimes df^j\right) \\ &\left(\sum(\nabla e_j)df^j + \sum(d^{\nabla}\nabla e_j)f^j + \sum \nabla e_j df^j + \sum e_j \otimes ddf^j\right)\end{aligned}$$

$$\sum f^j(d^{\nabla}\nabla e_j) = \sum f^j$$

So we seem to have a map $f^j e_j \mapsto \Omega^k_j f^j e_k$.

$$\begin{aligned}e_r\Omega^r_j = d^{\nabla}\nabla e_j &= d^{\nabla}(e_k\omega^k_j) \\ &= \nabla e_k \wedge \omega^k_j + e_k d\omega^k_j \\ &= e_r\omega^r_k \wedge \omega^k_j + e_k d\omega^k_j \\ &= e_r\omega^r_k \wedge \omega^k_j + e_r d\omega^r_j \\ &= e_r(d\omega^r_j + \omega^r_k \wedge \omega^k_j)\end{aligned}$$

$$d^{\nabla}\nabla e = d^{\nabla}(e\omega) = \nabla e \wedge \omega + ed\omega$$

¿From this we get $0 = d(A^{-1}A)A^{-1} = (dA^{-1})AA^{-1} + A^{-1}dAA^{-1}$
$dA^{-1} = A^{-1}dAA^{-1}$

$$\begin{aligned}
\Omega_j^r &= d\omega_j^r + \omega_k^r \wedge \omega_j^k \\
\Omega &= d\omega + \omega \wedge \omega \\
\Omega' &= d\omega' + \omega' \wedge \omega' \\
\Omega' &= d\left(A^{-1}\omega A + A^{-1}dA\right) + \left(A^{-1}\omega A + A^{-1}dA\right) \wedge \left(A^{-1}\omega A + A^{-1}dA\right) \\
&= d\left(A^{-1}\omega A\right) + d\left(A^{-1}dA\right) + A^{-1}\omega \wedge \omega A + A^{-1}\omega \wedge dA \\
&+ A^{-1}dAA^{-1} \wedge \omega A + A^{-1}dA \wedge A^{-1}dA \\
&= d\left(A^{-1}\omega A\right) + dA^{-1} \wedge dA + A^{-1}\omega \wedge \omega A + A^{-1}\omega \wedge dA \\
&+ A^{-1}dAA^{-1}\omega A + A^{-1}dA \wedge A^{-1}dA \\
&= dA^{-1}\omega A + A^{-1}d\omega A - A^{-1}\omega dA + dA^{-1} \wedge dA + A^{-1}\omega \wedge \omega A + A^{-1}\omega \wedge dA \\
&+ A^{-1}dAA^{-1} \wedge \omega A + A^{-1}dA \wedge A^{-1}dA \\
&= A^{-1}d\omega A + A^{-1}\omega \wedge \omega A \\
\Omega' &= A^{-1}\Omega A
\end{aligned}$$

$\omega' = A^{-1}\omega A + A^{-1}dA$

These are interesting equations let us approach things from a more familiar setting so as to interpret what we have.

H.1 Curvature

An important fact about covariant derivatives is that they don't need to commute. If $\sigma : M \to E$ is a section and $X \in \mathfrak{X}(M)$ then $\nabla_X \sigma$ is a section also and so we may take it's covariant derivative $\nabla_Y \nabla_X \sigma$ with respect to some $Y \in \mathfrak{X}(M)$. In general, $\nabla_Y \nabla_X \sigma \neq \nabla_X \nabla_Y \sigma$ and this fact has an underlying geometric interpretation which we will explore later. A measure of this lack of commutativity is the **curvature operator** which is defined for a pair $X, Y \in \mathfrak{X}(M)$ to be the map $F(X,Y) : \Gamma(E) \to \Gamma(E)$ defined by

$$F(X,Y)\sigma := \nabla_X \nabla_Y \sigma - \nabla_Y \nabla_X \sigma - \nabla_{[X,Y]}\sigma.$$

or

$$[\nabla_X, \nabla_Y]\sigma - \nabla_{[X,Y]}\sigma$$

Appendix I

Notation and font usage guide

Category	Space or object	Typical elements	Typical morphisms
Vector Spaces	V , W , $\mathbb{R}^n$	v,w, x, y	A, B, K, λ, L
Banach Spaces	E,F,M,N, V,W, $\mathbb{R}^n$	v, w, x, y etc.	A, B, K, λ, L
Open sets in vector spaces	U, V, O, U_α	p, q, x, y, v, w	f, g, φ, ψ
Differentiable manifolds	M, N, P, Q	p, q, x, y	f, g, φ, ψ
Open sets in manifolds	U, V, O, U_α	p, q, x, y	$f, \varphi, \psi, \mathtt{x}_\alpha, \mathtt{y}_\alpha$
Bundles	$E \to M$	v, w, ξ, p, q, x	$(\bar{f}, f), (g, \mathrm{id}), h$
Sections of bundles	$\Gamma(M, E)$	$s, s_1, \sigma, ...$	f^*
Sections over open sets	$\Gamma(U, E) = \mathcal{S}_M^E(U)$	$s, s_1, \sigma, ...$	f^*
Lie Groups	G, H, K	g, h, x, y	h, f, g
Lie Algebras	$\mathfrak{g}, \mathfrak{h}, \mathfrak{k}, \mathfrak{a}, \mathfrak{b}$	v, x, y, z, ξ	h, g, df, dh
Fields	$\mathbb{F}, \mathbb{R}, \mathbb{C}, \mathbb{K}$	t, s, x, y, z, r	f, g, h
Vector Fields	$\mathfrak{X}_M(U)$,$\mathfrak{X}(M)$	X, Y, Z	f^*, f_*

Also we have the following notations

$C^\infty(U)$ or $\mathcal{F}(U)$	Smooth functions on U
$C_c^\infty(U)$ or $\mathcal{D}(U)$	"...." with compact support in U
T_pM	Tangent space at p
TM, with $\tau_M : TM \to M$	Tangent bundle of M
$T_pf : T_pM \to T_{f(p)}N$	Tangent map of $f : M \to N$ at p
$Tf : TM \to TN$	Tangent map of $f : M \to N$
T_p^*M	Cotangent space at p
T^*M, with $\pi_M : T^*M \to M$	Cotangent bundle of M
$J_x(M, N)_y$	k-jets of maps $f :: M, x \to N, y$
$\mathfrak{X}(U), \mathfrak{X}_M(U)$ (or $\mathfrak{X}(M)$)	Vector field over U (or over M)
$(\mathtt{x}, U), (\mathtt{x}_\alpha, U_\alpha), (\psi_\beta, U_\beta), (\varphi, U)$	Typical charts
$T_s^r(\mathrm{V})$ r-contravariant s-covariant	Tensors on V
$\mathfrak{T}_s^r(M)$ r-contravariant s-covariant	Tensor fields on M

d	exterior derivative, differential
∇	covariant derivative
M, g	Riemannian manifold with metric tensor g
M, ω	Symplectic manifold with symplectic form ω
$L(\mathrm{V}, \mathrm{W})$	Linear maps from V to W. Assumed bounded if V, W are Banach
$L_s^r(\mathrm{V}, \mathrm{W})$	r-contravariant, s-covariant multilinear maps $\mathrm{V}^{*r} \times \mathrm{V}^s \rightarrow \mathrm{W}$

In keeping with this we will later think of the space $L(\mathrm{V}, \mathrm{W})$ of linear maps from V to W as being identified with $\mathrm{W} \otimes \mathrm{V}^*$ rather than $\mathrm{V}^* \otimes \mathrm{W}$ (this notation will be explained in detail). Thus $(w \otimes \alpha)(v) = w\alpha(v) = \alpha(v)w$. This works nicely since if $w = (w^1, ..., w^m)^t$ is a column and $\alpha = (a_1, ..., a_n)$ then the linear transformation $w \otimes \alpha$ defined above has as matrix

$$w\alpha = \begin{bmatrix} w^1\alpha_1 & w^1\alpha_2 & \cdots & w^1\alpha_n \\ w^2\alpha_1 & w^2\alpha_2 & \cdots & w^2\alpha_n \\ \vdots & \vdots & & \vdots \\ w^m\alpha_1 & w^m\alpha_2 & \cdots & w^m\alpha_n \end{bmatrix}$$

while if $\mathrm{V} = \mathrm{W}$ the number $\alpha(w)$ is just the 1×1 matrix αw. To be consistent with this and with tensor notation we will write a matrix which is to be thought of as a linear map with one index down and one up. For instance, if $A = (a_j^i)$ then the i-th entry of $w = Av$ is

$$w^i = \sum_j a_j^i v^j.$$

A basis for a vector space is usually thought of as an ordered set of vectors (which are linearly independent and span) but it is not always convenient to index the basis elements by a set of integers such as $I = \{1, 2, ..., n\}$. For example, a basis for the space $\mathbb{R}_n^n$ of all $n \times n$ vectors (also denoted $\mathbb{M}_{n \times n}$) is more conveniently indexed by $I = \{1, 2, ..., n\} \times \{1, 2, ..., n\}$. We write elements of this set as ij rather than (i, j) so that a basis would be written $\{E_{ij}\}_{(ij) \in I^2}$. Thus a matrix which represents a linear transformation from $\mathbb{R}_n^n$ to $\mathbb{R}_n^n$ would be written something like $A = (A_{kl}^{ij})$ and given another such matrix say$B = (B_{kl}^{ij})$, the matrix multiplication $C = AB$ is given by

$$C_{kl}^{ij} = \sum_{a,b} A_{ab}^{ij} B_{kl}^{ab}.$$

The point is that it is convenient to define a basis for a (real) vector space to be an indexed set of vectors $\{v_\alpha\}_{\alpha \in \mathcal{I}}$ for some general index set $\mathcal{I}$ such that every vector $v \in \mathrm{V}$ has a unique representation as a sum

$$v = \sum c^\alpha v_\alpha$$

where all but a finite number of the coefficients $c^\alpha \in \mathbb{R}$ are zero. Since vector spaces use scalars which commute there is no harm in write $\sum v_\alpha c^\alpha$ instead of $\sum c^\alpha v_\alpha$. But notice that only the first (rather strange) expression is consistent with our matrix conventions. Thus $\sum v_\alpha c^\alpha = \sum c^\alpha v_\alpha = cv$ (not vc). In noncommutative geometry, these seemingly trivial issues acquire a much more serious character.

Remark I.1 *The use of superscripts is both useful and annoying. Useful because of the automatic bookkeeping features of expressions (like the above) but annoying because of potential confusion with powers. For example, the sphere would have to be denoted as the set of all* (x^1, x^2) *such that* $(x^1)^2 + (x^2)^2 = 1$. *This expression looks rather unfortunate to the author. In cases like this we might just use subscripts instead and hope the reader realizes from context that we have not employed any index lowering transformation such as* $x^i g_{ij} = x_i$ *but rather simply mean that* $x_i = x^i$. *Context must be the guide. Thus we write the equation for an n-sphere more comfortably as* $\sum_{i=1}^n x_i^2 = 1$.

I.1 Positional index notation and consolidation

With regard to the tensor algebras, recall our convention of covariant variables to the left and all the contravariant variables to the right. Also recall that when we formed the tensor product of, for example, $\tau \in \mathfrak{T}^1{}_1$ with $\sigma \in \mathfrak{T}^2{}_0$ we did not get a map $\mathfrak{X}^* \times \mathfrak{X} \times \mathfrak{X}^* \times \mathfrak{X}^* \to C^\infty$ which would be from a space which we could denote by $\mathfrak{T}^1{}_1{}^2{}_0$. Instead, we were formed what we once called the consolidated tensor product which put all the covariant variables together on the left. Thus the result was in $\mathfrak{T}^3{}_1$.

Putting all of the covariant variables on the left is just a convention and we could have done it the other way and used the space $\mathfrak{T}_s{}^r$ instead of $\mathfrak{T}^r{}_s$. But have we lost anything by this consolidation? In accordance with the above convention we have sometimes written the components of a tensor $\tau \in \mathfrak{T}^r{}_s$ as $\tau^{i_1 \dots i_r}{}_{j_i \dots j_s}$ instead of $\tau^{i_1 \dots i_r}_{j_i \dots j_s}$. But one sometimes sees components written such as, say, $\tau^i{}_j{}^k$ which is interpreted by many authors to refer to a multilinear map $\mathfrak{X}^* \times \mathfrak{X} \times \mathfrak{X}^* \to C^\infty$. This is indeed one possible meaning but we shall employ a different interpretation on most occasions. For example, if one defines a tensor whose components are at first written $\tau^i{}_{j\,k}$ and then later we see $\tau^i{}_j{}^k$, this may simply mean that an index operation has taken place. Index raising (and lowering) is defined and studied later on but basically the idea is that if there is a special 2−covariant tensor g_{ij} (metric or symplectic) defined which is a nondegenerate bilinear form on each tangent space then we can form the tensor g_{ij} defined by $\mathrm{g}_{ir}\mathrm{g}^{rj} = \delta^i_j$. Then from $\tau^i{}_{j\,k}$ we form the tensor whose components are $\tau^i{}_j{}^k := \tau^i{}_{j\,r}\mathrm{g}^{rk}$. Now some authors insist that this must mean that the new tensor is a multilinear map $\mathfrak{X}^* \times \mathfrak{X} \times \mathfrak{X}^* \to C^\infty$ but there is another approach which says that the index being in a new "unnatural" position is just a reminder that we have defined a new tensor from a previously defined tensor in certain way but that the result is still map $\mathfrak{X}^* \times \mathfrak{X}^* \times \mathfrak{X} \to C^\infty$. We have thus consolidated, putting covariant variables on the left. So $\tau^i{}_j{}^k$ actually refers to a new tensor $\widetilde{\tau} \in \mathfrak{T}^2{}_1$ whose components are $\widetilde{\tau}^{ik}{}_j := \tau^i{}_j{}^k := \tau^i{}_{j\,r}\mathrm{g}^{rk}$. We might also use the index position to indicate how the tensor was constructed. Many schemes are possible.

Why is it that in practice the interpretation of index position and the choice to consolidate or not seems to make little difference. The answer can best be understood by noticing that when one is doing a component calculation the indices take care of themselves. For example, in the expression $\tau^a{}_{bc}\,\theta_a X^b Y^c$ all one needs to know is how the components $\tau^a{}_{bc}$ are defined (e.g. $\tau_{rbc} g^{ar}\,\tau^a{}_{bc}$) and that X^b matches up with the b in $\tau^a{}_{bc}$ and so on. We have $\tau^a{}_{bc}\,\theta_a Y^c X^b = \tau^a{}_{bc}\,\theta_a X^b \theta_a Y^c$ and it matters not which variable comes first. One just has to be consistent. As another example, just notice how nicely it works out to say that the tensor product of $\tau^a{}_{bc}$ and $\eta^a{}_{bc}$ is the tensor $\tau^a{}_{bc}\,\eta^d{}_{ef}$. Does $\tau^a{}_{bc}\,\eta^d{}_{ef}$ refer to a multilinear map $\mathfrak{X}^* \times \mathfrak{X} \times \mathfrak{X} \times \mathfrak{X}^* \times \mathfrak{X} \times \mathfrak{X} \to C^\infty$ or to a multilinear map $\mathfrak{X}^* \times \mathfrak{X}^* \times \mathfrak{X} \times \mathfrak{X} \times \mathfrak{X} \times \mathfrak{X} \to C^\infty$? Does it matter?

Appendix J

Review of Curves and Surfaces

First of all a C^k map σ from an open set $U \subset \mathbb{R}^m$ into another Euclidean space $\mathbb{R}^n$ is given by n functions (say $\sigma_1, ..., \sigma_n$) of m variables (say $u^1, ..., u^m$) such that the functions and all possible partial derivatives of order $\leq k$ exist and are continuous throughout U. The total derivative at $p \in U$ of such a map is a linear map represented by the matrix of first partials:

$$\begin{bmatrix} \frac{\partial \sigma_1}{\partial u^1}(p) & \cdots & \frac{\partial \sigma_1}{\partial u^m}(p) \\ \vdots & & \vdots \\ \frac{\partial \sigma_n}{\partial u^1}(p) & \cdots & \frac{\partial \sigma_n}{\partial u^m}(p) \end{bmatrix}$$

and if this map is full rank for each $p \in U$ then we say that the

map is an C^k **immersion**. A C^k map $\phi : U \subset \mathbb{R}^n \to V \subset \mathbb{R}^n$ which is a homeomorphism and such that ϕ^{-1} is also a C^k map is called a **diffeomorphism**. By convention, if $k = 0$ then a diffeomorphism is really just a homeomorphism between open set of Euclidean space. (Here we have given rough and ready definitions that will refined in the next section on calculus on general Banach spaces).

J.0.1 Curves

Let O be an open set in $\mathbb{R}$. A continuous map $c : O \to \mathbb{R}^n$ is C^k if the $k-$th derivative $c^{(k)}$ exist on all of O and is a continuous. If I is a subset of $\mathbb{R}$ then a map $c : I \to \mathbb{R}^n$ is said to be C^k if there exists a C^k extension $\widetilde{c} : O \to \mathbb{R}^n$ for some open set O containing I. We are particularly interested in the case where I is an interval. This interval may be open, closed, half open etc. We also allow intervals for which one of the "end points" a or b is $\pm\infty$.

Definition J.1 *Let $O \subset \mathbb{R}$ be open. A continuous map $\mathbf{c} : O \to \mathbb{R}^n$ is said to be piecewise C^k if there exists a discrete sequence of points $\{t_n\} \subset O$ with*

$t_i < t_{i+1}$ such that $\mathbf{c}$ *restricted to each* $(t_i, t_{i+1}) \cap O$ *is* C^k. *If* I *is subset of* $\mathbb{R}$, *then a map* $\mathbf{c} : I \to \mathbb{R}^n$ *is said to be piecewise* C^k *if there exists a piecewise extension of* $\mathbf{c}$ *to some open set containing* I.

Definition J.2 *A* ***parametric curve*** *in* $\mathbb{R}^n$ *is a piecewise differentiable map* $\mathbf{c} : I \to \mathbb{R}^n$ *where* I *is either an interval or finite union of intervals. If* $\mathbf{c}$ *is an interval then we say that* $\mathbf{c}$ *is a* ***connected parametric curve***.

Definition J.3 *If* $\mathbf{c} : I \to \mathbb{R}^n$ *is a* C^1-*parametric curve then* $\mathbf{c}'$ *is well defined on* I *except, possibly, at a discrete set of points in* I *and is called the velocity of* $\mathbf{c}$. *If* $\|\mathbf{c}'\| = 1$ *where defined then we say that* $\mathbf{c}$ *is a* ***unit speed*** *curve. If* $\mathbf{c} : I \to \mathbb{R}^n$ *is a* C^2-*parametric curve then* $\mathbf{c}''$ *is a piecewise continuous map and is referred to as the acceleration of* c. *We call the vector* $c'(t)$ *the velocity at time* t *and call* $c''(t)$ *the acceleration at time* t.

Definition J.4 *A parametric curve is called* ***regular*** *if* c' *is defined and nonzero on all of* I.

We shall now restrict our attention to curves which are piecewise C^∞.

Definition J.5 *A elementary parametric curve is a regular curve* $c : I \to \mathbb{R}^n$ *such that* I *is an open connected interval.*

When should two curves be considered geometrically equivalent? Part of the answer comes from considering reparametrization:

Definition J.6 *If* $\mathbf{c} : I_1 \to \mathbb{R}^n$ *and* $b : I_2 \to \mathbb{R}^n$ *are curves then we say that* b *is a positive (resp. negative) reparametrization of* $\mathbf{c}$ *if there exists a bijection* $h : I_2 \to I_1$ *with* $\mathbf{c} \circ h = b$ *such that* h *is smooth and* $h'(t) > 0$ *(resp.* $h'(t) > 0$*) for all* $t \in I_2$.

Now we can think of two parametric curves as being in some sense the same if one is a C^k reparametrization of the other. On the other hand this form of congruence is appropriate to the topological and differentiable structure of $\mathbb{R}^n$.

Definition J.7 *Two parametric curves,* $\mathbf{c} : I_1 \to \mathbb{R}^n$ *and* $b : I_2 \to \mathbb{R}^n$, *are said to be congruent if there is a Euclidean motion* $T : \mathbb{R}^n \to \mathbb{R}^n$ *such that* $T \circ \mathbf{c}$ *is a reparametrization of* b.

We distinguish between a C^1 map $c : I \to \mathbb{R}^n$ and its image (or trace) $c(I)$ as a subset of $\mathbb{R}^n$. The geometric curve is the set $c(I)$ itself while the parameterized curve is the map in question. Of course, two sets, S_1 and S_2, are congruent (equivalent) if there exists a Euclidean motion $T : \mathbb{R}^n \to \mathbb{R}^n$ such that $T(S_1) = S_2$. The question now arises as to what our point of view should be. Are we intending to study certain subsets of $\mathbb{R}^n$ or are we ultimately interested in maps into $\mathbb{R}^n$. Sets or Maps? In the study of curves and surfaces both points of view are valuable. Unfortunately the two approaches are not always clearly distinguished in the literature. If subset S is the trace $\mathbf{c}(I)$ of some parametric curve then we say that $\mathbf{c}$ parameterizes the set S. Then one way to study this set geometrically is to study those aspects of the set which may be described in terms of a parameterization $\mathbf{c}$ but which would remain the same in some sense if we where to choose a different parameterization. If the parametrizations we use are bijections onto the image S then this approach works fairly well. On the other hand if we are interested in including self intersections then we have to be a bit careful. We can consider two maps $\mathbf{c} : I_1 \to \mathbb{R}^n$ and $b : I_2 \to \mathbb{R}^n$ to be congruent (in a generalized sense) if there exists a Euclidean motion $T : \mathbb{R}^n \to \mathbb{R}^n$ such that $T \circ \mathbf{c} = b$ is a reparametrization of $\mathbf{c}$. As we know from elementary calculus, every C^1 parameterized curve has a parameterization by arc length. In this case we often use the letter s to denote the independent variable. Now if two curves $\mathbf{c}_1$ and $\mathbf{c}_2$ are both unit speed reparametrizations of the same parameterized curve $\mathbf{c}$ then there is a transformation $s' = s + s_0$ so that $\mathbf{c}_2(s') = \mathbf{c}_1(s + s_0)$ where s_0 is a constant. Whenever we have a unit speed curve we say that the curve (or more precisely, the geometric curve $S = \mathbf{c}(I)$) is parameterized by arc length. Thus arc length parameterization is unique up to this kind of simple change of variables: $s = s' + s_0$

Given a parameterized curve $\mathbf{c} : I \to \mathbb{R}^n$ one may, in principal, immediately define the arc length of the curve from a reference point $p_0 = \mathbf{c}(t_0)$ to another point $p = \mathbf{c}(t)$. The formula familiar from elementary calculus is

$$l_{\mathbf{c},t_0}(t) = \int_{t_0}^{t} |\dot{\mathbf{c}}(\tau)|\, d\tau.$$

We will assume for simplicity that $|\dot{\mathbf{c}}(t)| > 0$ all $t \in I$. We will sometimes abbreviate $l_{\mathbf{c},t_0}$ to l whenever there is no danger of confusion. Since, $\frac{ds}{dt} = |\dot{\mathbf{c}}(t)| > 0$ we may invert and obtain $t = l^{-1}(s)$ with $t(t_0) = 0$ and $\frac{dt}{ds}(s) = 1/\frac{ds}{dt}(l^{-1}(s))$. Having done this we reparameterize our curve using l^{-1} and obtain $\widetilde{\mathbf{c}}(s) := \mathbf{c} \circ l^{-1}(s)$. We will abuse notation by using the same symbol $\mathbf{c}$ for this new parameterized curve. Now the unit tangent to the curve is defined by $\frac{d\mathbf{c}}{ds}(s) := \mathbf{T}(s)$. Notice that

$$\begin{aligned}
\sqrt{\mathbf{T}(s) \cdot \mathbf{T}(s)} &= |\mathbf{T}(s)| \\
&= \left| \frac{d\mathbf{c}}{ds}(s) \right| = \left| \frac{d\mathbf{c}}{dt} \frac{dt}{ds} \right| \\
&= \left| \frac{d\mathbf{c}}{dt} \right| \frac{dt}{ds} = |\dot{\mathbf{c}}(t)|\, |\dot{\mathbf{c}}(t)|^{-1} = 1
\end{aligned}$$

and so $\mathbf{T}$ is a unit vector. If $\frac{d\mathbf{T}}{ds}(s)$ is identically zero over some finite interval then $\mathbf{c}$ is easily seen to be a straight line. Because of this we will may as well assume that $\left|\frac{d\mathbf{T}}{ds}(s)\right| > 0$ for all $t \in I$. Now we have define $\mathbf{N}(s)$ as $\frac{d\mathbf{T}}{ds}(s) / \left|\frac{d\mathbf{T}}{ds}(s)\right|$ so we automatically get that $\mathbf{N}(s) = |\kappa| \frac{d\mathbf{T}}{ds}(s)$ where $|\kappa| := \left|\frac{d\mathbf{T}}{ds}(s)\right|$. The function $|\kappa|$ is often denoted simply as κ and is called the unsigned curvature. Observe that since $1 = |\mathbf{T}(s)|^2 = \mathbf{T}(s) \cdot \mathbf{T}(s)$ we get

$$0 = 2\frac{d\mathbf{T}}{ds}(s) \cdot \mathbf{T}(s)$$

and then $\mathbf{N}(s) \cdot \mathbf{T}(s) = 0$ so $\mathbf{N}(s)$ is a unit vector normal to the tangent $\mathbf{T}(s)$. The next logical step is to consider $\frac{d\mathbf{N}}{ds}$. Once again we separate out the case where $\frac{d\mathbf{N}}{ds}$ is identically zero in some interval. In this case we see that $\mathbf{N}$ is constant and it is not hard to see that the curve must remain in the fixed plane determined by $\mathbf{N}$. This plane is also oriented by the $\mathbf{T}$ and $\mathbf{N}$. The case of a curve in an oriented plane is equivalent to a curve in $\mathbb{R}^2$ and will be studied separately below. For now we assume that $\left|\frac{d\mathbf{N}}{ds}\right| > 0$ on I.

J.0.2 Curves in $\mathbb{R}^3$

In this case we may define the unit binormal vector to be the vector at $c(s)$ such that $\mathbf{T}, \mathbf{N}, \mathbf{B}$ is a positively oriented triple of orthonormal unit vectors. By positively oriented we mean that

$$\det[\mathbf{T}, \mathbf{N}, \mathbf{B}] = 1.$$

We now show that $\frac{d\mathbf{N}}{ds}$ is parallel to $\mathbf{B}$. For this it suffices to show that $\frac{d\mathbf{N}}{ds}$ is normal to both $\mathbf{T}$ and $\mathbf{N}$. First, we have $\mathbf{N}(s) \cdot \mathbf{T}(s) = 0$. If this equation is differentiated we obtain $2\frac{d\mathbf{N}}{ds} \cdot \mathbf{T}(s) = 0$. On the other hand we also have $1 = \mathbf{N}(s) \cdot \mathbf{N}(s)$ which differentiates to give $2\frac{d\mathbf{N}}{ds} \cdot \mathbf{N}(s) = 0$. From this we see that there must be a function $\tau = \tau(s)$ such that $\frac{d\mathbf{N}}{ds} := \tau\mathbf{B}$. This is a function of arc length but should really be thought of as a function on the curve. In a sense that we shall eventually make precise, τ is an invariant of the geometric curve. This invariant is called the **torsion**.

Theorem J.1 *If on some interval I we have* $|\dot{\mathbf{c}}(s)| > \mathbf{0}$, $\left|\frac{d\mathbf{T}}{ds}\right| > 0$, *and* $\left|\frac{d\mathbf{N}}{ds}\right| > 0$

with $\mathbf{T}$ *and* $\mathbf{N}$ *defined as above then*

$$\frac{d}{ds}[\mathbf{T},\mathbf{N},\mathbf{B}] = [\mathbf{T},\mathbf{N},\mathbf{B}] \begin{bmatrix} 0 & |\kappa| & 0 \\ -|\kappa| & 0 & \tau \\ 0 & \tau & 0 \end{bmatrix}$$

or in other words

$$\begin{array}{rlll} \frac{d\mathbf{T}}{ds} = & & |\kappa|\,\mathbf{N} & \\ \frac{d\mathbf{N}}{ds} = & -|\kappa|\,\mathbf{T} & & \tau\mathbf{B} \\ \frac{d\mathbf{B}}{ds} = & & \tau\mathbf{N} & \end{array}$$

Proof. Since $F = [\mathbf{T},\mathbf{N},\mathbf{B}]$ is by definition an orthogonal matrix we have $F(s)F^t(s) = I$. It is also clear that there is some matrix function $A(s)$ such that $F' = F(s)A(s)$. Also, Differentiating we have $\frac{dF}{ds}(s)F^t(s) + F(s)\frac{dF}{ds}^t(s) = 0$ and so

$$FAF^t + FA^tF^t = 0$$
$$A + A^t = 0$$

since F is invertible. Thus $A(s)$ is antisymmetric. But we already have established that $\frac{d\mathbf{T}}{ds} = |\kappa|\,\mathbf{N}$ and $\frac{d\mathbf{B}}{ds} = \tau\mathbf{N}$ and so the result follows. ■

As indicated above, it can be shown that the functions $|\kappa|$ and τ completely determine a sufficiently regular curve up to reparameterization and rigid motions of space. This is not hard to establish but we will save the proof for a later time. The three vectors form a vector field along the curve c. At each point $p = \mathbf{c}(s)$ along the curve c the provide and oriented orthonormal basis (or frame) for vectors based at p. This basis is called the Frenet frame for the curve. Also, $|\kappa|(s)$ and $\tau(s)$ are called the (unsigned) curvature and torsion of the curve at $\mathbf{c}(s)$. While, $|\kappa|$ is never negative by definition we may well have that $\tau(s)$ is negative. The curvature is, roughly speaking, the reciprocal of the radius of the circle which is tangent to $\mathbf{c}$ at $\mathbf{c}(s)$ and best approximates the curve at that point. On the other hand, τ measures the twisting of the plane spanned by $\mathbf{T}$ and $\mathbf{N}$ as we move along the curve. If $\gamma : I \to \mathbb{R}^3$ is an arbitrary speed curve then we define $|\kappa|_\gamma(t) := |\kappa| \circ h^{-1}$ where $h : I' \to I$ gives a unit speed reparameterization $\mathbf{c} = \gamma \circ h : I' \to \mathbb{R}^n$. Define the torsion function τ_γ for γ by $\tau \circ h^{-1}$. Similarly we have

$$\mathbf{T}_\gamma(t) := \mathbf{T} \circ h^{-1}(t)$$
$$\mathbf{N}_\gamma(t) := \mathbf{N} \circ h^{-1}(t)$$
$$\mathbf{B}_\gamma(t) := B \circ h^{-1}(t)$$

Exercise J.1 *If* $\mathbf{c} : I \to \mathbb{R}^3$ *is a unit speed reparameterization of* $\gamma : I \to \mathbb{R}^3$ *according to* $\gamma(t) = \mathbf{c} \circ h$ *then show that*

1. $\mathbf{T}_\gamma(t) = \gamma' / \|\gamma'\|$

2. $\mathbf{N}_\gamma(t) = \mathbf{B}_\gamma(t) \times \mathbf{T}_\gamma(t)$
3. $\mathbf{B}_\gamma(t) = \frac{\gamma' \times \gamma''}{\|\gamma' \times \gamma''\|}$
4. $|\kappa|_\gamma = \frac{\|\gamma' \times \gamma''\|}{\|\gamma''\|^3}$
5. $\tau_\gamma = \frac{(\gamma' \times \gamma'') \cdot \gamma'''}{\|\gamma' \times \gamma''\|^2}$

Exercise J.2 *Show that* $\gamma'' = \frac{dv}{dt}\mathbf{T}_\gamma + v^2 |\kappa|_\gamma \mathbf{N}_\gamma$ *where* $v = \|\gamma'\|$.

For a curve confined to a plane we haven't got the opportunity to define $\mathbf{B}$ or τ. However, we can obtain a more refined notion of curvature.

We now consider the special case of curves in $\mathbb{R}^2$. Here it is possible to define a signed curvature which will be positive when the curve is turning counterclockwise. Let $J : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $J(a, b) := (-b, a)$. The signed curvature $\kappa_{2,\gamma}$ of γ is given by

$$\kappa_\gamma(t) := \frac{\gamma''(t) \cdot J\gamma'(t)}{\|\gamma'(t)\|^3}.$$

Exercise J.3 *If* γ *is a parameterized curve in* $\mathbb{R}^2$ *then* $\kappa_\gamma \equiv 0$ *then* γ *(parameterizes) a straight line. If* $\kappa_\gamma \equiv k_0 > 0$ *(a constant) then* γ *parameterizes a portion of a circle of radius* $1/k_0$.

The unit tangent is $\mathbf{T} = \frac{\gamma'}{\|\gamma'\|}$. We shall redefine the normal $\mathbf{N}$ to a curve to be such that $\mathbf{T}, \mathbf{N}$ is consistent with the orientation given by the standard basis of $\mathbb{R}^2$. In fact, we have $\mathbf{N} = J\mathbf{c}'(s) = J\mathbf{T}$.

Exercise J.4 *If* $\mathbf{c} : I \to \mathbb{R}^2$ *is a unit speed curve then*

1. $\frac{d\mathbf{T}}{ds}(s) = \kappa_{\mathbf{c}}(s)\mathbf{N}(s)$
2. $\mathbf{c}''(s) = \kappa_{\mathbf{c}}(s)\,(J\mathbf{T}(s))$

J.1 Frenet Frames along curves in $\mathbb{R}^n$

We have already discussed the Frenet frames for curves in $\mathbb{R}^3$. We generalize this now for $n > 3$. We assume that $\gamma : I \to \mathbb{R}^n$ is a regular curve so that $\|\gamma'\| > 0$. For convenience we will assume that γ is infinitely differentiable (smooth). An **adapted moving orthonormal frame along** γ is a orthonormal set $\mathbf{E}_1(t), ..., \mathbf{E}_n(t)$ of smooth vector fields along γ such $\mathbf{E}_1(t) = \gamma'/\|\gamma'\|$. The moving frame will be called positively oriented if there is be a smooth matrix function $Q(t)$ of $n \times n$ orthogonal matrices of determinant 1 such that

$$[\mathbf{e}_1, ..., \mathbf{e}_n]Q(t) = [\mathbf{E}_1(t), ..., \mathbf{E}_n(t)]$$

where $\mathbf{e}_1, ..., \mathbf{e}_n$ is the standard basis of $\mathbb{R}^n$. Let us refer to such a moving frame simply as an orthonormal frame along the given curve.

We shall call a curve γ in R^n **fully regular** if $\gamma'(t), \gamma''(t), ..., \gamma^{(n)}(t)$ is a linearly independent set for each t in the domain of the curve. Now for a fully regular curve the existence of a moving orthonormal frame an easily proved. One essentially applies a Gram-Schmidt process: If $\mathbf{E}_1(t), ..., \mathbf{E}_k(t)$ are already defined then

$$\mathbf{E}_{k+1}(t) := \gamma^{(k+1)}(t) - \sum_{j=1}^{k} \gamma^{(j)}(t) \cdot \mathbf{E}_j(t).$$

This at least gives us an orthonormal moving frame and we leave it as an exercise to show that the $\mathbf{E}_k(t)$ are all smooth. Furthermore by making one possible adjustment of sign on $\mathbf{E}_n(t)$ we may guarantee that the moving frame is a positively oriented moving frame. So far we have a moving frame along our fully regular curve which is clearly quite nicely related to the curve itself. By our construction we have a nice list of properties:

1. For each k, $1 \leq k \leq n$ the vectors $\mathbf{E}_1(t), ..., \mathbf{E}_k(t)$ are in the linear span of $\gamma'(t), ..., \gamma^{(k)}(t)$ so that there is a upper triangular matrix function $L(t)$ such that
$$[\gamma'(t), ..., \gamma^{(n)}(t)]L(t) = [\mathbf{E}_1(t), ..., \mathbf{E}_n(t)]$$
2. For each k, $1 \leq k \leq n-1$ the vectors $\mathbf{E}_1(t), ..., \mathbf{E}_k(t)$ have the same orientation as $\gamma'(t), ..., \gamma^{(k)}(t)$. Thus the matrix $L(t)$ has its first $n-1$ diagonal elements all positive.
3. $[\mathbf{E}_1(t), ..., \mathbf{E}_n(t)]$ is positively oriented as a basis of $\mathbb{R}^n$.

Exercise J.5 *Convince yourself that the moving frame we have constructed is unique one with these properties.*

We call this moving frame the **Frenet frame** along γ. However, we can deduce more nice properties. The derivative of each $\mathbf{E}_i(t)$ is certainly expressible as a linear combination of the basis $\mathbf{E}_1(t), ..., \mathbf{E}_n(t)$ and so we may write

$$\frac{d}{dt}\mathbf{E}_i(t) = \sum_{j=1}^{n} \omega_{ij} \mathbf{E}_j(t).$$

Of course, $\omega_{ij} = \frac{d}{dt}\mathbf{E}_i(t) \cdot \mathbf{E}_j(t)$ but since $\mathbf{E}_i(t) \cdot \mathbf{E}_j(t) = \delta_{ij}$ we conclude that $\omega_{ij} = -\omega_{ji}$ so the matrix $\omega = (\omega_{ij})$ is antisymmetric. Furthermore, for $1 \leq j < n$ we have $\mathbf{E}_j(t) = \sum_{k=1}^{j} L_{kj}\gamma^{(k)}(t)$ where $L = (L_{kj})$ is the upper triangular matrix mentioned above. using the fact that L, $\frac{d}{dt}L$ and L^{-1} are all upper triangular we have

$$\frac{d}{dt}\mathbf{E}_j(t) = \sum_{k=1}^{j} \left(\frac{d}{dt}L_{kj}\right) \gamma^{(k)}(t) + \sum_{k=1}^{j} L_{kj}\gamma^{(k+1)}(t)$$

but $\gamma^{(k+1)}(t) = \sum_{r=1}^{k+1} \left(L^{-1}\right)_{(k+1)r} \mathbf{E}_r(t)$ and $\gamma^{(k)}(t) = \sum_{r=1}^{k} \left(L^{-1}\right)_{kr} \mathbf{E}_r(t)$ so that

$$
\begin{aligned}
\frac{d}{dt}\mathbf{E}_j(t) &= \sum_{k=1}^{j}\left(\frac{d}{dt}L_{kj}\right)\gamma^{(k)}(t) + \sum_{k=1}^{j} L_{kj}\gamma^{(k+1)}(t) \\
&= \sum_{k=1}^{j}\left(\frac{d}{dt}L_{kj}\right)\sum_{r=1}^{k}\left(L^{-1}\right)_{kr}\mathbf{E}_r(t) \\
&+ \sum_{k=1}^{j} L_{kj}\sum_{r=1}^{k+1}\left(L^{-1}\right)_{(k+1)r}\mathbf{E}_r(t)
\end{aligned}
$$

From this we see that $\frac{d}{dt}\mathbf{E}_j(t)$ is in the span of $\{\mathbf{E}_r(t)\}_{1\leq r\leq j+1}$. So $\omega = (\omega_{ij})$ can have no nonzero entries below the subdiagonal. But ω is antisymmetric and so we have no choice but to conclude that ω has the form

$$
\begin{bmatrix}
0 & \omega_{12} & 0 & \cdots & 0 \\
-\omega_{12} & 0 & \omega_{23} & & \\
 & -\omega_{23} & 0 & & \\
 & & & \ddots & \omega_{n-1,n} \\
 & & & -\omega_{n-1,n} & 0
\end{bmatrix}
$$

J.2 Surfaces

J.2.1 C^k Singular elementary surfaces.

This approach is seldom examined explicitly but is a basic conceptual building block for all versions of the word surface. At first we define a C^k singular surface to simply be a C^k map $\mathbf{x} : U \to \mathbb{R}^n$ where U is an open set in the plane $\mathbb{R}^2$. We are actually studying the maps themselves even though we often picture its image $\mathbf{x}(U)$ and retain a secondary interest in this image. The image may be a very irregular and complicated set and is called the trace of the singular elementary surface. A morphism[1] $\mu : \mathbf{x}_1 \to \mathbf{x}_2$ between two such elementary surfaces $\mathbf{x}_1 : U_1 \to \mathbb{R}^n$ and $\mathbf{x}_2 : U_2 \to \mathbb{R}^m$ is a pair of C^k maps $h : U_1 \to U_2$ and $\bar{h} : \mathbb{R}^n \to \mathbb{R}^m$ such that the following diagram commutes

$$
\begin{array}{ccc}
\mathbb{R}^n & \overset{\bar{h}}{\to} & \mathbb{R}^m \\
\mathbf{x}_1 \uparrow & & \mathbf{x}_2 \uparrow \\
U_1 & \overset{h}{\to} & U_2
\end{array}.
$$

If $m = n$ and both h and $\bar{h}$ are C^k diffeomorphisms then we say that the morphism is an isomorphism and that the elementary surfaces $\mathbf{x}_1 : U_1 \to \mathbb{R}^n$ and $\mathbf{x}_2 : U_2 \to \mathbb{R}^n$ are **equivalent** or **isomorphic**.

[1]Don't let the use of the word "morphism" here cause to much worry. So far the use of this word doesn't represent anything deep.

J.2.2 Immersed elementary surfaces.

Here we change things just a bit. An **immersed C^k elementary surface** (also called a **regular elementary surface**) is just a map as above but with the restriction that such $k > 0$ and that the derivative has full rank on all of U.

Remark J.1 *We may also need to deal with elementary surfaces which are regular on part of U and so we say that $\mathbf{x}$ is regular at p if the Jacobian matrix has rank 2 at p and also we say that $\mathbf{x}$ is regular on a set $A \subset U$ if it is regular at each $p \in A$. Often, A will be of the form $U \backslash N$ where N is a set consisting of a finite number of points or maybe some small set in some other set. For example N might be the image (trace) of a regular curve.*

What should be the appropriate notion of morphism between two such maps $\mathbf{x}_1 : U_1 \to \mathbb{R}^n$ and $\mathbf{x}_2 : U_2 \to \mathbb{R}^n$? Clearly, it is once again a pair of maps $h : U_1 \to U_2$ and $\bar{h} : \mathbb{R}^n \to \mathbb{R}^n$ but we must not end up creating maps which do not satisfy our new defining property. There is more than one approach that might work (or at least be self consistent) but in view of our ultimate goals we take the appropriate choice to be a pair of C^k diffeomorphisms $h : U_1 \to U_2$ and $\bar{h} : \mathbb{R}^n \to \mathbb{R}^n$ such that the following diagram commutes:

$$\begin{array}{ccc} \mathbb{R}^n & \overset{\bar{h}}{\to} & \mathbb{R}^n \\ \mathbf{x}_1 \uparrow & & \mathbf{x}_2 \uparrow \\ U_1 & \overset{h}{\to} & U_2 \end{array} .$$

J.2.3 Embedded elementary surface

This time we add the requirement that the maps $\mathbf{x} : U \to \mathbb{R}^n$ actually be regular elementary surfaces which are injective and in fact homeomorphism onto their images $\mathbf{x}(U)$.

Such a map $\mathbf{x}_1 : U_1 \to \mathbb{R}^n$ is called an **embedded elementary surface**. A morphism between two such embeddings is a pair of C^k diffeomorphisms with the obvious commutative diagram as before. All such a morphisms are in fact isomorphisms. (In some categories all morphisms are isomorphisms.)

J.2.4 Surfaces as Subspaces

The common notion of a surfaces is not a map but rather a set. It is true that in many if not most cases the set is the image of one or more maps. One approach is to consider a surface as a 2-dimensional smooth submanifold (defined shortly) of some $\mathbb{R}^n$. Of course for the surfaces that we see around us and are most compatible with intuition sit in $\mathbb{R}^3$ (or more precisely $\mathbf{E}^3$)

J.2.5 Geometric Surfaces

Here we mean the word geometric in the somewhat narrow sense of keeping track of such things as curvatures, volumes, lengths etc. There is a sort of division into extrinsic geometry and intrinsic geometry which is a theme which survives in more general settings such as Riemannian geometry.

Each of the ideas of the previous section has an obvious extension that starts with maps from open sets in $\mathbb{R}^m$ into $\mathbb{R}^n$. The case $m = 1$ is the study of curves. In each case we can take a topological and then differentiable view point. We can also consider the maps as the main object to be studied under the appropriate equivalences or we can think of sets a that which we study. The theory of differentiable manifolds which is one of the main topics of this book provides the tools to study our subject from each of these related views. Now when we actually start to get to the extra structures found in Riemannian, semi-Riemannian and symplectic geometry we still have the basic choice to center attention on maps (immersions, embedding, etc.) or sets but now the morphisms and therefore the notion of equivalence is decidedly more intricate. For example, if we are considering curves or surfaces as maps into some $\mathbb{R}^n$ then the notion of equivalence takes advantage of the metric structure and the structure provided by the natural inner product. Basically, is we have a map $\mathbf{x} : X \to \mathbb{R}^n$ from some open set X in $\mathbb{R}^m$, or even more generally, some abstract smooth surface or differentiable manifold. We might well decide to consider two such maps $\mathbf{x}_1 : X_1 \to \mathbb{R}^n$ and $\mathbf{x}_2 : X_2 \to \mathbb{R}^n$ to be equivalent if there is a diffeomorphism $h : X_1 \to X_2$ (the reparameterization) and a Euclidean motion $A : \mathbb{R}^n \to \mathbb{R}^n$ such that the following diagram commutes:

$$\begin{array}{ccc} \mathbb{R}^n & \overset{A}{\to} & \mathbb{R}^n \\ \mathbf{x}_1 \uparrow & & \mathbf{x}_2 \uparrow \\ X_1 & \overset{h}{\to} & X_2 \end{array} .$$

In the next section we study some of the elementary ideas from the classical theory of curves and surfaces. Because we have not yet studied the notion of a differentiable manifold in detail or even given more than a rough definition of a smooth surface we will mainly treat surfaces which are merely elementary surfaces. These are also called parameterized surfaces and are actually maps rather than sets. The read will notice however that our constructions and notions of equivalence has one eye as it were on the idea of a curve or surface as a special kind of set with a definite shape. In fact, "shape" is the basic motivating idea in much of differential geometry.

J.3 Surface Theory

We have studied elementary surfaces theory earlier in the book. We shall now redo some of that theory but now that we are clear on what a differentiable manifold is we shall be able to treat surfaces in a decidedly more efficient way. In particular, surfaces now are first conceived of as submanifolds. Thus they are now sets rather than maps. Of course we can also study immersions which possible have self crossings with the new advantage that the domain space can be an abstract (or concrete) 2-dimensional manifold and so may have rather complicated topology. When we first studied surfaces as maps the domains were open sets in R^2. We called these elementary parameterized surfaces. Let S be a submanifold of $\mathbb{R}^3$. The inverse of a coordinate map $\psi : V \to U \subset \mathbb{R}^2$ is a parameterization $\mathbf{x} :U \to V \subset S$ of a portion V of our surface. Let (u_1, u_2) the coordinates of points in V. there will be several such parameterization that cover the surface. We reiterate, what we are now calling a parameterization of a surface is not necessarily an onto map and is just the inverse of a coordinate map. For example, we have the usual parameterization of the sphere

$$\mathbf{x}(\varphi, \theta) = (\cos\theta \sin\varphi, \sin\theta \sin\varphi, \cos\varphi).$$

This parameterization is deficient at the north and south poles but covers all but a set of measure zero. For some purposes this is good enough and we can just use this one parameterization.

A curve on a surface S may be given by first letting $t \mapsto (u_1(t), u_2(t))$ be a smooth curve into U and then composing with $\mathbf{x} :U \to S$. For concreteness let the domain of the curve be the interval $[a, b]$. By the ordinary chain rule

$$\dot{\mathbf{x}} = \dot{u}_1 \partial_1 \mathbf{x} + \dot{u}_2 \partial_2 \mathbf{x}$$

and so the length of such a curve is

$$\begin{aligned} L &= \int_a^b |\dot{\mathbf{x}}(t)|\, dt = \int_a^b |\dot{u}_1 \partial_1 \mathbf{x} + \dot{u}_2 \partial_2 \mathbf{x}|\, dt \\ &= \int_a^b (g_{ij} \dot{u}_1 \dot{u}_2)^{1/2} dt \end{aligned}$$

where $g_{ij} = \partial_i \mathbf{x} \cdot \partial_j \mathbf{x}$. Let $p = \mathbf{x}(u_1, u_2)$ be arbitrary in $V \subset S$. The bilinear form g_p given on each $T_pS \subset T\mathbb{R}^3$ where $p = \mathbf{x}(u_1, u_2)$ given by

$$g_p(v, w) = g_{ij} v^i w^j$$

for $v_p = v^1 \partial_1 \mathbf{x} + v^2 \partial_2 \mathbf{x}$ gives a tensor g is called the first fundamental form or metric tensor. The classical notation is $ds^2 = \sum g_{ij} du_j du_j$ which does, whatever it's shortcomings, succinctly encode the first fundamental form. For example, if we parameterize the sphere $S^2 \subset \mathbb{R}^3$ using the usual spherical coordinates φ, θ we have

$$ds^2 = d\varphi^2 + \sin^2(\varphi) d\theta^2$$

from which the length of a curve $c(t) = \mathbf{x}(\varphi(t), \theta(t))$ is given by

$$L(c) = \int_{t_0}^{t} \sqrt{(\frac{d\varphi}{dt})^2 + \sin^2\varphi(t)(\frac{d\theta}{dt})^2} dt.$$

Now it may seem that we have something valid only in a single parameterization. Indeed the formulas are given using a single chart and so for instance the curve should not stray from the chart domain V. On the other hand, the expression $g_p(v, w) = g_{ij} v^i w^j$ is an invariant since it is just the length of the vector v as it sits in $\mathbb{R}^3$. So, as the reader has no doubt anticipated, $g_{ij}v^i w^j$ would give the same answer now matter what chart we used. By breaking up a curve into segments each of which lies in some chart domain we may compute it's length using a sequence of integrals of the form $\int (g_{ij}\dot{u}_1\dot{u}_2)^{1/2}dt$. It is a simple consequence of the chain rule that the result is independent of parameter changes. We also have a well defined notion of surface area of on S. This is given by

$$Area(S) := \int_S dS$$

and where dS is given locally by $\sqrt{g(u^1,u^2)}du^1du^2$ where $g := \det(g_{ij})$.

We will need to be able to produce normal fields on S. In a coordinate patch we may define

$$\begin{aligned} N &= \partial_1\mathbf{x}(u_1,u_2) \times \partial_2\mathbf{x}(u_1,u_2) \\ &= \det \begin{bmatrix} \frac{\partial x^1}{\partial u^1} & \frac{\partial x^1}{\partial u^2} & \mathbf{i} \\ \frac{\partial x^2}{\partial u^1} & \frac{\partial x^2}{\partial u^2} & \mathbf{j} \\ \frac{\partial x^3}{\partial u^1} & \frac{\partial x^3}{\partial u^2} & \mathbf{k} \end{bmatrix}. \end{aligned}$$

The unit normal field is then $\mathbf{n} = N/|N|$. Of course, $\mathbf{n}$ is defined independent of coordinates up to sign because there are only two possibilities for a normal direction on a surface in $\mathbb{R}^3$. The reader can easily prove that if the surface is orientable then we may choose a global normal field. If the surface is a closed submanifold (no boundary) then the two choices are characterized as inward and outward.

We have two vector bundles associated with S that are of immediate interest. The first one is just the tangent bundle of S which is in this setting embedded into the tangent bundle of $\mathbb{R}^3$. The other is the normal bundle NS which has as its fiber at $p \in S$ the span of either normal vector $\pm\mathbf{n}$ at p. The fiber is denoted N_pS. Our plan now is to take the obvious connection on $T\mathbb{R}^3$, restrict it to S and then decompose into tangent and normal parts. Restricting to the tangent and normal bundles appropriately,what we end up with is three connections. The obvious connection on $\mathbb{R}^3$ is simply $D_\xi(\sum_{i=1}^3 Y^i \frac{\partial}{\partial x^i}) := dY^i(\xi)\frac{\partial}{\partial x^i}$ which exist simply because we have a global distinguished coordinate frame $\{\frac{\partial}{\partial x^i}\}$. The fact that this standard frame is orthonormal with respect to the dot product on $\mathbb{R}^3$ is of significance here. We have both of the following:

1. $D_\xi(X\cdot Y) = D_\xi X \cdot Y + X \cdot D_\xi Y$ for any vector fields X and Y on $\mathbb{R}^3$ and any tangent vector ξ.

2. $D_\xi \circ D_v = D_v \circ D_\xi$ (This means the connection has no "torsion" as we define the term later).

Now the connection on the tangent bundle of the surface is defined by projection. Let ξ be tangent to the surface at p and Y a tangent vector field on the surface. Then by definition

$$\nabla_\xi Y = (D_\xi Y)^{\tan}$$

where $(D_\xi Y)^{\tan}(p)$ is the projection of $D_\xi Y$ onto the tangent planes to the surface at p. This gives us a map $\nabla : TS \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ which is easily seen to be a connection. Now there is the left over part $(D_\xi Y)^\perp$ but as a map $(\xi, Y) \mapsto (D_\xi Y)^\perp$ this does *not* give a connection. On the other hand, if η is a normal field, that is, a section of the normal bundle NS we define $\nabla^\perp_\xi \eta := (D_\xi \eta)^\perp$. The resulting map $\nabla^\perp : TS \times \Gamma(S, NS) \to \Gamma(S, NS)$ given by $(\xi, \eta) \mapsto \nabla^\perp_\xi \eta$ is indeed a connection on the normal bundle. Here again there is a left over part $(D_\xi \eta)^{\tan}$. What about these two other left over pieces $(D_\xi \eta)^{\tan}$ and $(D_\xi Y)^\perp$? These pieces measure the way the surface bends in $\mathbb{R}^3$. We define the shape operator at a point $p \in S$ with respect to a unit normal direction in the following way. First choose the unit normal field $\mathbf{n}$ in the chosen direction as we did above. Now define the shape operator $S(p) : T_pS \to T_pS$ by

$$S(p)\xi = -\nabla_\xi \mathbf{n}.$$

To see that the result is really tangent to the sphere just notice that $\mathbf{n} \cdot \mathbf{n} = 1$ and so $\nabla^\perp_\xi \mathbf{n}$

$$\begin{aligned} 0 &= \xi 1 = \xi(\mathbf{n} \cdot \mathbf{n}) \\ &= 2D_\xi \mathbf{n} \cdot \mathbf{n} \end{aligned}$$

which means that $D_\xi \mathbf{n} \in T_pS$. Thus the fact, that $\mathbf{n}$ had constant length gave us $\overline{\nabla}_\xi \mathbf{n} = (D_\xi \mathbf{n})^{\tan}$ and we have made contact with one of the two extra pieces. For a general normal section η we write $\eta = f\mathbf{n}$ for some smooth function on the surface and then

$$\begin{aligned} (D_\xi \eta)^{\tan} &= (D_\xi f\mathbf{n})^{\tan} \\ &= (df(\xi)\mathbf{n} + fD_\xi \mathbf{n})^{\tan} \\ &= -fS(p)\xi. \end{aligned}$$

so we obtain

Lemma J.1 $S(p)\xi = f^{-1}(D_\xi f\mathbf{n})^{\tan}$

The next result tell us that $S(p) : T_pS \to T_pS$ is symmetric with respect to the first fundamental form.

Lemma J.2 *Let* $v, w \in T_pS$. *Then we have* $g_p(S(p)v, w) = g_p(v, S(p)w)$.

Proof. The way we have stated the result hide something simple. Namely, tangent vector to the surface are also vectors in $\mathbb{R}^3$ under the usual identification of $T\mathbb{R}^3$ with $\mathbb{R}^3$. With this in mind the result is just $S(p)v \cdot w = v \cdot S(p)w$. Now this is easy to prove. Note that $\mathbf{n} \cdot w = 0$ and so $0 = v(\mathbf{n} \cdot w) = \overline{\nabla}_v \mathbf{n} \cdot w + \mathbf{n} \cdot \overline{\nabla}_v w$.

But the same equation holds with v and w interchanged. Subtracting the two expressions gives

$$\begin{aligned}0 &= \overline{\nabla}_v \mathbf{n} \cdot w + \mathbf{n} \cdot \overline{\nabla}_v w \\ &- \left(\overline{\nabla}_w \mathbf{n} \cdot v + \mathbf{n} \cdot \overline{\nabla}_w v\right) \\ &= \overline{\nabla}_v \mathbf{n} \cdot w - \overline{\nabla}_w \mathbf{n} \cdot v + \mathbf{n} \cdot (\overline{\nabla}_v w - \overline{\nabla}_w v) \\ &= \overline{\nabla}_v \mathbf{n} \cdot w - \overline{\nabla}_w \mathbf{n} \cdot v\end{aligned}$$

from which the result follows. ∎

Since $S(p)$ is symmetric with respect to the dot product there are eigenvalues κ_1, κ_2 and eigenvectors $v_{\kappa_1}, v_{\kappa_2}$ such that $v_{\kappa_i} \cdot S(p) v_{\kappa_j} = \delta_{ij} \kappa_i$. Let us calculate in a special coordinate system containing our point p obtained by projecting onto the tangent plane there. Equivalently, we rigidly move the surface until p is at the origin of $\mathbb{R}^3$ and is tangent to the x, y plane. Then the surface is parameterized near p by $(u^1, u^2) \mapsto (u^1, u^2, f(u^1, u^2))$ for some smooth function f with $\frac{\partial f}{\partial u^1}(0) = \frac{\partial f}{\partial u^2}(0) = 0$. At the point p, which is now the origin, we have $g_{ij}(0) = \delta_{ij}$. Since S is now the graph of the function f the tangent space $T_p S$ is identified with the x, y plane. A normal field is given by $\operatorname{grad} F = \operatorname{grad}(f(x, y) - z) = (\frac{\partial f}{\partial u^1}, \frac{\partial f}{\partial u^2}, -1)$ and the unit normal is

$$\mathbf{n}(u^1, u^2) = \frac{1}{\sqrt{(\frac{\partial f}{\partial u^1})^2 + (\frac{\partial f}{\partial u^2})^2 + 1}} (\frac{\partial f}{\partial u^1}, \frac{\partial f}{\partial u^2}, -1)$$

Letting $r(u^1, u^2) := ((\frac{\partial f}{\partial u^1})^2 + (\frac{\partial f}{\partial u^2})^2 + 1)^{1/2}$ and using lemma J.1 we have $S(p)\xi = r^{-1}(\overline{\nabla}_\xi r\mathbf{n})^{\tan} = r^{-1}(\overline{\nabla}_\xi N)^{\tan}$ where $N := (\frac{\partial f}{\partial u^1}, \frac{\partial f}{\partial u^2}, -1)$. Now at the origin $r = 1$ and so $\xi \cdot S(p)\xi = \overline{\nabla}_\xi N \cdot \xi = \frac{\partial}{\partial u^k} \frac{\partial F}{\partial u^i} \xi^k \xi^i = \frac{\partial^2 F}{\partial u^k \partial u^i} \xi^k \xi^i$ from which we get the following:

$$\xi \cdot S(p) v = \sum_{ij} \xi^i v^j \frac{\partial f}{\partial u^i \partial u^j}(0)$$

valid for these special type of coordinates and only at the central point p. Notice that this means that once we have the surface positioned as a graph over the x, y-plane and parameterized as above then

$$\xi \cdot S(p) v = D^2 f(\xi, v) \text{ at } 0.$$

Here we must interpret ξ and v on the right hand side to be (ξ^1, ξ^2) and (v^1, v^2) where as on the left hand side $\xi = \xi^1 \frac{\partial \mathbf{x}}{\partial u^1} + \xi^2 \frac{\partial \mathbf{x}}{\partial u^2}$, $v = v^1 \frac{\partial \mathbf{x}}{\partial u^1} + v^2 \frac{\partial \mathbf{x}}{\partial u^2}$.

Exercise J.6 *Position S to be tangent to the x, y plane as above. Let the x, z plane intersect S in a curve c_1 and the y, z plane intersect S in a curve c_2. Show that by rotating we can make the coordinate vectors $\frac{\partial}{\partial u^1}, \frac{\partial}{\partial u^2}$ be eigenvectors for $S(p)$ and that the curvatures of the two curves at the origin are κ_1 and κ_2.*

We have two important invariants at any point p. The first is the Gauss curvature $K := \det(S) = \kappa_1\kappa_2$ and the second is the mean curvature $H = \frac{1}{2}trace(S) = \frac{1}{2}(\kappa_1 + \kappa_2)$.

The sign of H depends on which of the two normal directions we have chosen while the sign of κ does not. In fact, the Gauss curvature turns out to be "intrinsic" to the surface in the sense that it remains constant under any deformation of the surface the preserves lengths of curves. More on this below but first let us establish a geometric meaning for H. First of all, we may vary the point p and then S becomes a function of p and the same for H (and K).

Theorem J.2 *Let S_t be a family of surfaces given as the image of maps $h_t : S \to \mathbb{R}^3$ and given by $p \mapsto p + t\mathbf{v}$ where $\mathbf{v}$ is a section of $T\mathbb{R}^3\big|_S$ with $\mathbf{v}(0) = 1$ and compact support. Then*

$$\left.\frac{d}{dt}\right|_{t=0} area(S_t) = -\int_S (\mathbf{v} \cdot H\mathbf{n})dS$$

More generally, the formula is true if $h : (-\epsilon, \epsilon) \times S \to \mathbb{R}^3$ is a smooth map and $\mathbf{v}(p) := \frac{d}{dt}\big|_{t=0} h(t, p)$.

Exercise J.7 *Prove the above theorem by first assuming that $\mathbf{v}$ has support inside a chart domain and then use a partition of unity argument to get the general case.*

Surface S is called a **minimal surface** if $H \equiv 0$ on S. It follows from theorem J.2 that if S_t is a family of surfaces given as in the theorem that if S_0 is a minimal surface then 0 is a critical point of the function $a(t) := area(S_t)$. Conversely, if 0 is a critical point for all such variations of S then S is a minimal surface.

Exercise J.8 *(doody this one in Maple) Show that Sherk's surface, which is given by $e^z \cos(y) = \cos x$, is a minimal surface. If you haven't seen this surface do a plot of it using Maple, Mathematica or some other graphing software. Do the same for the helicoid $y \tan z = x$.*

Now we move on to the Gauss curvature K. Here the most important fact is that K may be written in terms of the first fundamental form. The significance of this is that if S_1 and S_2 are two surfaces and if there is a map $\phi : S_1 \to S_2$ that preserves the length of curves, then κ^{S_1} and κ^{S_2} are the same in the sense that $K^{S_1} = K^{S_2} \circ \phi$. In the following theorem, "$g_{ij} = \delta_{ij}$ to first order" means that $g_{ij}(0) = \delta_{ij}$ and $\frac{\partial g_{ij}}{\partial u}(0) = \frac{\partial g_{ij}}{\partial v}(0) = 0$.

Theorem J.3 (Gauss's Theorema Egregium) *Let $p \in S$. There always exist coordinates u, v centered at p (so $u(p) = 0, v(p) = 0$) such that $g_{ij} = \delta_{ij}$ to first order at 0 and for which we have*

$$K(p) = \frac{\partial^2 g_{12}}{\partial u \partial v}(0) - \frac{1}{2}\frac{\partial^2 g_{22}}{\partial u^2}(0) - \frac{1}{2}\frac{\partial^2 g_{11}}{\partial v^2}(0).$$

Proof. In the coordinates described above which give the parameterization $(u,v) \mapsto (u,v,f(u,v))$ where p is the origin of $\mathbb{R}^3$ we have

$$\begin{bmatrix} g_{11}(u,v) & g_{12}(u,v) \\ g_{21}(u,v) & g_{22}(u,v) \end{bmatrix} = \begin{bmatrix} 1+(\frac{\partial f}{\partial x})^2 & \frac{\partial f}{\partial x}\frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial x}\frac{\partial f}{\partial y} & 1+(\frac{\partial f}{\partial y})^2 \end{bmatrix}$$

from which we find after a bit of straightforward calculation

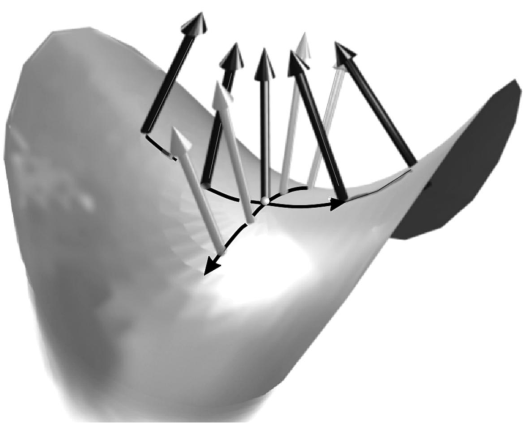
$$\begin{aligned} &\frac{\partial^2 g_{12}}{\partial u \partial v}(0) - \frac{1}{2}\frac{\partial^2 g_{22}}{\partial u^2}(0) - \frac{1}{2}\frac{\partial^2 g_{11}}{\partial v^2}(0) \\ &= \frac{\partial^2 f}{\partial u^2}\frac{\partial^2 f}{\partial v^2} - \frac{\partial^2 f}{\partial u \partial v} = \det D^2 f(0) \\ &= \det S(p) = K(p). \end{aligned}$$

■

Note that if we have any other coordinate system s,t centered at p then writing $(u,v) = (x^1,x^2)$ and $(s,t) = (\bar{x}^1,\bar{x}^2)$ we have the transformation law

$$\bar{g}_{ij} = g_{kl}\frac{\partial x^k}{\partial \bar{x}^i}\frac{\partial x^l}{\partial \bar{x}^j}$$

which means that if we know the metric components in any coordinate system then we can get them, and hence $K(p)$, at any point in any coordinate system. The conclusion is the that the metric determines the Gauss curvature. We say that K is an **intrinsic invariant**.



Negative Gauss Curvature

Now every field $\bar{X} \in \mathfrak{X}(\overline{M})\big|_M$ is uniquely written as $X^{\tan} + X^{\perp}$ where $X^{\tan} \in \mathfrak{X}(M)$ and $X^{\perp} \in \mathfrak{X}(M)^{\perp}$. Now for any $\bar{Y} \in \mathfrak{X}(\overline{M})\big|_M$ and $X \in \mathfrak{X}(M)$ we have the decomposition

$$\bar{\nabla}_X \bar{Y} = \left(\bar{\nabla}_X \bar{Y}\right)^{\tan} + \left(\bar{\nabla}_X \bar{Y}\right)^{\perp}$$

and writing $\bar{Y} = Y^{\tan} + Y^{\perp}$ we have

$$\begin{aligned}
&\left(\bar{\nabla}_X\left(Y^{\tan}+Y^{\perp}\right)\right)^{\tan}+\left(\bar{\nabla}_X\left(Y^{\tan}+Y^{\perp}\right)\right)^{\perp}\\
&=\left(\bar{\nabla}_X Y^{\tan}\right)^{\tan}+\left(\bar{\nabla}_X Y^{\perp}\right)^{\tan}\\
&+\left(\bar{\nabla}_X Y^{\tan}\right)^{\perp}+\left(\bar{\nabla}_X Y^{\perp}\right)^{\perp}
\end{aligned}$$

Now we will show that $\langle\left(\bar{\nabla}_X Y^{\tan}\right)^{\perp}, Y^{\perp}\rangle = -\langle Y^{\tan}, \left(\bar{\nabla}_X Y^{\tan}\right)^{\tan}\rangle$. Indeed, $\langle Y^{\tan}, Y^{\perp}\rangle = 0$ and so

$$\begin{aligned}
0 &= \bar{\nabla}_X\langle Y^{\tan}, Y^{\perp}\rangle = \langle\bar{\nabla}_X Y^{\tan}, Y^{\perp}\rangle + \langle Y^{\tan}, \bar{\nabla}_X Y^{\perp}\rangle\\
&= \langle\left(\bar{\nabla}_X Y^{\tan}\right)^{\perp}, Y^{\perp}\rangle + \langle Y^{\tan}, \left(\bar{\nabla}_X Y^{\tan}\right)^{\tan}\rangle
\end{aligned}$$

Notice that $Y^{\tan}$ and $Y^{\perp}$ may vary independently. We now want to interpret what we have so far.

Definition J.8 *For $X_1, X_2 \in \mathfrak{X}(M)$ we define $\nabla_{X_1}X_2 := \left(\bar{\nabla}_{X_1}X_2\right)^{\tan}$.*

$(X_1, X_2) \to \nabla_{X_1}X_2$ defines a connection on M (i.e. a connection on the bundle TM). Thus $(X_1, X_2) \to \nabla_{X_1}X_2$ is $C^\infty(M)$-linear in X_1. It follows, as usual, that for any $v \in T_pM$ the map $v \mapsto \nabla_v X$ is well defined. The connection is actually the Levi-Civita connection for M. To verify this we just need to check that $\nabla_{X_1}X_2 - \nabla_{X_2}X_1 = [X_1, X_2]$ for all $X_1, X_2 \in \mathfrak{X}(M)$ and that

$$\nabla_X\langle Y, Z\rangle = \langle\nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle$$

for all $X, Y, Z \in \mathfrak{X}(M)$. But both of these follow directly from the corresponding facts for $\bar{\nabla}$.

Definition J.9 *For any $X \in \mathfrak{X}(M)$, $Y \in \mathfrak{X}(M)^{\perp}$ we define $\nabla^{\perp}_X Y := \left(\bar{\nabla}_X T\right)^{\perp}$*

$(X, Y) \to \nabla^{\perp}_X Y$ defines a connection on the vector bundle $TM^{\perp} \to M$ which is again something easily deduced.

Exercise J.9 *Check the details, show that $\bar{\nabla}$ restricted to the bundle $T\bar{M}\big|_M$ is a connection and that $\nabla^{\perp}$ is a connection on $TM^{\perp}$. Show also that not only does $\nabla_{X_1}X_2 := \left(\bar{\nabla}_{X_1}X_2\right)^{\tan}$ define a connection of TM but that this connection is the Levi-Civita connection for M were M has the semi-Riemannian metric given by restriction of the metric on $\bar{M}$.*

Definition J.10 *Let $II : \mathfrak{X}(M)\times\mathfrak{X}(M) \to \mathfrak{X}(M)^{\perp}$ be defined by $II(X_1, X_2) := \left(\bar{\nabla}_{X_1}X_2\right)^{\perp}$. This $\mathfrak{X}(M)^{\perp}$-valued bilinear map called the second fundamental tensor. The set of all elements $u \in TM^{\perp}$ of unit length is called the unit normal bundle of M in $\bar{M}$. It is denoted by $TM^{\perp}_1$.*

We now have the decomposition

$$\bar{\nabla}_{X_1}X_2 = \nabla_{X_1}X_2 + II(X_1, X_2)$$

for any $X_1, X_2 \in \mathfrak{X}(M)$.

Lemma J.3 ***Proposition J.1*** *II is $C^\infty(M)$ bilinear and symmetric.*

Proof. The linearity in the first variable is more or less obvious. We shall be content to show that

$$\begin{aligned} II(X_1, fX_2) &= \left(\bar{\nabla}_{X_1} fX_2\right)^{\perp} \\ &= \left((X_1 f)\, X_2 + f\bar{\nabla}_{X_1} X_2\right)^{\perp} \\ &= f\left(\bar{\nabla}_{X_1} X_2\right)^{\perp} = fII(X_1, X_2). \end{aligned}$$

Symmetry: $II(X_1, X_2) - II(X_2, X_1) = \left(\bar{\nabla}_{X_1} X_2 - \bar{\nabla}_{X_2} X_1\right)^{\perp}$
$= ([X_1, X_2])^{\perp} = 0$ ∎

Definition J.11 *Let $p \in M$. For each unit vector u normal to M at p we have a map called the* ***shape operator*** *S_u defined by $S_u(v) := \left(\bar{\nabla}_v U\right)^{\tan}$ where U is any unit normal field defined near p such that $U(p) = u$.*

Now for any $Z \in \mathfrak{X}(M)^{\perp}$ and $X \in \mathfrak{X}(M)$ we consider $\bar{\nabla}_X Z$. We decompose this as $\bar{\nabla}_X Z = \left(\bar{\nabla}_X Z\right)^{\tan} + \left(\bar{\nabla}_X Z\right)^{\perp} = S_Z(X) + \nabla^{\perp}_X Z$

J.3.1 Symmetry and Noether's theorem for field theory

Appendix K

Save

A transitive $G-$space, M, is essentially a group theoretic object in a sense that we now describe. The reader will understand the situation much better if s/he does the following exercises before proceeding to the abstract situation.

Exercise K.1 *Show that the action of $Aff^+(\mathbf{A}^2)$ on $\mathbf{A}^2$ is transitive and effective but not free.*

Exercise K.2 *Fix a point x_0 (say $(0,0)$) in $\mathbf{A}^2$ and show that $H := \{g \in Aff^+(\mathbf{A}^2) : gx_0 = x_0\}$ is a closed subgroup of $Aff^+(\mathbf{A}^2)$ isomorphic to* $\mathrm{SL}(2)$.

Exercise K.3 *Let $H \cong \mathrm{SL}(2)$ be as in the previous exercise. Show that there is a natural 1-1 correspondence between the cosets of $Aff^+(\mathbf{A}^2)/H$ and the points of $\mathbf{A}^2$.*

Exercise K.4 *Show that the bijection of the previous example is a homeomorphism if we give $Aff^+(\mathbf{A}^2)/H$ its quotient topology.*

Exercise K.5 *Let S^2 be the unit sphere considered a subset of $\mathbb{R}^3$ in the usual way. Let the group $SO(3)$ act on S^2 by rotation. This action is clearly continuous and transitive. Let $n = (0,0,1)$ be the "north pole". Show that if H is the (closed) subgroup of $SO(3)$ consisting of all $g \in SO(3)$ such that $g \cdot n = n$ then $x = g \cdot n \mapsto g$ gives a well defined bijection $S^2 \cong SO(3)/H$. Note that $H \cong \mathrm{SO}(2)$ so we may write $S^2 \cong \mathrm{SO}(3)/\mathrm{SO}(2)$.*

Exercise K.6 *Show that the bijection of the previous example is a homeomorphism if we give* $\mathrm{SO}(3)/H$ *its quotient topology.*

Definition K.1 *Let M be a transitive G-space. Fix $x_0 \in M$. The isotropy subgroup for the point x_0 is*

$$G_{x_0} := \{g \in G : gx_0 = x_0\}$$

Theorem K.1 *Let M be a transitive (left) G-space and fix $x_0 \in M$. Let G_{x_0} be the corresponding **isotropy subgroup** Then we have a natural bijection*

$$G \cdot x_0 \cong G/G_{x_0}$$

given by $g \cdot x_0 \mapsto gG_{x_0}$. In particular, if the action is transitive then $G/G_{x_0} \cong M$ and x_0 maps to H.

For the following discussion we let M be a transitive (left) G-space and fix $x_0 \in M$. The action of G on M may be transferred to an equivalent action on G/G_{x_0} via the bijection of the above theorem. To be precise, if we let Φ denote the bijection of the theorem $G/G_{x_0} \cong M$ and let λ denote the actions of G on M and G respectively, then λ' is defined so that the following diagram commutes:

$$\begin{array}{cccc} & G \times G/G_{x_0} & \xrightarrow{\lambda'} & G/G_{x_0} \\ id_G \times \Phi & \downarrow & & \downarrow \Phi \\ & G \times M & \xrightarrow{\lambda} & M \end{array}$$

This action turns out to be none other than the most natural action of G on G/G_{x_0}. Namely,

$$(g, xG_{x_0}) \mapsto gxG_{x_0}.$$

We have left out consideration of the topology. That is, what topology should we take on G/G_{x_0} so that all of the maps of the above diagram are continuous? If we just use the bijection to transfer the topology of M over to G/G_{x_0}. In this case, all of the actions are continuous. On the other hand, G/G_{x_0} has the quotient topology and we must face the possibility that the two topologies do not coincide! In order to not interrupt discussion we will defer dealing with this question until chapter 16. For now we just comment that for each of the examples discussed in this chapter the two topologies are easily seen to coincide.

Exercise K.7 *Prove theorem K.1 and the above statement about the equivalent action of G on G/G_{x_0}.*

Now any closed subgroup $H \subset G$ is the isotropy subgroup of the coset H for the natural action $G \times G/H \to G/H$ (where $(g, xH) \mapsto gxH$). The upshot of all this is that we may choose to study coset spaces G/H as $G-$spaces since every transitive $G-$spaces is equivalent to one of this form.

K.0.2 Euclidean space

As we have mentioned, if we have an affine space A modelled on an n-dimensional vector space V then if we are also given an inner product $\langle , \rangle$ on V, we may at once introduce several new structures. For one thing, we have a notion of length of vectors and this can be used to define a distance function on A itself. From the definition of affine space we know that for any two elements $p, q \in A$, there is a unique translation vector v such that $p + v = q$. If we denote this v suggestively by $q - p$ then the distance is $\text{dist}(p, q) := \|q - p\|$ where $\|.\|$ is the norm defined by the inner product $\|v\| = \langle v, v\rangle^{1/2}$. There is also a special family of coordinates that are constructed by choosing an orthonormal basis $e_1, ..., e_n$ for V. Having chosen the orthonormal basis, we get a map $V \to \mathbb{R}^n$

$$v = v^1 e_1 + \cdots + v^n e_n \mapsto (v^1, ..., v^n).$$

Now picking one of the bijections $A \to V$ centered at some point and composing this with the map $V \to \mathbb{R}^n$ that we got from the basis we get the desired coordinate map $A \to \mathbb{R}^n$. All coordinate systems constructed in this way are related by affine transformations of the form $L : x \mapsto Lx_0 + Q(x - x_0)$ for some $L \in O(V)$. The set of all such transformations is a group which is called the **Euclidean motion** group $\mathrm{Euc}(A, V, \langle,\rangle)$. This group acts transitively on A. If we require L be orientation preserving then we have the proper **Euclidean motion** group $\mathrm{Euc}^+(A, V, \langle,\rangle)$ which also acts transitively on A.

Exercise K.8 *Show that for any point $x_0 \in A$, the isotropy subgroup $G_{x_0} \subset \mathrm{Euc}^+(A, V, \langle,\rangle)$ is isomorphic to $\mathrm{SO}(V)$.*

Exercise K.9 *Show that action of $G = \mathrm{Euc}(A, V, \langle,\rangle)$ on A is transitive and continuous. So that if H is the isotropy subgroup corresponding to some point $x_0 \in A$ then we have the ($G-$space) isomorphism $A \cong G/H$. Exhibit this isomorphism in the concrete case of $A = \mathbf{E}^2$ and $G = Euc(2)$.*

In the finite dimensional situation we will usually stick to the concrete case of $\mathrm{Euc}(n)$ (resp. $\mathrm{Euc}^+(n)$) acting on $\mathbf{E}^n$ and take $x_0 = 0 \in \mathbf{E}^n$. Using the matrix representation described above, the isotropy subgroup of $x_0 = 0$ consists of all matrices of the form $\begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}$ for $Q \in \mathrm{O}(n)$ (resp. $S\mathrm{O}(n)$). Identifying this isotropy subgroup with $\mathrm{O}(n)$ (resp. $S\mathrm{O}(n)$) we have

$$\mathbf{E}^n \cong \mathrm{Euc}(n)/\mathrm{O}(n) \cong \mathrm{Euc}^+(n)/S\mathrm{O}(n).$$

We also have the homogeneous space presentation

$$M^{1+3} \cong P/O(1,3)$$

where we have identified the isotropy subgroup of a point in M^{1+3} with $O(1,3)$.

Appendix L

Interlude: A Bit of Physics

L.1 The Basic Ideas of Relativity

We will draw an analogy with the geometry of the Euclidean plane. Recall that the abstract plane $\mathcal{P}$ is *not* the same thing as coordinate space $\mathbb{R}^2$ but rather there are many "good" bijections $\Psi : \mathcal{P} \to \mathbb{R}^2$ called coordinatizations such that points $p \in \mathcal{P}$ corresponding under Ψ to coordinates $(x(p), y(p))$ are temporarily identified with the pair $(x(p), y(p))$ is such a way that the distance between points is given by $dist(p,q) = \sqrt{(x(p) - x(q))^2 + (y(p) - y(q))^2}$ or

$$dist(p,q)^2 = \Delta x^2 + \Delta y^2$$

for short. Now the numbers Δx and Δy separately have no absolute meaning since a different good-coordinatization $\Phi : \mathcal{P} \to \mathbb{R}^2$ would give something like $(X(p), Y(p))$ and then for the same two points p, q we expect that in general $\Delta x \neq \Delta X$ and $\Delta y \neq \Delta Y$. On the other hand, the notion of distance is a geometric reality that should be independent of coordinate choices and so we always have $\Delta x^2 + \Delta y^2 = \Delta X^2 + \Delta Y^2$. But what is a "good coordinatization"? Well, one thing can be said for sure and that is if x, y are good then X, Y will be good also if

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \cos\theta & \pm\sin\theta \\ \mp\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$$

for some θ and numbers T_1, T_2. The set of all such transformations form a group under composition is called the Euclidean motion group. Now the idea of points on an abstract plane is easy to imagine but it is really just a "set" of objects with some logical relations; an idealization of certain aspects of our experience. Similarly, we now encourage the reader to make the following idealization. Imagine the set of all ideal local events or possible events as a sort of 4-dimensional plane. Imagine that when we speak of an event happening at location (x, y, z) in rectangular coordinates and at time t it is only because we have imposed some sort of coordinate system on the set of events that is implicit in our norms regarding measuring procedures etc. What if some other system were used to describe the same set of events, say, two explosions $e1$ and $e2$. You would not be surprised to find out that the spatial separations for the two events

$$\Delta X, \Delta Y, \Delta Z$$

would not be absolute and would not individually equal the numbers

$$\Delta x, \Delta y, \Delta z.$$

But how about ΔT and Δt. Is the time separation, in fixed units of seconds say, a real thing?

The answer is actually *no* according to the special theory of relativity. In fact, not even the quantities $\Delta X^2+\Delta Y^2+\Delta Z^2$ will agree with $\Delta x^2+\Delta y^2+\Delta y^2$ under certain circumstances! Namely, if two observers are moving relative to each other at constant speed, there will be objectively irresolvable disagreements. Who is right? There simply is no fact of the matter. The objective or absolute quantity is rather

$$-\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta y^2$$

which always equals $-\Delta T^2 + \Delta X^2 + \Delta Y^2 + \Delta Y^2$ for good coordinates systems. But what is a good coordinate system? It is one in which the equations of physics take on their simplest form. Find one, and then all others are related by the Poincaré group of linear transformations given by

$$\begin{pmatrix} X \\ Y \\ Z \\ T \end{pmatrix} = A \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \\ z_0 \\ t_0 \end{pmatrix}$$

where the matrix A is a member of the Lorentz group. The Lorentz group is characterized as that set $O(1,3)$ of matrices A such that

$$A^T \eta A = \eta \quad \text{where}$$
$$\eta := \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

This is exactly what makes the following true.

Fact If $(t, \tilde{\mathbf{x}})$ and $(T, \tilde{\mathbf{X}})$ are related by $(t, \tilde{\mathbf{x}})^t = A(T, \tilde{\mathbf{X}})^t + (t_0, \tilde{\mathbf{x}}_0)^t$ for $A \in O(1,3)$ then $-t^2 + |\tilde{\mathbf{x}}|^2 = T^2 + \left|\tilde{\mathbf{X}}\right|^2$.

A 4−vector is described relative to any inertial coordinates $(t, \vec{x})$ by a 4-tuple $v = (v^0, v^1, v^2, v^3)$ such that its description relative to $(T, \tilde{\mathbf{X}})$ as above is given by

$$\mathsf{V}^t = A\mathsf{v}^t \text{ (contravariant).}$$

Notice that we are using superscripts to index the components of a vector (and are not to be confused with exponents). This due to the following convention: vectors written with components up are called contravariant vectors while those with indices down are called covariant. Contravariant and covariant vectors transform differently and in such a way that the contraction of a contravariant with a covariant vector produces a quantity that is the same in any inertial coordinate system. To change a contravariant vector to its associated covariant form one uses the matrix η introduced above which is called the Lorentz metric tensor. Thus $(v^0, v^1, v^2, v^3)\eta = (-v^0, v^1, v^2, v^3) := (v_0, v_1, v_2, v_3)$ and thus the pseudo-length $v_i v^i = -(v^0)^2 + (v^1)^2 + (v^2)^2 + (v^3)^2$ is an invariant with respect to coordinate changes via the Lorentz group or even the Poincaré group. Notice also that $v_i v^i$ actually means $\sum_{i=0}^{3} v_i v^i$ which is in turn the same thing as

$$\sum \eta_{ij} v^i v^j$$

The so called **Einstein summation convention** say that when an index is repeated once up and once down as in $v_i v^i$, then the summation is implied.

Minkowski Space

One can see from the above that lurking in the background is an inner product space structure: If we fix the origin of space time then we have a vector space with scalar product $\langle v, v\rangle = v_i v^i = -(v^0)^2 + (v^1)^2 + (v^2)^2 + (v^3)^2$. This scalar product space (indefinite inner product!) is called Minkowski space. The scalar product just defined is the called the **Minkowski metric** or **the Lorentz metric.**.

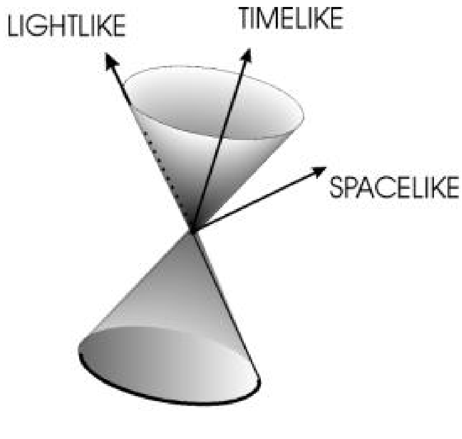
Definition L.1 *A 4-vector v is called **space-like** if and only if $\langle v, v\rangle > 0$, **time-like** if and only if $\langle v, v\rangle < 0$ and **light-like** if and only if $\langle v, v\rangle = 0$. The set of all light-like vectors at a point in Minkowski space form a double cone in $\mathbb{R}^4$ referred to as the **light cone**.*

Remark L.1 (Warning) *Sometimes the definition of the Lorentz metric given is opposite in sign from the one we use here. Both choices of sign are popular. One consequence of the other choice is that time-like vectors become those for which $\langle v, v\rangle > 0$.*

Definition L.2 *At each point of $x \in \mathbb{R}^4$ there is a set of vectors parallel to the 4-axes of $\mathbb{R}^4$. We will denote these by $\partial_0, \partial_1, \partial_2$,and ∂_3 (suppressing the point at which they are based).*

Definition L.3 *A vector v based at a point in $\mathbb{R}^4$ such that $\langle \partial_0, v\rangle < 0$ will be called **future pointing** and the set of all such forms the interior of the "future" light-cone.*

One example of a 4-vector is the momentum 4-vector written $\mathsf{p} = (E, \overrightarrow{\mathbf{p}})$ which we will define below. We describe the motion of a particle by referring



to its career in space-time. This is called its world-line and if we write c for the speed of light and use coordinates $(x^0, x^1, x^2, x^3) = (ct, x, y, z)$ then a world line is a curve $\gamma(s) = (x^0(s), x^1(s), x^2(s), x^3(s))$ for some parameter. The momentum 4-vector is then $\mathsf{p} = mc\mathsf{u}$ where u is the unite vector in the direction of the 4-velocity $\gamma'(s)$. The action functional for a free particle in Relativistic mechanics is invariant with respect to Lorentz transformations described above. In the case of a free particle of mass m it is

$$A_L = -\int_{s_1}^{s_2} mc\, |\langle\dot{\gamma}(s), \dot{\gamma}(s)\rangle|^{1/2}\, ds.$$

The quantity c is a constant equal to the speed of light in any inertial coordinate system. Here we see the need the absolute value since for a timelike path to assume that $\langle\dot{\gamma}(s), \dot{\gamma}(s)\rangle \leq 0$ by our convention. Define

$$\tau(s) = \int_{s_1}^{s_2} |\langle\dot{\gamma}(s), \dot{\gamma}(s)\rangle|^{1/2}\, ds$$

Then $\frac{d}{ds}\tau(s) = cm\, |\langle\dot{\gamma}(s), \dot{\gamma}(s)\rangle|^{1/2} \geq 0$ so we can reparameterize by τ:

$$\gamma(\tau) = (x^0(\tau), x^1(\tau), x^2(\tau), x^3(\tau)).$$

The parameter τ is called the proper time of the particle in question. A stationary curve will be a geodesic or straight line in $\mathbb{R}^4$.

Let us return to the Lorentz group. The group of rotations generated by rotations around the spatial x, y, and z-axes are a copy of $SO(3)$ sitting inside $SO(1,3)$ and consists of matrices of the form

$$\begin{bmatrix} 1 & 0 \\ 0 & R \end{bmatrix}$$

where $R \in SO(3)$. Now a rotation of the x, y-plane about the z, t-plane[1] for example has the form

$$\begin{pmatrix} cT \\ X \\ Y \\ Z \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) & 0 \\ 0 & \sin(\theta) & \cos(\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}$$

where as a Lorentz "rotation" of the $t, x-$plane about the $y, z-$plane has the form

$$\begin{pmatrix} cT \\ X \\ Y \\ Z \end{pmatrix} = \left[\begin{matrix} \cosh(\beta) & \sinh(\beta) & 0 & 0 \\ \sinh(\beta) & \cosh(\beta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \right]$$

$= [cT, X, Y, Z] = [(\cosh \beta)\, ct + (\sinh \beta)\, x, (\sinh \beta)\, ct + (\cosh \beta)\, x, y, z]$ Here the parameter β is usually taken to be the real number given by

$$\tanh \beta = v/c$$

where v is a velocity indicating that in the new coordinates the observer in travelling at a velocity of magnitude v in the x direction as compared to an observer in the original before the transformation. Indeed we have for an observer motionless at the spatial origin the T, X, Y, Z system the observers path is given in T, X, Y, Z coordinates as $T \mapsto (T, 0, 0, 0)$

$$\begin{aligned} \frac{dX}{dt} &= c \sinh \beta \\ \frac{dX}{dx} &= \cosh \beta \\ v = \frac{dx}{dt} &= \frac{\sinh \beta}{\cosh \beta} = c \tanh \beta. \end{aligned}$$

Calculating similarly, and using $\frac{dY}{dt} = \frac{dZ}{dt} = 0$ we are lead to $\frac{dy}{dt} = \frac{dz}{dt} = 0$. So the t, x, y, z observer sees the other observer (and hence his frame) as moving in the x direction at speed $v = c \tanh \beta$. The transformation above is called a **Lorentz boost** in the x-direction.

Now as a curve parameterized by the parameter τ (the proper time) the 4-**momentum** is the vector

$$\mathsf{p} = mc \frac{d}{d\tau} \mathsf{x}(\tau).$$

In a specific inertial (Lorentz) frame

$$\mathsf{p}(t) = \frac{dt}{d\tau} \frac{d}{dt} mc\mathsf{x}(t) = \left(\frac{mc^2}{c\sqrt{1-(v/c)^2}}, \frac{mc\dot{x}}{\sqrt{1-(v/c)^2}}, \frac{mc\dot{y}}{\sqrt{1-(v/c)^2}}, \frac{mc\dot{z}}{\sqrt{1-(v/c)^2}} \right)$$

[1] It is the z,t-plane rather than just the z-axis since we are in four dimensions and both the z-axis and the t-axis would remain fixed.

which we abbreviate to $\mathsf{p}(t) = (E/c, \overrightarrow{\mathbf{p}})$ where $\overrightarrow{\mathbf{p}}$ is the 3-vector given by the last there components above. Notice that

$$m^2c^2 = \langle \mathsf{p}(t), \mathsf{p}(t) \rangle = E^2/c^2 + \left|\overrightarrow{\mathbf{p}}\right|^2$$

is an invariant quantity but the pieces E^2/c^2 and $\left|\overrightarrow{\mathbf{p}}\right|^2$ are dependent on the choice of inertial frame.

What is the energy of a moving particle (or tiny observer?) in this theory? We claim it is the quantity E just introduced. Well, if the Lagrangian is any guide we should have from the point of view of inertial coordinates and for a particle moving at speed v in the positive $x-$direction

$$\begin{aligned} E &= v\frac{\partial L}{\partial v} = v\frac{\partial}{\partial v}\left(-mc^2\sqrt{1-(v/c)^2}\right) - (-mc^2\sqrt{1-(v/c)^2}) \\ &= m\frac{c^3}{\sqrt{(c^2-v^2)}} = m\frac{c^2}{\sqrt{(1-(v/c)^2)}}. \end{aligned}$$

Expanding in powers of the dimensionless quantity v/c we have $E = mc^2 + \frac{1}{2}mv^2 + O\left((v/c)^4\right)$. Now the term $\frac{1}{2}mv^2$ is just the nonrelativistic expression for kinetic energy. What about the mc^2? If we take the Lagrangian approach seriously, this must be included as some sort of energy. Now if v had been zero then we would still have a "rest energy" of mc^2! This is interpreted as the energy possessed by the particle by virtue of its mass. A sort of energy of being as it were. Thus we have $v = 0$ here and the famous equation for the equivalence of mass and energy follows:

$$E = mc^2.$$

If the particle is moving then mc^2 is only part of the energy but we can define $E_0 = mc^2$ as the "rest energy". Notice however, that although the length of the momentum 4-vector $m\dot{\mathsf{x}}$=p is always m;

$$\left|\langle \mathsf{p}, \mathsf{p} \rangle\right|^{1/2} = m\left|-(dt/d\tau)^2 + (dx/d\tau)^2 + (dy/d\tau)^2 + (dz/d\tau)^2\right|^{1/2} = m$$

and is therefore conserved in the sense of being constant one must be sure to remember that the mass of a body consisting of many particles is *not* the sum of the individual particle masses.

L.2 Electricity and Magnetism

Up until now it has been mysterious how any object of matter could influence any other. It turns out that most of the forces we experience as middle sized objects pushing and pulling on each other is due to a single electromagnetic force. Without the help of special relativity there appears to be two forces; electric and magnetic. Elementary particles that carry electric charges such as electrons or protons, exert forces on each other by means of a field. In a particular Lorentz frame, the electromagnetic field is described by a skew-symmetric matrix of functions called the electromagnetic field tensor:

$$(F_{\mu\nu}) = \begin{bmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & B_z & -B_y \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{bmatrix}.$$

Where we also have the forms $F_\mu^\nu = \eta^{s\nu} F_{\mu s}$ and $F^{\mu\nu} = \eta^{s\mu} F_s^\nu$. This tensor can be derived from a potential $\mathsf{A} = (A_0, A_1, A_2, A_3)$ by $F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}$. The contravariant form of the potential is $(A_0, -A_1, -A_2, -A_3)$ is a four vector often written as

$$\mathsf{A} = (\phi, \overrightarrow{\mathbf{A}}).$$

The action for a charged particle in an electromagnetic field is written in terms of A in a manifestly invariant way as

$$\int_a^b -mc d\tau - \frac{e}{c} A_\mu dx^\mu$$

so writing $\mathsf{A} = (\phi, \overrightarrow{\mathbf{A}})$ we have

$$S(\tilde{\mathbf{x}}) = \int_a^b (-mc\frac{d\tau}{dt} - e\phi(t) + \overrightarrow{\mathbf{A}} \cdot \frac{d\tilde{\mathbf{x}}}{dt})dt$$

so in a given frame the Lagrangian is

$$L(\tilde{\mathbf{x}}, \frac{d\tilde{\mathbf{x}}}{dt}, t) = -mc^2\sqrt{1 - (v/c)^2} - e\phi(t) + \overrightarrow{\mathbf{A}} \cdot \frac{d\tilde{\mathbf{x}}}{dt}.$$

Remark L.2 *The system under study is that of a particle in a field and does not describe the dynamics of the field itself. For that we would need more terms in the Lagrangian.*

This is a time dependent Lagrangian because of the $\phi(t)$ term but it turns out that one can re-choose A so that the new $\phi(t)$ is zero and yet still have $F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}$. This is called change of gauge. Unfortunately, if we wish to express things in such a way that a constant field is given by a constant potential then we cannot make this choice. In any case, we have

$$L(\overrightarrow{\mathbf{x}}, \frac{d\overrightarrow{\mathbf{x}}}{dt}, t) = -mc^2\sqrt{1 - (v/c)^2} - e\phi + \overrightarrow{\mathbf{A}} \cdot \frac{d\overrightarrow{\mathbf{x}}}{dt}$$

and setting $\overrightarrow{\mathbf{v}} = \frac{d\tilde{\mathbf{x}}}{dt}$ and $\left|\overrightarrow{\mathbf{v}}\right| = v$ we get the follow form for energy

$$\overrightarrow{\mathbf{v}} \cdot \frac{\partial}{\partial \overrightarrow{\mathbf{v}}} L(\tilde{\mathbf{x}}, \overrightarrow{\mathbf{v}}, t) - L(\tilde{\mathbf{x}}, \overrightarrow{\mathbf{v}}, t) = \frac{mc^2}{\sqrt{1 - (v/c)^2}} + e\phi.$$

Now this is not constant with respect to time because $\frac{\partial L}{\partial t}$ is not identically zero. On the other hand, this make sense from another point of view; the particle is interacting with the field and may be picking up energy from the field.

The Euler-Lagrange equations of motion turn out to be

$$\frac{d\tilde{\mathbf{p}}}{dt} = e\tilde{\mathbf{E}} + \frac{e}{c}\overrightarrow{\mathbf{v}} \times \tilde{\mathbf{B}}$$

where $\tilde{\mathbf{E}} = -\frac{1}{c}\frac{\partial \tilde{\mathbf{A}}}{\partial t} - \operatorname{grad}\phi$ and $\tilde{\mathbf{B}} = \operatorname{curl}\tilde{\mathbf{A}}$ are the electric and magnetic parts of the field respectively. This decomposition into electric and magnetic parts is an artifact of the choice of inertial frame and may be different in a different frame. Now the momentum $\tilde{\mathbf{p}}$ is $\frac{m\overrightarrow{\mathbf{v}}}{\sqrt{1-(v/c)^2}}$ but a speeds $v << c$ this becomes nearly equal to $m\mathbf{v}$ so the equations of motion of a charged particle reduce to

$$m\frac{d\overrightarrow{\mathbf{v}}}{dt} = e\tilde{\mathbf{E}} + \frac{e}{c}\overrightarrow{\mathbf{v}} \times \tilde{\mathbf{B}}.$$

Notice that is the particle is not moving, or if it is moving parallel the magnetic field $\tilde{\mathbf{B}}$ then the second term on the right vanishes.

The electromagnetic field equations.

We have defined the 3-vectors $\tilde{\mathbf{E}} = -\frac{1}{c}\frac{\partial \tilde{\mathbf{A}}}{\partial t} - \operatorname{grad}\phi$ and $\tilde{\mathbf{B}} = \operatorname{curl}\tilde{\mathbf{A}}$ but since the curl of a gradient we see that $\operatorname{curl}\tilde{\mathbf{E}} = -\frac{1}{c}\frac{\partial \tilde{\mathbf{B}}}{\partial t}$. Also, from $\tilde{\mathbf{B}} = \operatorname{curl}\tilde{\mathbf{A}}$ we get $\operatorname{div}\tilde{\mathbf{B}} = \mathbf{0}$. This easily derived pair of equations is the first two of the four famous Maxwell's equations. Later we will see that the electromagnetic field tensor is really a differential 2-form F and these two equations reduce to the statement that the (exterior) derivative of F is zero:

$$dF = 0$$

Exercise L.1 *Apply Gauss' theorem and Stokes' theorem to the first two Maxwell's equations to get the integral forms of the equations. What do these equations say physically?*

One thing to notice is that these two equations do not determine $\frac{\partial}{\partial t}\tilde{\mathbf{E}}$.

Now we have not really written down a action or Lagrangian that includes terms that represent the field itself. When that part of the action is added in we get

$$S = \int_a^b (-mc - \frac{e}{c}A_\mu \frac{dx^\mu}{d\tau})d\tau + a\int_V F^{\nu\mu}F_{\nu\mu}dx^4$$

where in so called Gaussian system of units the constant a turns out to be $\frac{-1}{16\pi c}$. Now in a particular Lorentz frame and recalling 20.5 we get $= a\int_V F^{\nu\mu}F_{\nu\mu}dx^4 = \frac{1}{8\pi}\int_V \left|\tilde{\mathbf{E}}\right|^2 - \left|\tilde{\mathbf{B}}\right|^2 dtdxdydz.$

In order to get a better picture in mind let us now assume that there is a continuum of charged particle moving through space and that volume density of charge at any given moment in space-time is ρ so that if $dxdydz = dV$ then ρdV is the charge in the volume dV. Now we introduce the four vector $\rho\mathsf{u} = \rho(d\mathsf{x}/d\tau)$ where u is the velocity 4-vector of the charge at (t, x, y, z). Now recall that $\rho d\mathsf{x}/d\tau = \frac{d\tau}{dt}(\rho, \rho\overrightarrow{\mathbf{v}}) = \frac{d\tau}{dt}(\rho, \tilde{\mathbf{j}}) = \mathsf{j}$. Here $\tilde{\mathbf{j}} = \rho\overrightarrow{\mathbf{v}}$ is the charge current density as viewed in the given frame a vector field varying smoothly from point to point. Write $\mathsf{j} = (j^0, j^1, j^2, j^3)$.

Assuming now that the particle motion is determined and replacing the discrete charge e be the density we have applying the variational principle with the region $U = [a, b] \times V$ says

$$\begin{aligned} 0 &= -\delta\left(\int_V \int_a^b \frac{\rho dV}{c} dV A_\mu \frac{dx^\mu}{d\tau} d\tau + a\int_U F^{\nu\mu} F_{\nu\mu} dx^4\right) \\ &= -\delta\left(\frac{1}{c}\int_U j^\mu A_\mu + aF^{\nu\mu}F_{\nu\mu}dx^4\right) \end{aligned}$$

Now the Euler-Lagrange equations become

$$\frac{\partial\mathcal{L}}{\partial A_\nu} - \partial_\mu \frac{\partial\mathcal{L}}{\partial(\partial_\mu A_\nu)} = 0$$

where $\mathcal{L}(A_\mu, \partial_\mu A_\eta) = \frac{\rho}{c} A_\mu \frac{dx^\mu}{dt} + aF^{\nu\mu}F_{\nu\mu}$ and $F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}$. If one is careful to remember that $\partial_\mu A_\nu = \frac{\partial A_\nu}{\partial x^\mu}$ is to be treated as an independent variable one can arrive at some complicated looking equations and then looking at the matrix 20.5 we can convert the equations into statements about the fields $\tilde{\mathbf{E}}$, $\tilde{\mathbf{B}}$,and $(\rho, \tilde{\mathbf{j}})$. We will not carry this out since we instead discover a much more efficient formalism for dealing with the electromagnetic field. Namely, we will use differential forms and the Hodge star operator. At any rate the last two of Maxwell's equations read

$$\begin{aligned} \operatorname{curl} \tilde{\mathbf{B}} &= 0 \\ \operatorname{div} \tilde{\mathbf{E}} &= 4\pi\rho. \end{aligned}$$

L.2.1 Maxwell's equations.

$\mathbb{R}^{3,1}$ is just $\mathbb{R}^4$ but with the action of the symmetry group $O(1, 3)$.

Recall the electromagnetic field tensor

$$(F_{\mu\nu}) = \begin{bmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -B_z & B_y \\ -E_y & B_z & 0 & -B_x \\ -E_z & -B_y & B_x & 0 \end{bmatrix}.$$

Let us work in units where $c = 1$. Since this matrix is skew symmetric we can form a 2-form called the electromagnetic field 2-form:

$$F = \frac{1}{2}\sum_{\mu,\nu} F_{\mu\nu} dx^\mu \wedge dx^\nu = \sum_{\mu<\nu} F_{\mu\nu} dx^\mu \wedge dx^\nu.$$

Let write $E = E_x dx + E_y dy + E_z dz$ and $B = B_x dy \wedge dz + B_y dz \wedge dx + B_z dx \wedge dy$. One can check that we now have

$$F = B + E \wedge dt.$$

Now we know that F comes from a potential $A = A_\nu dx^\nu$. In fact, we have

$$\begin{aligned} dA = d(A_\nu dx^\nu) &= \sum_{\mu<\nu} (\frac{\partial}{\partial x^\mu} A_\nu - \frac{\partial}{\partial x^\nu} A_\mu) dx^\mu \wedge dx^\nu \\ &= \sum_{\mu<\nu} F_{\mu\nu} dx^\mu \wedge dx^\nu = F. \end{aligned}$$

Thus we automatically have $dF = ddA = 0$. Now what does $dF = 0$ translate into in terms of the E and B? We compute:

$$\begin{aligned} 0 &= dF = d(B + E \wedge dt) = dB + dE \wedge dt \\ &= (\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z}) + d\,(E_x dx + E_y dy + E_z dz) \wedge dt \\ &= (\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z}) dx \wedge dy \wedge dz + \frac{\partial B}{\partial t} \wedge dt + \\ &+ \left[(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}) dy \wedge dz + (\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x}) dz \wedge dx + (\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}) dy \wedge dx \right] \wedge dt \end{aligned}$$

From this we conclude that

$$\begin{aligned} \operatorname{div}(\tilde{\mathbf{B}}) &= 0 \\ \operatorname{curl}(\tilde{\mathbf{E}}) + \frac{\partial \tilde{\mathbf{B}}}{\partial t} &= 0 \end{aligned}$$

which is Maxwell's first two equations. Thus Maxwell's first two equations end up being equivalent to just the single equation

$$dF = 0$$

which was true just from the fact that $dd = 0$ since we assuming that there is a potential A! This equation does not involve the scalar product structure encoded by the matrix η.

As for the second pair of Maxwell's equations, they too combine to give a single equation. The appropriate star operator is given by

Definition L.4 *Define $\epsilon(\mu)$ to be entries of the diagonal matrix $\eta = diag(-1, 1, 1, 1)$. Let $*$ be defined on $\Omega^k(\mathbb{R}^4)$ by letting $*(dx^{i_1} \wedge \cdots \wedge dx^{i_k}) = \pm\epsilon(j_1)\epsilon(j_2)\cdots\epsilon(j_k) dx^{j_1} \wedge \cdots \wedge dx^{j_{n-k}}$ where $dx^{i_1} \wedge \cdots \wedge dx^{i_k} \wedge dx^{j_1} \wedge \cdots \wedge dx^{j_{n-k}} = \pm dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3$. (Choose the sign to that makes the last equation true and then the first is true by definition). Extend $*$ linearly to a map $\Omega^k(\mathbb{R}^4) \to \Omega^{4-k}(\mathbb{R}^4)$. More simply and explicitly*

Exercise L.2 *Show that $* \circ *$ acts on $\Omega^k(\mathbb{R}^4)$ by $(-1)^{k(4-k)+1}$.*

Exercise L.3 *Show that if F is the electromagnetic field tensor defined above then*

$$*F = (*F)_{\mu\nu}\, dx^\mu \wedge dx^\nu$$

where $(*F)_{\mu\nu}$ are the components of the matrix

$$(*F)_{\mu\nu} = \begin{bmatrix} 0 & B_x & B_y & B_z \\ -B_x & 0 & E_z & -E_y \\ -B_y & -E_z & 0 & E_x \\ -B_z & E_y & -E_x & 0 \end{bmatrix}$$

Now let J be the differential 1-form constructed from the 4-current $\mathsf{j} = (\rho, \tilde{\mathbf{j}})$ introduced in section F.0.8 by letting $(j_0, j_1, j_2, j_3) = (-\rho, \tilde{\mathbf{j}})$ and then setting $J = j_\mu dx^\mu$.

Now we add the second equation

$$*d*F = J.$$

Exercise L.4 *Show that the single differential form equation $*d*F = J$ is equivalent to Maxwell's second two equations*

$$\operatorname{curl}(\tilde{\mathbf{B}}) = \frac{\partial \tilde{\mathbf{E}}}{\partial t} + \tilde{\mathbf{j}}$$
$$\operatorname{div}(\tilde{\mathbf{E}}) = \rho.$$

In summary, we have that Maxwell's 4 equations (in free space) in the formalism of differential forms and the Hodge star operator are simply the pair

$$dF = 0$$
$$*d*F = J.$$

The first equation is equivalent to Maxwell's first two equations and interestingly does not involve the metric structure of space $\mathbb{R}^3$ or the metric structure of spacetime $\mathbb{R}^{1,3}$. The second equation above is equivalent to Maxwell's second two equations an through the star operator essentiality involves the Metric structure of $\mathbb{R}^{1,3}$.

Now an interesting thing happens if the Lagrangian density is invariant under some set of transformations. Suppose that $\delta\phi$ is an infinitesimal "internal" symmetry of the Lagrangian density so that $\delta S(\delta\phi) = 0$ even though $\delta\phi$ does not vanish on the boundary. Then if ϕ is already a solution of the field equations then

$$0 = \delta S = \int_U \partial_\mu \left(\delta\phi \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) d^4x$$

for all regions U. This means that $\partial_\mu \left(\delta\phi \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) = 0$ so if we define $j^\mu = \delta\phi \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}$ we get

$$\partial_\mu j^\mu = 0$$

or

$$\frac{\partial}{\partial t} j^0 = -\nabla \cdot \overrightarrow{\mathbf{j}}$$

where $\overrightarrow{\mathbf{j}} = (j^1, j^2, j^3)$ and $\nabla \cdot \overrightarrow{\mathbf{j}} = \operatorname{div}(\overrightarrow{\mathbf{j}})$ is the spatial divergence. This looks like some sort of conservation. Indeed, if we define the total charge at any time t by

$$Q(t) = \int j^0 d^3x$$

the assuming $\overrightarrow{\mathbf{j}}$ shrinks to zero at infinity then the divergence theorem gives

$$\begin{aligned} \frac{d}{dt} Q(t) &= \int \frac{\partial}{\partial t} j^0 d^3x \\ &= -\int \nabla \cdot \overrightarrow{\mathbf{j}}\, d^3x = 0 \end{aligned}$$

so the charge $Q(t)$ is a conserved quantity. Let $Q(U,t)$ denote the total charge inside a region U. The charge inside any region U can only change via a flux through the boundary:

$$\begin{aligned} \frac{d}{dt} Q(U,t) &= \int_U \frac{\partial}{\partial t} j^0 d^3x \\ &= \int_{\partial U} \overrightarrow{\mathbf{j}} \cdot \mathbf{n} dS \end{aligned}$$

which is a kind of "local conservation law". To be honest the above discussion only takes into account so called internal symmetries. An example of an internal symmetry is given by considering a curve of linear transformations of $\mathbb{R}^k$ given as matrices $C(s)$ with $C(0) = I$. Then we vary ϕ by $C(s)\phi$ so that $\delta\phi = \frac{d}{ds}\big|_0 C(s)\phi = C'(0)\phi$. Another possibility is to vary the underlying space so that $C(s, .)$ is now a curve of transformations of $\mathbb{R}^4$ so that if $\phi_s(x) = \phi(C(s,x))$ is a variation of fields then we must take into account the fact that the domain of integration is also varying:

$$L(\phi_s, \partial\phi_s) = \int_{U_s \subset \mathbb{R}^4} \mathcal{L}(\phi_s, \partial\phi_s) d^4x$$

We will make sense of this later.

	Global	local	Sheaf notation
functions on M	$C^\infty(M)$	$C^\infty(U)$	$C^\infty M$
Vector fields on M	$\mathfrak{X}(M)$	$\mathfrak{X}(U)$	$\mathfrak{X}_M$
Sections of E	$\Gamma(E)$	$\Gamma(U,E)$	$-_E$
Forms on M	$\Omega(M)$	$\Omega(U)$	Ω_M
Tensor fields on M	$\mathfrak{T}^k_l(M)$	$\mathfrak{T}^k_l(U)$	$\mathfrak{T}^k_l M$

Example L.1 *For matrix Lie groups there is a simple way to compute the Maurer-Cartan form. Let g denote the identity map $G \to G$ so that $dg : TG \to TG$ is also the identity map. Then*

$$\omega_G = g^{-1} dg.$$

For example, let $G = \mathrm{Euc}(n)$ presented as matrices of the form

$$\begin{bmatrix} 1 & 0 \\ x & Q \end{bmatrix} \text{ for } Q \in \mathrm{O}(n) \text{ and } x \in \mathbb{R}^n.$$

Then $g = \begin{bmatrix} 1 & 0 \\ x & Q \end{bmatrix}$ *and* $dg = d \begin{bmatrix} 1 & 0 \\ x & Q \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ dx & dQ \end{bmatrix}$.

$$\begin{aligned} g^{-1} dg &= \begin{bmatrix} 1 & 0 \\ -x & Q^t \end{bmatrix} \begin{bmatrix} 0 & 0 \\ dx & dQ \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ Q^t dx & dQ \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ (Q^t)_r dx^r & (Q^t)^{ik} dQ_{kj} \end{bmatrix} \end{aligned}$$

Index

Bibliography

[A] J. F. Adams, Stable Homotopy and Generalized Homology, Univ. of Chicago Press, 1974.

[Arm] M. A. Armstrong, Basic Topology, Springer-Verlag, 1983.

[At] M. F. Atiyah, K-Theory, W.A.Benjamin, 1967.

[A,B,R] Abraham, R., Marsden, J.E., and Ratiu, T., *Manifolds, tensor analysis, and applications*, Addison Wesley, Reading, 1983.

[Arn] Arnold, V.I., *Mathematical methods of classical mechanics*, Graduate Texts in Math. **60**, Springer-Verlag, New York, 2nd edition (1989).

[A] Alekseev, A.Y., *On Poisson actions of compact Lie groups on symplectic manifolds*, J. Diff. Geom. **45** (1997), 241-256.

[Bott and Tu]

[Bry]

[Ben] D. J. Benson, *Representations and Cohomology, Volume II: Cohomology of Groups and Modules,* Cambridge Univ. Press, 1992.

[1] R. Bott and L. Tu, *Differential Forms in Algebraic Topology*, Springer-Verlag GTM 82,1982.

[Bre] G. Bredon, *Topology and Geometry*, Springer-Verlag GTM 139, 1993.

[Chav1]

[Chav2]

[Drin] Drinfel'd, V.G., *On Poisson homogeneous spaces of Poisson-Lie groups*, Theor. Math. Phys. **95** (1993), 524-525.

[Dieu] J. Dieudonn´e, *A History of Algebraic and Differential Topology* 1900-1960, Birkh¨auser,1989.

[Do] A. Dold, *Lectures on Algebraic Topology*, Springer-Verlag, 1980.

[Dug] J. Dugundji, *Topology*, Allyn & Bacon, 1966.

[Eil,St] S. Eilenberg and N. Steenrod, *Foundations of Algebraic Topology*, Princeton Univ. Press, 1952.

[Fen] R. Fenn, *Techniques of Geometric Topology*, Cambridge Univ. Press, 1983.

[Fr,Q] M. Freedman and F. Quinn, *Topology of 4-Manifolds*, Princeton Univ. Press, 1990.

[Fult] W. Fulton, *Algebraic Topology: A First Course*, Springer-Verlag, 1995.

[G1] Guillemin, V., and Sternberg, S., *Convexity properties of the moment mapping*, Invent. Math. 67 (1982), 491-513.

[G2] Guillemin, V., and Sternberg, S., *Symplectic Techniques in Physics*, Cambridge Univ. Press, Cambridge, 1984.

[Gu,Hu,We] Guruprasad, K., Huebschmann, J., Jeffrey, L., and Weinstein, A., *Group systems, groupoids, and moduli spaces of parabolic bundles*, Duke Math. J. 89 (1997), 377-412.

[Gray] B. Gray, *Homotopy Theory*, Academic Press, 1975.

[Gre,Hrp] M. Greenberg and J. Harper, *Algebraic Topology: A First Course*, Addison-Wesley, 1981.

[2] P. J. Hilton, *An Introduction to Homotopy Theory*, Cambridge University Press, 1953.

[Hilt2] P. J. Hilton and U. Stammbach, *A Course in Homological Algebra*, Springer-Verlag, 1970.

[Huss] D. Husemoller, *Fibre Bundles*, McGraw-Hill, 1966 (later editions by Springer-Verlag).

[Hu] Huebschmann, J., *Poisson cohomology and quantization*, J. Reine Angew. Math. **408** (1990), 57-113.

[KM]

[Kirb,Seib] R. Kirby and L. Siebenmann, *Foundational Essays on Topological Manifolds, Smoothings, and Triangulations*, Ann. of Math.Studies 88, 1977.

[L1] Lang, S. *Foundations of Differential Geometry*, Springer-Verlag GTN vol 191

[M,T,W] Misner,C. Wheeler, J. and Thorne, K. *Gravitation*, Freeman 1974

[Mil] Milnor, J., *Morse Theory*, Annals of Mathematics Studies **51**, Princeton U. Press, Princeton, 1963.

[MacL] S. MacLane, *Categories for the Working Mathematician*, Springer-Verlag GTM 5, 1971.

[Mass] W. Massey, *Algebraic Topology: An Introduction*, Harcourt, Brace & World, 1967 (reprinted by Springer-Verlag).

[Mass2] W. Massey, *A Basic Course in Algebraic Topology*, Springer-Verlag, 1993.

[Maun] C. R. F. Maunder, *Algebraic Topology*, Cambridge Univ. Press, 1980 (reprinted by Dover Publications).

[Miln1] J. Milnor, *Topology from the Differentiable Viewpoint*, Univ. Press of Virginia, 1965.

[Mil,St] J. Milnor and J. Stasheff, *Characteristic Classes*, Ann. of Math. Studies 76, 1974.

[Roe] Roe,J. Elliptic Operators, *Topology and Asymptotic methods*, Longman, 1988

[Spv] Spivak, M. *A Comprehensive Introduction to Differential Geometry*, (5 volumes) Publish or Perish Press, 1979.

[St] Steenrod, N. *Topology of fiber bundles*, Princeton University Press, 1951.

[Va] Vaisman, I., *Lectures on the Geometry of Poisson Manifolds*, Birkhäuser, Basel, 1994.

[We1] Weinstein, A., *Lectures on Symplectic Manifolds*, Regional conference series in mathematics **29**, Amer. Math. Soc.,Providence,1977.

[We2] Weinstein, A., *The local structure of Poisson manifolds*, J. Diff. Geom. **18** (1983), 523–557.

[We3] Weinstein, A., *Poisson structures and Lie algebras*, Astérisque, hors série (1985), 421–434.

[3] J. A. Álvarez López, *The basic component of the mean curvature of Riemannian foliations,* Ann. Global Anal. Geom. **10**(1992), 179–194.

[BishCr] R. L. Bishop and R. J. Crittenden, *Geometry of Manifolds*, New York:Academic Press, 1964.

[Chavel] I. Chavel, *Eigenvalues in Riemannian Geometry,* Orlando: Academic Press, 1984.

[Cheeger] J. Cheeger, *A lower bound for the smallest eigenvalue of the Laplacian*, Problems in analysis (Papers dedicated to Salomon Bochner, 1969), pp. 195–199, Princeton, N. J.: Princeton Univ. Press, 1970.

[ChEbin] J. Cheeger and D. Ebin, *Comparison Theorems in Riemannian Geometry*, Amsterdam: North-Holland, 1975.

[Cheng] S.Y. Cheng, *Eigenvalue Comparison Theorems and its Geometric Applications* Math. Z. **143**, 289–297, (1975).

[Chern1] S.S. Chern, "*What is geometry*", The American Mathematical Monthly, **97**, no. 8 (1990) 679-686.

[El-KacimiHector] A. El Kacimi–Alaoui and G. Hector, *Décomposition de Hodge basique pour un feuilletage Riemannien,* Ann. Inst. Fourier, Grenoble **36**(1986), no. 3 , 207–227.

[Gray2] A. Gray *Comparison Theorems for the Volumes of Tubes as Generalizations of the Weyl Tube Formula* Topology **21, no. 2**, 201–228, (1982).

[HeKa] E. Heintz and H. Karcher, *A General Comparison Theorem with Applications to Volume estimates for Submanifolds* Ann. scient. Èc. Norm Sup., 4^e sèrie t. **11**, 451–470, (1978).

[KamberTondeur] F. W. Kamber and Ph. Tondeur, *De Rham-Hodge theory for Riemannian foliations*, Math. Ann. **277**(1987), 415–431.

[Lee] J. Lee, *Eigenvalue Comparison for Tubular Domains* Proc. of the Amer. Math. Soc. **109 no. 3**(1990).

[Min-OoRuhTondeur] M. Min-Oo, E. A. Ruh, and Ph. Tondeur, *Vanishing theorems for the basic cohomology of Riemannian foliations*, J. reine angew. Math. **415**(1991), 167–174.

[Molino] P. Molino, *Riemannian foliations,* Progress in Mathematics, Boston: Birkhauser, 1988.

[NishTondeurVanh] S. Nishikawa, M. Ramachandran, and Ph. Tondeur, *The heat equation for Riemannian foliations,* Trans. Amer. Math. Soc. **319**(1990), 619–630.

[O'Neill] B. O'Neill, *Semi-Riemannian Geometry*, New York:Academic Press, 1983.

[PaRi] E. Park and K. Richardson, *The basic Laplacian of a Riemannian foliation*, Amer. J. Math. **118**(1996), no. 6, pp. 1249–1275.

[Ri1] K. Richardson, *The asymptotics of heat kernels on Riemannian foliations*, to appear in Geom. Funct. Anal.

[Ri2] K. Richardson, *Traces of heat kernels on Riemannian foliations*, preprint.

[Steenrod] N. Steenrod, *The topology of fibre bundles*, Princeton University Press, Princeton, New Jersey, 1951.

[Tondeur1] Ph. Tondeur, *Foliations on Riemannian manifolds,* New York:Springer Verlag, 1988.

[Tondeur2] Ph. Tondeur, *Geometry of Foliations*, Monographs in Mathematics, vol. 90, Basel: Birkhäuser, 1997.

[NRC] *GravitationalPhysics-ExploringtheStructureofSpaceandTime.* National Academy Press 1999.