INSTITUTE OF PHYSICS

SERIES IN HIGH ENERGY PHYSICS, COSMOLOGY AND GRAVITATION

GEOMETRY AND PHYSICS OF BRANES

EDITED BY

U BRUZZO

V GORINI

U MOSCHELLA

IoP

GEOMETRY AND PHYSICS OF BRANES

Series in High Energy Physics, Cosmology and Gravitation

Other books in the series

Electron–Positron Physics at the Z
M G Green, S L Lloyd, P N Ratoff and D R Ward

Non-accelerator Particle Physics
Paperback edition
H V Klapdor-Kleingrothaus and A Staudt

Ideas and Methods of Supersymmetry and Supergravity
***or* A Walk Through Superspace**
Revised edition
I L Buchbinder and S M Kuzenko

Pulsars as Astrophysical Laboratories for Nuclear and Particle Physics
F Weber

Classical and Quantum Black Holes
Edited by P Fré, V Gorini, G Magli and U Moschella

Particle Astrophysics
Revised paperback edition
H V Klapdor-Kleingrothaus and K Zuber

The World in Eleven Dimensions
Supergravity, Supermembranes and M-Theory
Edited by M J Duff

Gravitational Waves
Edited by I Ciufolini, V Gorini, U Moschella and P Fré

Modern Cosmology
Edited by S Bonometto, V Gorini and U Moschella

The Mathematical Theory of Cosmic Strings
Cosmic Strings in the Wire Approximation
M R Anderson

The Galactic Black Hole
Lectures on General Relativity and Astrophysics
H Falcke and F W Hehl

GEOMETRY AND PHYSICS OF BRANES

Edited by

Ugo Bruzzo

Sissa, Trieste, Italy

Vittorio Gorini and Ugo Moschella

*Department of Chemical, Mathematical and Physical Sciences,
University of Insubria at Como, Italy*

IoP

INSTITUTE OF PHYSICS PUBLISHING
BRISTOL AND PHILADELPHIA

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

ISBN 0 7503 0863 X

Library of Congress Cataloging-in-Publication Data are available

Commissioning Editor: James Revill
Production Editor: Simon Laurenson
Production Control: Sarah Plenty
Cover Design: Victoria Le Billon
Marketing: Nicola Newey and Verity Cooke

Published by Institute of Physics Publishing, wholly owned by The Institute of Physics, London

Institute of Physics Publishing, Dirac House, Temple Back, Bristol BS1 6BE, UK

US Office: Institute of Physics Publishing, The Public Ledger Building, Suite 929, 150 South Independence Mall West, Philadelphia, PA 19106, USA

Typeset in LaTeX 2$_\varepsilon$ by Text 2 Text, Torquay, Devon
Printed in the UK by MPG Books Ltd, Bodmin, Cornwall

Contents

Preface

This book brings together the contents of the courses given at the doctoral school on ‘Geometry and Physics of Branes’ which took place in the spring of 2001 at the Centre for Scientific Culture ‘Alessandro Volta’ located in the beautiful environment of Villa Olmo in Como, Italy. The school was the result of a twinning between the Graduate School in Contemporary Relativity and Gravitational Physics, which is organized yearly by SIGRAV-Societa’ Italiana di Relativita’ e Gravitazione (Italian Society of Relativity and Gravitation), and the School on Algebraic Geometry and Physics organized every year (in alternation with a Workshop on the same subject) by the Mathematical Physics Group of the International School for Advanced Studies (SISSA-ISAS) in Trieste.

The central topic of the school was the concept of the brane in string theory, from both physical and mathematical viewpoints. Rather than attempting to make a (forcefully superficial) general overview of the mathematics and physics of branes, the philosophy underlying the choice of lectures was to provide an introduction to some lines of research, related to the notion of branes in string theory, which are presently the object of strong interest in the mathematical and physical communities.

Qualitatively, a brane is a state of string theory which corresponds to an extended solitonic configuration of the string theory. Sometimes these can be related to classical solutions of the low-energy limit of the string theory (which is a supergravity theory) which are charged with respect to some gauge potential. However, in other situations (technically, when the branes have charges in the Ramond–Ramond sector) these classical solutions describe membranes over which the open strings terminate. These are the D-branes. The contribution by A Lerda (An elementary introduction to branes in string theory) is a remarkably lucid introduction to these notions.

The discovery of open unoriented string models in the late 1980s prompted an interest in conformal field theory on open and unoriented surfaces. Another source of interest in such theories comes from two-dimensional quantum field theory in the presence of a boundary. The article by Y S Stanev (Two-dimensional conformal field theory on open and unoriented surfaces) develops the basics of this theory. The emphasis is on the construction of the correlation functions and partition functions.

The contribution by C Gómez and P Resco (Topics in string tachyon dynamics) concerns the role of tachyons in string theory, in particular the emergence of the so-called tachyon condensation phenomenon in several situations. Topics touched upon include tachyon condensation in open-string theory, its contribution to the value of the cosmological constant in closed string theory, its relevance to the study of the confinement problem for the gauge degrees of freedom, its connection with the bound states of a brane–antibrane system and a possible description in terms of K-theory.

Mirror symmetry has motivated the huge interest of mathematicians in string theory. The solitonic states of type IIB string theory correspond to 3-branes which can be described as special Lagrangian submanifolds of the compactification (Calabi–Yau) manifold X carrying a $U(1)$ bundle. With these geometric data, by means of the Floer cohomology of X regarded as a symplectic manifold, one constructs an A^{∞} category, the so-called Fukaya category of X. The dual type IIA string theory admits brane configurations which are complex submanifolds of the compatification space Y supporting stable bundles. In this case the category naturally attached to these data is the category of coeherent sheaves on Y or, rather, an A^{∞}-deformation of it. Kontsevitch has conjectured that there is an equivalence, in some proper sense, between the two categories. Fukaya's paper (Deformation theory, homological algebra, and mirror symmetry) fits within the author's ambitious programme to build a comprehensive mathematical setting to study this conjecture and is about a homology theory naturally attached to the deformations of vector bundles.

The contribution by A Grassi and M Rossi (Large N dualities and transitions in geometry) is about the so-called Gopakumar–Vafa conjecture and a possible strategy to prove it. After the 't Hooft proposal, according to which for large N there should be some duality between $SU(N)$ gauge theory and closed-string theory, Gopakumar and Vafa conjectured a duality between the $SU(N)$ Chern–Simons theory on S^3 and a IIA string theory compactified on a Calabi–Yau three-fold Y; the geometric relation between the two theories is that Y may be subjected to a procedure which makes it into T^*S^3. A possible way to prove this duality is to consider M-theory compactified on a manifold with special (G_2) holonomy

The School was made possible by funding from several sources, including the International School for Advanced Studies in Trieste, the University of Insubria (Como-Varese), the Department of Chemistry, Physics and Mathematics of the same University and the Physics Departments of the Universities of Milan, Pavia and Turin. We are grateful to the other members of the scientific organizing committee Mauro Carfora, Pietro Fre', Alberto Lerda and Augusto Sagnotti and to the scientific coordinator of the Centro Volta, Giulio Casati, for their invaluable help in the organization. We also acknowledge the essential support of the secretarial conferece staff of the Centro Volta, in particular of Chiara Stefanetti.

Ugo Bruzzo, Vittorio Gorini and Ugo Moschella
15 May 2002

PART 1

AN ELEMENTARY INTRODUCTION TO BRANES IN STRING THEORY

Alberto Lerda

This page intentionally left blank

Chapter 1

Introduction

Alberto Lerda
Dipartimento di Scienze e Tecnologie Avanzate
Università del Piemonte Orientale 'A Avogadro'
I-15100 Alessandria (Italy)

In recent years there has been a remarkable improvement in our understanding of string theory. One of the key ingredients of this progress has been the concept of duality [1], originally formulated for the supersymmetric gauge field theories and later extended to string theory [2]. Among other things, the idea of duality has led to the conclusion that the five consistent and perturbatively inequivalent superstring theories in ten dimensions are actually related to one another by non-perturbative maps. As a consequence of these relations, the five superstrings can be interpreted as five different perturbative expansions of a single underlying theory, called M-theory [3]. This M-theory, whose intrinsic fundamental formulation is not yet known, also admits another perturbative limit where it becomes a unique supergravity model in 11 dimensions. In this way a very tight and fruitful relationship between string theory and supergravity has been established which has increasingly led to very interesting (and largely unexpected) developments.

Under the action of the so-called string duality groups, a discrete version of the continuous duality groups already known in supergravity, the usual perturbative string states are mapped into solitonic configurations which represent extended objects with p spatial dimensions. These can be particles ($p = 0$), strings ($p = 1$), membranes ($p = 2$) or, in general, p-branes. Thus, we can legitimately say that modern string theory is not only a theory of strings! In fact, the existence of p-dimensional extended objects is required in order to provide the degrees of freedom needed by the non-perturbative string dualities. However, from a supergravity point of view, p-branes naturally appear as classical solutions of the various low-energy string effective actions that carry a non-vanishing charge with respect to some $(p + 1)$-form gauge potential. Thus, it is tempting to identify these supergravity p-branes with the configurations required by string dualities. For this reason in recent years much attention has been devoted to the study of these supergravity branes and their properties. The simplest of them are discussed in detail in [4] where one can also find the references to the original papers. The classical solutions with a non-vanishing electric or magnetic charge under the Neveu-Schwarz–Neveu-Schwarz (NS–NS) 2-form correspond, respectively, to the fundamental string and the solitonic 5-brane or, in the dual formulation, to the solitonic string and the fundamental 5-brane. In contrast, classical solutions with a non-vanishing charge under the various $(p + 1)$-forms of the Ramond–Ramond (R–R) sector have no relation with the perturbative closed string or its solitons. In fact, as recognized by J Polchinski [5], these solutions correspond to membranes on which open strings can end, with Dirichlet boundary conditions in the transverse directions and the usual Neumann boundary conditions in the longitudinal directions. For this reason they are called Dirichlet branes or D-branes for short (extensive reviews on D-branes are listed in [6]).

It turns out that the tension of these D-branes is proportional to the *inverse* of the string coupling constant; thus they are non-perturbative configurations of string theory which, however, can be studied in a very explicit way thanks to their description in terms of open strings with Dirichlet boundary conditions. For example, the interaction between two such D-branes can be computed by evaluating a one-loop open-string annulus diagram. However, since the early days of string theory it has been known that an annulus diagram of open strings can be equivalently rewritten as a tree-level cylinder diagram in a closed string theory where a closed string is generated from the vacuum, propagates and then annihilates again in the vacuum. The state that describes the emission (or absorption) of a closed string from the vacuum is called a *boundary state*; and it was originally introduced in the early days of dual models [7] to factorize the planar and non-planar open string diagrams at one loop in the closed string channel. In the mid-1980s, when the BRST formulation of string theory was developed, the boundary state was again considered in a series of papers by Callan *et al* [8] where, among other things, the ghost contribution was added and the generalization with an Abelian external gauge field was constructed. The extension of the boundary state to the case of Dirichlet boundary conditions was initiated in another series of papers by Green *et al* [9] in the early 1990s, before it became clear that these Dirichlet configurations are associated with the configurations required by the string dualities. More recently, the boundary state has been extensively used to describe the properties and interactions of the D-branes, both in flat and in curved backgrounds (see, for example, [10–12] or the reviews in [13] and the references therein). In particular, in [10, 12] it has been shown that the boundary state encodes all relevant properties of the classical D-branes since it correctly reproduces the couplings of the Dirac–Born–Infeld action as well as the large-distance behaviour of the classical Dp-brane supergravity solutions.

We would like to emphasize that the twofold interpretation of the D-branes, as classical supergravity solutions and as spacetime defects where open strings can terminate, is a direct consequence of a duality between open and closed strings which allows a double interpretation of the annulus/cylinder diagram. This twofold nature of the D-branes is their most important and intriguing feature; indeed because of this they play a crucial role both from a gravitational point of view (i.e. in a theory of closed strings) and from a gauge field theory point of view (i.e. in a theory of open strings). This open/closed string duality is at the heart of the gauge/gravity correspondence which has recently been uncovered since Maldacena's well-known conjecture [14–16] and which is perhaps one of the most exciting developments of string theory.

In this contribution we are going to present an elementary introduction to the branes of string theory and, in particular, to the boundary state description of the D-branes. These lecture notes are not intended to be an exhaustive presentation but rather their aim is merely to provide some very basic material that may serve as a background for more advanced topics in brane theory (for more extended and complete reviews see, for example, [4, 6, 13, 16, 17]). In particular, in chapter 2 we will review the supergravity effective actions of type II string theories, the classical field equations that follow from these actions and the simplest brane solutions. In chapter 3 we review the boundary state formalism to describe Dirichlet branes and discuss the case in which an external field is present on their world-volume. In chapter 4, using the boundary state we discuss the D-brane effective action and finally, in chapter 5, we show how supergravity classical D-brane solutions can be recovered from the boundary state.

Chapter 2

Branes in string theory

In this chapter we are going to present explicitly the simplest brane configurations of string theory. In particular, we will discuss the fundamental string solution (or F1), the solitonic Neveu-Schwarz 5-brane solution (or NS5) and the so-called Dp-branes in a flat ten-dimensional spacetime. We will not discuss branes in curved backgrounds and we will limit our considerations to the branes of type II string theories. A more extensive and complete discussion can be found, for example, in [4, 17, 18].

2.1 The superstring effective actions of type II

The various branes in which we are interested are classical solutions of the field equations that arise from the low-energy string effective actions of type II. Type I string theories can be divided into two: type IIA and type IIB. Both are defined in ten dimensions and have 32 real supercharges corresponding to $\mathcal{N} = 2$ supersymmetry in $d = 10$. In type IIA the two supersymmetries have opposite chirality, while in type IIB they have the same chirality.

2.1.1 Type IIA

The massless bosonic content of the type IIA string theory consists of a graviton $G_{\mu\nu}$,[1] an antisymmetric two-index tensor $B^{(2)}_{\mu\nu}$ (also called the Kalb–Ramond field) and a dilaton ϕ from the Neveu-Schwarz–Neveu-Schwarz sector (NS–NS), a vector $C^{(1)}_{\mu}$ and an antisymmetric three-index tensor $C^{(3)}_{\mu\nu\rho}$ from the Ramond–Ramond sector (R–R). These fields correspond to a total of 128 physical degrees of freedom, of which 35 are associated with the graviton, 28 with the two-form

[1] Our conventions for indices, forms and Hodge duals are the following: $\mu, \nu, \dots = 0, \dots, 9$, signature $(-, +^9)$, $\varepsilon^{0\dots9} = -\varepsilon_{0\dots9} = +1$, $\omega^{(n)} = \frac{1}{n!}\omega_{\mu_1\dots\mu_n}\, \mathrm{d}x^{\mu_1} \wedge \dots \wedge \mathrm{d}x^{\mu_n}$, and $*\omega^{(n)} = \frac{\sqrt{-\det G}}{n!(10-n)!}\varepsilon_{\nu_1\dots\nu_{10-n}\mu_1\dots\mu_n}\omega^{\mu_1\dots\mu_n}\, \mathrm{d}x^{\nu_1} \wedge \dots \wedge \mathrm{d}x^{\nu_{10-n}}$.

$B^{(2)}$, one with the dilaton, eight with the one-form $C^{(1)}$ and 56 with the three-form $C^{(3)}$.

The dynamics of these fields is described by the following action (in the string frame):

$$S_{\mathrm{IIA}} = \frac{1}{2\kappa_{10}^2}\Big\{ \int \mathrm{d}^{10}x\, \mathrm{e}^{-2\phi}\sqrt{-\det G}\,R(G) \\ + \int [\mathrm{e}^{-2\phi}(4\mathrm{d}\phi \wedge {}^*\mathrm{d}\phi - \tfrac{1}{2}H^{(3)} \wedge {}^*H^{(3)}) - \tfrac{1}{2}F^{(2)} \wedge {}^*F^{(2)} \\ - \tfrac{1}{2}\widetilde{F}^{(4)} \wedge {}^*\widetilde{F}^{(4)} - \tfrac{1}{2}B^{(2)} \wedge F^{(4)} \wedge F^{(4)}]\Big\}. \tag{2.1}$$

In this expression κ_{10} is the gravitational coupling constant in ten dimensions given by

$$\kappa_{10} = 8\pi^{7/2}\alpha'^2 g_s \tag{2.2}$$

where $\sqrt{\alpha'}$ is the fundamental string length and g_s is the string coupling constant which is related to the vacuum expectation value of the dilaton according to $g_s = \langle \mathrm{e}^{\phi}\rangle$.[2] Furthermore,

$$H^{(3)} = \mathrm{d}B^{(2)} \qquad F^{(2)} = \mathrm{d}C^{(1)} \qquad F^{(4)} = \mathrm{d}C^{(3)} \\ \widetilde{F}^{(4)} = F^{(4)} + C^{(1)} \wedge H^{(3)}. \tag{2.3}$$

The action in (2.1) is the truncation to the purely bosonic sector of the type IIA supergravity action. It is interesting to observe that all the terms arising from the NS–NS sector are multiplied by a factor of $\mathrm{e}^{-2\phi}$, while the terms arising from the R–R sector do not contain any coupling with the dilaton. This is a distinctive feature of the so-called string frame, the one in which the action (2.1) is written. In order to remove the dilaton factor from the curvature term and to avoid mixed graviton–dilaton propagators, it is convenient to rewrite the action in the more conventional Einstein frame. This is achieved simply by means of the following redefinition of the metric tensor

$$G_{\mu\nu}(\text{string frame}) = \mathrm{e}^{\phi/2} g_{\mu\nu}(\text{Einstein frame}). \tag{2.4}$$

Using this relation and after some straightforward algebra, one finds that the effective action of the type IIA string in the Einstein frame is

$$S_{\mathrm{IIA}} = \frac{1}{2\kappa_{10}^2}\Big\{ \int \mathrm{d}^{10}x\, \sqrt{-\det g}\,R(g) \\ - \tfrac{1}{2}\int [\mathrm{d}\phi \wedge {}^*\mathrm{d}\phi + \mathrm{e}^{-\phi}H^{(3)} \wedge {}^*H^{(3)} + \mathrm{e}^{+\frac{3}{2}\phi}F^{(2)} \wedge {}^*F^{(2)} \\ + \mathrm{e}^{+\frac{1}{2}\phi}\widetilde{F}^{(4)} \wedge {}^*\widetilde{F}^{(4)} + B^{(2)} \wedge F^{(4)} \wedge F^{(4)}]\Big\}. \tag{2.5}$$

[2] By explicitly including the string coupling constant in the definition of κ_{10} we implicitly declare that the field ϕ that appears in the action (2.1) represents only the *fluctuation* of the dilaton around its vacuum expectation value.

In this frame the curvature term has the standard form of the Einstein–Hilbert action and the dilaton field also has a canonical normalization factor of $-\frac{1}{2}$. The price one has to pay for this is the appearance of non-vanishing couplings between the various antisymmetric tensors and the dilaton. The difference between the antisymmetric tensor of the NS–NS sector and those of the R–R sector is now in the sign of the dilaton exponent, which is negative for the former and positive for the latter.

2.1.2 Type IIB

The massless bosonic content of the chiral type IIB superstring consists of a graviton $G_{\mu\nu}$, an antisymmetric two-index tensor $B^{(2)}_{\mu\nu}$, a dilaton ϕ from the NS–NS sector (which is the same as in the type IIA case), a zero-form $C^{(0)}$, a 2-form $C^{(2)}$ and a four-form $C^{(4)}$ with a self-dual field strength from the R–R sector. These fields again correspond to a total of 128 physical degrees of freedom, of which 35 are associated with the graviton, 28 with the 2-form $B^{(2)}$, one with the dilaton, one with the zero-form $C^{(0)}$, 28 with the 2-form $C^{(2)}$ and 35 with the four-form $C^{(4)}$.

In the string frame the effective action of the type IIB string can be written as

$$
\begin{aligned}
S_{\mathrm{IIB}} = \frac{1}{2\kappa_{10}^2}\Bigg\{ & \int \mathrm{d}^{10}x\, \mathrm{e}^{-2\phi}\sqrt{-\det G}\,R(G) \\
& + \int [\mathrm{e}^{-2\phi}(4\mathrm{d}\phi \wedge {}^*\mathrm{d}\phi - \tfrac{1}{2}H^{(3)} \wedge {}^*H^{(3)}) - \tfrac{1}{2}F^{(1)} \wedge {}^*F^{(1)} \\
& - \tfrac{1}{2}\widetilde{F}^{(3)} \wedge {}^*\widetilde{F}^{(3)} - \tfrac{1}{4}\widetilde{F}^{(5)} \wedge {}^*\widetilde{F}^{(5)} - \tfrac{1}{2}C^{(4)} \wedge H^{(3)} \wedge F^{(3)}]\Bigg\}
\end{aligned}
\tag{2.6}
$$

where

$$
H_{(3)} = \mathrm{d}B_{(2)} \qquad F_{(1)} = \mathrm{d}C_{(0)} \qquad F_{(3)} = \mathrm{d}C_{(2)} \qquad F_{(5)} = \mathrm{d}C_{(4)} \tag{2.7}
$$

and

$$
\widetilde{F}_{(3)} = F_{(3)} + C_{(0)} \wedge H_{(3)} \qquad \widetilde{F}_{(5)} = F_{(5)} + C_{(2)} \wedge H_{(3)}. \tag{2.8}
$$

The gravitational coupling constant κ_{10} is defined in (2.2).

The structure of the type IIB action (2.6) is very similar to that in type IIA theory (see equation (2.1)), the only difference being in the field content of the R–R sector. We note that the self-duality constraint

$$
{}^*\widetilde{F}^{(5)} = \widetilde{F}^{(5)} \tag{2.9}
$$

has to be imposed only at the level of the field equations and not inside the action. In other words, the field equations that follow from (2.6) are consistent with the self-duality of $\widetilde{F}^{(5)}$ but they do not imply it. Therefore, this condition has to be imposed as an extra condition on the solutions of the field equations. Clearly this procedure is satisfactory only at the classical level and a more careful treatment is needed at the quantum level.

By rescaling the metric according to (2.4), we can rewrite the effective action of the type IIB theory in the Einstein frame, where it becomes

$$
\begin{aligned}
S_{\mathrm{IIB}} = \frac{1}{2\kappa_{10}^2}\Bigg\{ & \int \mathrm{d}^{10}x\,\sqrt{-\det g}\,R(g) \\
& - \tfrac{1}{2}\int [\mathrm{d}\phi\wedge{}^*\mathrm{d}\phi + \mathrm{e}^{-\phi}H^{(3)}\wedge{}^*H^{(3)} + \mathrm{e}^{+2\phi}F^{(1)}\wedge{}^*F^{(1)} \\
& + \mathrm{e}^{+\phi}\widetilde{F}^{(3)}\wedge{}^*\widetilde{F}^{(3)} + \tfrac{1}{2}\,\widetilde{F}^{(5)}\wedge{}^*\widetilde{F}^{(5)} + C^{(4)}\wedge H^{(3)}\wedge F^{(3)}]\Bigg\}.
\end{aligned}
\tag{2.10}
$$

As in type IIA theory, the type IIB effective action in the Einstein frame also has non-trivial exponential couplings between the dilaton and antisymmetric tensors; as before the sign in the dilaton exponent is negative for the NS–NS 2-form and positive for the forms of the R–R sector. Note, however, that there is no dilaton coupling associated with the four-form of the R–R sector.

The action (2.10) possesses an amusing $SL(2,\mathbf{R})$ symmetry, which is manifested [19] if we introduce the complex scalar field

$$\tau = C^{(0)} + \mathrm{i}\mathrm{e}^{-\phi} \tag{2.11}$$

a 2×2 matrix

$$\mathcal{M} = \frac{1}{\operatorname{Im}\tau}\begin{pmatrix} |\tau|^2 & -\operatorname{Re}\tau \\ -\operatorname{Re}\tau & 1 \end{pmatrix} = \mathrm{e}^{\phi}\begin{pmatrix} |\lambda|^2 & \chi \\ \chi & 1 \end{pmatrix} \tag{2.12}$$

and assemble the two three-forms $H^{(3)}$ and $F^{(3)}$ into a doublet as follows

$$\mathcal{F}(3) = \begin{pmatrix} H^{(3)} \\ F^{(3)} \end{pmatrix}. \tag{2.13}$$

In fact, using these definitions, the action (2.10) can be rewritten in the following way:

$$
\begin{aligned}
S_{\mathrm{IIB}} = \frac{1}{2\kappa_{10}^2}\Bigg\{ & \int \mathrm{d}^{10}x\,\sqrt{-\det g}\left[R(g) + \frac{1}{4}\operatorname{Tr}(\partial_\mu\mathcal{M}\partial^\mu\mathcal{M}^{-1})\right] \\
& - \tfrac{1}{2}\int [\mathcal{F}_i^{(3)}\mathcal{M}^{ij}\wedge{}^*\mathcal{F}_j^{(3)} + \tfrac{1}{2}\widetilde{F}^{(5)}\wedge{}^*\widetilde{F}^{(5)} \\
& + \tfrac{1}{2}\epsilon^{ij}C^{(4)}\wedge\mathcal{F}_i^{(3)}\wedge\mathcal{F}_j^{(3)}]\Bigg\}
\end{aligned}
\tag{2.14}
$$

with $i, j = 1, 2$ and thus, it is not difficult to check that this expression is invariant under the following $SL(2,\mathbf{R})$ transformations:

$$\mathcal{M} \to \mathcal{M}' = \Lambda\mathcal{M}\Lambda^{\mathrm{t}} \tag{2.15}$$
$$\mathcal{F}^{(3)} \to \mathcal{F}^{(3)'} = (\Lambda^{\mathrm{t}})^{-1}\mathcal{F}^{(3)} \tag{2.16}$$
$$g_{\mu\nu} \to g'_{\mu\nu} = g_{\mu\nu} \tag{2.17}$$
$$C^{(4)} \to C^{(4)'} = C^{(4)} \tag{2.18}$$

where

$$\Lambda = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad ad - bc = 1 \tag{2.19}$$

is an $SL(2,\mathbf{R})$ matrix. Note that transformation (2.15) can also be written in a rational form as

$$\tau \to \tau' = \frac{a\tau + b}{c\tau + d}. \tag{2.20}$$

Let us now consider the particular $SL(2,\mathbf{R})$ matrix

$$\Lambda_0 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{2.21}$$

and for simplicity, but without loss of generality, put $C^{(0)} = 0$. Then, from (2.15) (or equivalently from (2.20)) we can see that, under a Λ_0 transformation,

$$\mathcal{M}|_{C^{(0)}} = \begin{pmatrix} \mathrm{e}^{-\phi} & 0 \\ 0 & \mathrm{e}^{\phi} \end{pmatrix} \to \mathcal{M}'|_{C^{(0)}} = \begin{pmatrix} \mathrm{e}^{\phi} & 0 \\ 0 & \mathrm{e}^{-\phi} \end{pmatrix} \tag{2.22}$$

i.e.

$$\phi \to -\phi \tag{2.23}$$

furthermore from (2.16) we can see that

$$\begin{pmatrix} H^{(3)} \\ F^{(3)} \end{pmatrix} \to \begin{pmatrix} F^{(3)} \\ -H^{(3)} \end{pmatrix}. \tag{2.24}$$

This is a very interesting result: in fact, under a Λ_0 transformation the two type IIB two-forms exchange roles and the dilaton changes sign. Since, as we mentioned earlier, the vacuum expectation value of the dilaton is related to the string coupling constant, we can see that (2.23) implies

$$g_s = \langle \mathrm{e}^{\phi} \rangle \to \langle \mathrm{e}^{-\phi} \rangle = \frac{1}{g_s}. \tag{2.25}$$

This is a weak/strong coupling duality, called an S duality, which is a symmetry of the effective action of the type IIB superstring. There is much of evidence that this duality is, in fact, a true symmetry of the full type IIB superstring theory and not just of its low-energy effective action (see, for example, [3]).

2.2 General construction

Let us now consider a truncation of the (bosonic) supergravity action (2.5) or (2.10) that contains only

- the metric $g_{\mu\nu}$,
- the dilaton ϕ and
- *one* of the antisymmetric tensors, say the $(p+1)$-form potential.

It can be easily shown that this is a consistent truncation, in the sense that the fields that are retained are not sources for the fields that are eliminated [4, 17]. In view of this fact, therefore we can safely consider the following truncated action:

$$S_n = \frac{1}{2\kappa_{10}^2}\left\{\int \mathrm{d}^{10}x\,\sqrt{-\det g}R(g) - \frac{1}{2}\int [\mathrm{d}\phi\wedge{}^*\mathrm{d}\phi + \mathrm{e}^{-a\phi}F^{(n)}\wedge{}^*F^{(n)}]\right\} \tag{2.26}$$

where $F^{(n)}$ is the field strength of the antisymmetric potential we have chosen (where, of course, $n = p+2$) and a is a coefficient that we can read from action (2.5) or (2.10). In particular, we see that

- if the chosen potential is the antisymmetric tensor of $B^{(2)}$ the NS–NS sector, then $p = 1, n = 3$ and $a = 1$; and
- if the chosen potential is one of the antisymmetric tensors of the R–R sector $C^{(p+1)}$, then $n = p+2$ and $a = (p-3)/2$, where $p = 0, 2, \ldots$ in type IIA theory and $p = -1, 1, 3, \ldots$ in type IIB theory.

From action (2.26) we can easily obtain the classical field equations. For the dilaton we have

$$\frac{1}{\sqrt{-\det g}}\partial_\mu\left(\sqrt{-\det g}g^{\mu\nu}\partial_\nu\phi\right) = -\frac{a}{2n!}\mathrm{e}^{-a\phi}F^2_{(n)} \tag{2.27}$$

for the antisymmetric potential we have

$$\partial_\nu\left(\sqrt{-\det g}\mathrm{e}^{-a\phi}F^{(n)\nu\mu_1\ldots\mu_{n-1}}\right) = 0 \tag{2.28}$$

and, finally, for the metric we have the Einstein equation

$$R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R = T_{\mu\nu} \tag{2.29}$$

where the energy–momentum tensor $T_{\mu\nu}$ is given by

$$T_{\mu\nu} = \frac{1}{2}\partial_\mu\phi\partial_\nu\phi + \frac{1}{2(n-1)!}\mathrm{e}^{-a\phi}\left(F^{(n)}_{\mu\ldots}F^{(n)\nu\ldots} - \frac{n-1}{8n}{F_{(n)}}^2\right). \tag{2.30}$$

Our goal is to find solutions of these equations that represent (classical) extended objects with p spatial dimensions. To simplify things we make the following ansatz:

- we require *Poincaré invariance* in the $(p+1)$ longitudinal directions; and
- we require *rotational invariance* in the remaining $(9-p)$ transverse directions.

In other words we require the solution to possess the following symmetry:

$$\mathrm{ISO}(1, p) \otimes \mathrm{SO}(9-p) \tag{2.31}$$

and accordingly, we split the spacetime coordinates $\{x^\mu\}$ into the longitudinal ones, denoted by $\{x^a\}$ with $a = 0, 1, \ldots, p$ and the transverse ones, denoted by $\{y^i\}$ with $i = p+1, \ldots, 9$. A metric that is compatible with these requirements, then has the following form:

$$\mathrm{d}s^2 = \mathrm{e}^{2A(r)}\,\mathrm{d}x^a\,\mathrm{d}x^b\,\eta_{ab} + \mathrm{e}^{2B(r)}\,\mathrm{d}y^i\,\mathrm{d}y^j\,\delta_{ij} \tag{2.32}$$

where A and B are functions only of the radial coordinate $r = \sqrt{y^i y_i}$ in the transverse space. For the dilaton we can simply take

$$\phi = f(r) \tag{2.33}$$

while for the antisymmetric tensor we posit

$$C^{(p+1)}_{01\ldots p} = \mathrm{e}^{C(r)} - 1. \tag{2.34}$$

The so-far arbitrary functions $A(r)$, $B(r)$, $C(r)$ and $f(r)$ are then uniquely determined by inserting the ansatz into (2.27)–(2.29) and solving the resulting differential equations (see, for example, [4, 17, 18] for details). It is worth mentioning that the ansatz (2.34) on the antisymmetric potential is of *electric* type. In fact, the corresponding field strength is

$$F^{(n)}_{i01\ldots p} \sim \partial_i C(r)\mathrm{e}^{C(r)}$$

which indeed describes an 'electric' configuration. However, one can also make a *magnetic* ansatz on the antisymmetric potential, which actually amounts to making an electric ansatz on the dual field strength. In other words, in the magnetic case one requires the ten-dimensional Hodge dual of $F^{(n)}$, i.e. the $(10-n)$-form ${}^*F^{(n)}$, to be of electric type. Note that the potential associated with an electric field strength ${}^*F^{(n)}$ is a $(7-p)$-form, which naturally couples with an extended object with $(6-p)$ spatial dimensions. Therefore, we can conclude that in the ten-dimensional spacetime where the superstring theory is defined, a p-brane and a $(6-p)$-brane are 'electromagnetically' dual to each other. This is a straightforward generalization of the familiar four-dimensional case, where instead the elementary electric charge and its dual magnetic monopole are both point-like.

2.3 Explicit solutions

We are now going to present the explicit form of the solution to the supergravity field equations in three specific cases: the fundamental string (F1), the solitonic Neveu-Schwarz 5-brane (NS5) and the Dp-branes.

2.3.1 Fundamental string

The simplest brane configuration is the fundamental string, which is the classical solution of the supergravity field equations (2.27)–(2.29) that is *electrically* charged under the 2-form $B^{(2)}$ of the NS–NS sector. We therefore look for a one-dimensional extended object, i.e. a string.[3] Therefore, according to our previous discussion, in this case we must set $p = 1$ and split the ten spacetime coordinates as follows

$$\begin{array}{ll} x^0, x^1 & \text{longitudinal coordinates} \\ y^2, \ldots, y^9 & \text{transverse coordinates.} \end{array}$$

By explicitly solving the classical field equations in this case, one can obtain the following results:

$$\mathrm{d}s^2 = H(r)^{-3/4}(\mathrm{d}x^a\,\mathrm{d}x^b\,\eta_{ab}) + H(r)^{1/4}(\mathrm{d}y^i\,\mathrm{d}y^j\,\delta_{ij}) \tag{2.35}$$

$$\mathrm{e}^{\phi} = H(r)^{-1/2} \tag{2.36}$$

$$B^{(2)} = (H(r)^{-1} - 1)\,\mathrm{d}x^0 \wedge \mathrm{d}x^1 \tag{2.37}$$

where the warp factor is given by

$$H(r) = 1 + \frac{L^6}{r^6} \tag{2.38}$$

with $r = \sqrt{y^i y^j \delta_{ij}}$ being the radial coordinate in the transverse space and the length L being defined by

$$L^6 = \frac{\kappa_{10}^2}{6\Omega_7}\frac{1}{\pi\alpha'}. \tag{2.39}$$

In this expression we have denoted by Ω_n the area of the unit n-dimensional sphere S_n, i.e.

$$\Omega_n = \frac{2\pi^{(n+1)/2}}{\Gamma\left(\dfrac{n+1}{2}\right)}. \tag{2.40}$$

Using (2.2) and recalling that $\sqrt{\alpha'}$ is the fundamental length of the string, it is easy to check that indeed the quantity L defined in (2.39) has the correct dimension of a length.

[3] Since the 2-form $B^{(2)}$ is common to both type IIA and type IIB the string we seek exists in both theories.

We note that (2.35) is the metric in the Einstein frame of a string electrically charged under the NS–NS 2-form $B_{(2)}$. In the string frame, however, the metric of this string configuration becomes

$$ds^2 = H(r)^{-1}(dx^a\, dx^b\, \eta_{ab}) + (dy^i\, dy^j\, \delta_{ij}) \tag{2.41}$$

while all other fields remain as before. From this result, it is possible to compute the tension M_1 of this string and its 'electric' charge $Q_{\rm el}$ under $B^{(2)}$. The tension, measured in string frame units, can be simply read from the warp factor $H(r)$, which essentially represents the gravitational potential produced by the string. More precisely, M_1 is the coefficient of the combination $2\kappa_{10}^2/(6\Omega_7)$ that plays the role of Newton's constant in this case. Thus, from (2.38) and (2.39), we obtain

$$M_1 = \frac{1}{2\pi\alpha'} \tag{2.42}$$

which is also the tension of the elementary string that gives rise to the supergravity effective theory. For this reason, the one-dimensional extended configuration described earlier is called the fundamental string (or F1-brane). The 'electric' charge $Q_{\rm el}$ of the fundamental string under $B^{(2)}$ can be simply obtained by applying Gauss's law, which in this case leads to

$$Q_{\rm el} = \frac{1}{2\kappa_{10}^2}\int_{S_7} {}^*dB^{(2)} = \frac{1}{2\pi\alpha'}. \tag{2.43}$$

It is important to observe that the tension and charge are related to one another and, in fact, in our normalizations they are equal, namely

$$M_1 = Q_{\rm el}. \tag{2.44}$$

This relation is not a coincidence but is a consequence of the so-called BPS property of the fundamental string, namely the fact that one-half of the 32 supersymmetries of the type II superstring are preserved by this solution. The BPS relation (2.44) also implies that the attractive gravitational force between two such fundamental strings is exactly balanced by the repulsive Coulomb force they experience; this precise cancellation of forces implies that these extended objects do not interact and can be safely piled on top of each other to form macroscopic configurations with small curvatures.

2.3.2 NS 5-brane

The NS 5-brane is the magnetic dual of the fundamental string considered in the previous section. Therefore it describes an extended object with five spatial dimensions. According to our general discussion, we must split the ten spacetime coordinates as follows:

$x^0, x^1, \ldots, x^5$	longitudinal coordinates
$y^6, \ldots, y^9$	transverse coordinates

Then, by explicitly solving the classical field equations in this case, one obtains the following results:

$$\mathrm{d}s^2 = H(r)^{-1/4}(\mathrm{d}x^a\,\mathrm{d}x^b\,\eta_{ab}) + H(r)^{3/4}(\mathrm{d}y^i\,\mathrm{d}y^j\,\delta_{ij}) \tag{2.45}$$

$$\mathrm{e}^{\phi} = H(r)^{1/2} \tag{2.46}$$

$$(\mathrm{d}B^{(2)})_{ijk} = \epsilon_{ijk\ell}\partial^{\ell} H(r) \tag{2.47}$$

where the warp factor is given by

$$H(r) = 1 + \frac{L^2}{r^2} \tag{2.48}$$

with $r = \sqrt{y^i y^j \delta_{ij}}$ being, as usual, the radial coordinate in the transverse space and the length L being defined by

$$L^2 = \frac{2\pi^2\alpha'}{\Omega_3}. \tag{2.49}$$

We note that (2.45) is the metric in the Einstein frame of a 5-brane magnetically charged under the NS–NS 2-form $B_{(2)}$. In the string frame, the metric of this configuration becomes

$$\mathrm{d}s^2 = (\mathrm{d}x^a\,\mathrm{d}x^b\,\eta_{ab}) + H(r)(\mathrm{d}y^i\,\mathrm{d}y^j\,\delta_{ij}) \tag{2.50}$$

while all other fields remain as before. It is interesting to observe that in the string frame the longitudinal world-volume of the NS 5-brane is flat and only the transverse directions are warped. This is exactly the opposite of what happens in the dual fundamental string solution (2.41), where the longitudinal spacetime is warped and the transverse space is flat. Note also that the dilaton in the NS 5-brane is opposite with respect to the dilaton of the F1 solution (compare equation (2.46) with equation (2.36)).

From the explicit form (2.50) of the metric in the string frame, we can now deduce the tension M_5 of this 5-brane and its 'magnetic' charge Q_{magn} under $B_{(2)}$. As before, the tension M_5, measured in string frame units, can be simply read from the warp factor $H(r)$ given in (2.48); in particular, M_5 is the coefficient in L^2 of the combination $2\kappa_{10}^2/(2\Omega_3)$ that plays the role of Newton's constant in this case. Thus, from (2.48) and (2.49), we obtain

$$M_5 = \frac{2\pi^2\alpha'}{\kappa_{10}^2} \sim \frac{1}{g_s^2}. \tag{2.51}$$

Note that, contrary to what happened for the fundamental string, in this case the tension clearly displays a *non-perturbative* behaviour, since it varies with the inverse square of the coupling constant. This is the typical behaviour of a solitonic configuration in field theory and, for this reason, the NS 5-brane solution is also known as the solitonic brane. The 'magnetic' charge Q_{magn} of the NS 5-brane can be simply obtained by applying Gauss's law (for the magnetic field) which, in this case, leads to

$$Q_{\mathrm{magn}} = \frac{1}{2\kappa_{10}^2} \int_{S_3} \mathrm{d}B_{(2)} = \frac{2\pi^2\alpha'}{\kappa_{10}^2}. \tag{2.52}$$

As in the fundamental string, now we also find that the tension and the charge are related to one another according to the BPS relation

$$M_5 = Q_{\mathrm{magn}} \tag{2.53}$$

implying again that half of the 32 supersymmetries of the type II theory are preserved by the solitonic 5-brane solution.

2.3.3 Dp-branes

In some sense the so-called D-branes are the most interesting and intriguing configurations of string theory. They are non-trivial solutions of the supergravity field equations of type IIA or IIB that are charged under one of the antisymmetric potentials of the R–R sector. It has been well known for a long time that no perturbative configuration of string theory can carry charge under the R–R potentials and thus the discovery of D-branes has represented a remarkable breakthrough in our understanding of string theory and, in particular, of its non-perturbative features. From the point of view of supergravity, the D-branes are very similar to the other brane-solutions we discussed earlier, the relevant differences being in the type of antisymmetric tensor that is switched on and in their space dimensions. However, from a string-theory point of view they are drastically different. Indeed, a Dp-brane is a $(p+1)$-extended object in the ten-dimensional spacetime defined by the distinctive property that open strings can terminate on it [5, 6]. In other words, a Dp-brane is a hypersurface spanned by open strings with Dirichlet boundary conditions in the $(9-p)$ transverse directions. Since the role of Dirichlet boundary conditions is crucial in this case, these branes have been called Dirichlet branes or simply D-branes.

Let us now present the explicit form of the Dp-brane solution with p even in type IIA and odd in type IIB. According to our general discussion, to describe a $(p+1)$-dimensional extended object we first split the ten spacetime coordinates as follows:

$$\begin{array}{ll} x^0, x^1, \ldots, x^p & \text{longitudinal coordinates} \\ y^{p+1}, \ldots, y^9 & \text{transverse coordinates} \end{array}$$

and then solve the supergravity field equations (2.27)–(2.29) using the ansatz

(2.32)–(2.34). In this way one can obtain the following results:

$$\mathrm{d}s^2 = H(r)^{-(7-p)/8}(\mathrm{d}x^a\,\mathrm{d}x^b\,\eta_{ab}) + H(r)^{(p+1)/8}(\mathrm{d}y^i\,\mathrm{d}y^j\,\delta_{ij}) \quad (2.54)$$
$$\mathrm{e}^{\phi} = H(r)^{(3-p)/4} \quad (2.55)$$
$$C^{(p+1)} = (H(r)^{-1} - 1)\,\mathrm{d}x^0 \wedge \ldots \wedge \mathrm{d}x^p \quad (2.56)$$

where the warp factor is given by

$$H(r) = 1 + \frac{L^{7-p}}{r^{7-p}} \quad (2.57)$$

with $r = \sqrt{y^i y^j \delta_{ij}}$ being the radial coordinate in the transverse space and the length L being defined by

$$L^{7-p} = \frac{2\kappa_{10}}{(7-p)\Omega_{8-p}}(\sqrt{\pi}(2\pi\sqrt{\alpha'})^{3-p}). \quad (2.58)$$

We note that (2.54) is the metric in the Einstein frame of a p-brane that is electrically charged under the $(p+1)$-form potential of the R–R sector. In the string frame, the metric of this configuration becomes

$$\mathrm{d}s^2 = H(r)^{-1/2}(\mathrm{d}x^a\,\mathrm{d}x^b\,\eta_{ab}) + H(r)^{1/2}(\mathrm{d}y^i\,\mathrm{d}y^j\,\delta_{ij}) \quad (2.59)$$

while all other fields remain as before. From this form we can see that the D-branes are somehow intermediate configurations between the fundamental string and the solitonic 5-brane. In fact, in the metric (2.59) both the longitudinal and transverse directions are warped (with inverse factors); this is to be contrasted with the metric of the fundamental string (2.41) where only the longitudinal directions are warped and with the one of the solitonic 5-brane (2.50) where only the transverse directions are warped. Later on we will see that the D-branes are intermediate configurations in another sense.

From the explicit solution (2.54)–(2.56), it is possible to compute the tension M_p of the p-brane and its ‘electric’ charge Q_p under $C^{(p+1)}$. As in the cases examined in the previous sections, the tension M_p, measured in string frame units, can be simply read from the warp factor $H(r)$ (2.57), which essentially represents the gravitational potential produced by brane. More precisely, M_p is the coefficient of the combination $2\kappa_{10}^2/((7-p)\Omega_{8-p})$ that plays the role of Newton’s constant in this case. Thus, from (2.57) and (2.58), we obtain [5]

$$M_p = \frac{\sqrt{\pi}(2\pi\sqrt{\alpha'})^{3-p}}{\kappa_{10}} \sim \frac{1}{g_s}. \quad (2.60)$$

This result clearly indicates that these Dp-branes are *non-perturbative* configurations of string theory; however, they are of a non-standard type since their tension scales with the inverse power of the coupling constant, while typical solitonic solutions are characterized instead by the inverse square of coupling constant (see, for example, equation (2.51)). Thus, from this point of view also we can say that the D-branes are somehow intermediate configurations between the (perturbative) fundamental string and the solitonic 5-brane. It is essentially for this reason that the D-branes can be studied in a very explicit way by means of open strings (with Dirichlet boundary conditions); and in fact they are extremely powerful tools that allow us to obtain precise information on some non-perturbative features of string theory.

Finally, let us compute the 'electric' charge Q_p of the Dp-brane under the R–R potential $C^{(p+1)}$. This can be simply obtained by applying Gauss's law which, in this case, leads to

$$Q_p = \frac{1}{2\kappa_{10}^2} \int_{S_{8-p}} {}^*\mathrm{d}C^{(p+1)} = \frac{\sqrt{\pi}(2\pi\sqrt{\alpha'})^{3-p}}{\kappa_{10}}. \tag{2.61}$$

Comparing with (2.60), we can see that for the Dp-brane solution also we have the BPS relation

$$M_p = Q_p. \tag{2.62}$$

This is a signal of the fact that one-half of the 32 supersymmetries of type II theory are preserved by the Dp-brane, or, put differently, that there is an exact cancellation between the attractive force of the NS–NS fields due to the tension M_p, and the repulsive Coulomb-like force of the R–R potential due to the charge Q_p.

We conclude this section by observing that from the explicit expression (2.61) we have

$$2\kappa_{10}^2\, Q_p\, Q_{6-p} = 2\pi. \tag{2.63}$$

This is a generalization of Dirac's quantization condition of the electric charge (suitably written for the type II string effective actions with coupling constant $2\kappa_{10}^2$), from which we can deduce that a Dp-brane and a D$(6-p)$-brane are electromagnetically dual to each other.

2.3.4 The geometry of the D3-brane of type IIB

In this section we recall some peculiar features of the spacetime geometry produced by the D3-branes of type IIB, which in the last few years have been extensively used in the so-called AdS/CFT correspondence [14–16]. Specializing the explicit solution (2.54)–(2.56) to the case $p = 3$, and considering a stack of N coincident D3 branes, we have

$$\mathrm{d}s^2 = H(r)^{-1/2}(\mathrm{d}x^a\, \mathrm{d}x^b\, \eta_{ab}) + H(r)^{1/2}(\mathrm{d}y^i\, \mathrm{d}y^j\, \delta_{ij}) \tag{2.64}$$

$$\mathrm{e}^{\phi} = 1 \tag{2.65}$$

$$C^{(4)} = (H(r)^{-1} - 1)\, \mathrm{d}x^0 \wedge \ldots \wedge \mathrm{d}x^3 \tag{2.66}$$

where the longitudinal coordinates are labelled by $a, b = 0, \ldots, 3$, the transverse coordinates by $i, j = 4, \ldots, 9$, and the warp factor is given by

$$H(r) = 1 + \frac{L^4}{r^4} \tag{2.67}$$

with $r = \sqrt{y^i y^j \delta_{ij}}$ being the radial coordinate in the transverse space and the length L being defined by

$$L^4 = N \frac{2\kappa_{10}\sqrt{\pi}}{4\Omega_5} = 4\pi N g_s \alpha'^2. \tag{2.68}$$

Note that since the dilaton is zero in the D3-brane solution, there is no difference between the Einstein frame and the string frame. As we mentioned before, due to the BPS no-force condition (2.62), the D3-branes can be piled up on top of each other to form a 'macroscopic' configuration; therefore, the potential produced by a stack of N coincident branes is simply N times the potential produced by a single brane. This explains the factor of N in (2.68).

Let us now consider the detailed form of the metric (2.64) at distances $r \gg L$, i.e. far away from the branes. In this region the harmonic function $H(r)$ in equation (2.67) can be approximated to one, so that the metric reduces to that of the flat ten-dimensional Minkowski spacetime. This is not unexpected since normally any field dies off at infinity, i.e. far away from its source. If we include the first-order correction, the flat geometry is modified by small terms which can be studied by standard perturbative methods, including string theory calculations of graviton scattering amplitudes.

Near the branes, i.e. for $r \ll L$, we have a very different scenario. In this region in fact, we can neglect the one in the harmonic function $H(r)$ of equation (2.67), so that the metric reduces to

$$\mathrm{d}s^2 \simeq \left(\frac{r^2}{L^2}\right) \mathrm{d}x^a \, \mathrm{d}x^b \, \eta_{ab} + \left(\frac{L^2}{r^2}\right) \mathrm{d}y^i \, \mathrm{d}y^j \, \delta_{ij}. \tag{2.69}$$

If we introduce spherical coordinates in the transverse space and write

$$\mathrm{d}y^i \, \mathrm{d}y^j \, \delta_{ij} = \mathrm{d}r^2 + r^2 \, \mathrm{d}\Omega_5^2 \tag{2.70}$$

where $\mathrm{d}\Omega_5^2$ is the metric of a unit 5-sphere S_5, we can easily see that equation (2.69) becomes

$$\mathrm{d}s^2 \simeq \left[\left(\frac{r^2}{L^2}\right) \mathrm{d}x^a \, \mathrm{d}x^b \, \eta_{ab} + \left(\frac{L^2}{r^2}\right) \mathrm{d}r^2\right] + L^2 \, \mathrm{d}\Omega_5^2. \tag{2.71}$$

Moreover, if we define $z = L^2/r$, the part of the previous metric in square brackets can be rewritten as

$$\left(\frac{L^2}{z^2}\right) (\mathrm{d}x^a \, \mathrm{d}x^b \, \eta_{ab} + \mathrm{d}z^2). \tag{2.72}$$

This is one of the standard forms in which the metric of a five-dimensional anti-de Sitter spacetime of radius L is usually written. Therefore, at distances $r \ll L$ the geometry produced by N D3-branes appears as the product of a five-dimensional anti-de Sitter spacetime AdS_5 times a five-dimensional sphere S_5, both with radius L.

In view of this analysis, we can say that the D3-branes of type IIB string theory are classical non-perturbative solutions that interpolate between

- the flat Minkowski spacetime in ten dimensions for $r \gg L$

and

- the $AdS_5 \times S_5$ spacetime for $r \ll L$.

Note that in the asymptotic region $r \gg L$, the ten spacetime coordinates are naturally split into $4 + 6$, as it is appropriate for a D3-brane, while in the near brane region they are split into $5 + 5$, since the radial coordinate r (or the closely related z coordinate) 'transmigrates' to join the longitudinal parameters.

The peculiar geometry of the $AdS_5 \times S_5$ spacetime has been intensively investigated in recent years in the light of Maldacena's celebrated conjecture [14, 15], which states that the type IIB string in an $AdS_5 \times S_5$ background is dual to the $\mathcal{N} = 4$ superconformal Yang–Mills theory in a flat four-dimensional Minkowski spacetime in the strong coupling limit. This remarkable duality, which has been successfully tested in numerous examples, allows us to perform classical (super)gravity calculations in an $AdS_5 \times S_5$ spacetime in order to obtain quantum results for the dual four-dimensional Yang–Mills theory in the strong coupling regime. Analysis of this gauge/gravity correspondence, of its applications and extensions is well beyond the purpose of these lectures and thus we simply refer to the existing reviews on this subject [16].

Chapter 3

The boundary state description of D-branes

As we mentioned in the introduction, the D-branes are characterized by the fact that open strings can end on them. Thus, a Dp-brane is a $(p+1)$-dimensional hyperplane spanned by open strings which have the standard Neumann boundary conditions in the $(p+1)$ longitudinal directions and Dirichlet boundary conditions in the remaining $(9-p)$ transverse directions. In this chapter we are going to present an alternative (though completely equivalent) description based instead on *closed* strings which are emitted (or absorbed) by world-sheets with boundaries on which the string coordinates obey the appropriate boundary conditions. As we shall see, this description based on the use of the so-called boundary state turns out to be extremely useful for practical applications; moreover it allows us to establish a very clear relation between the stringy description of D-branes to the supergravity description presented in the previous chapter. A more extensive review of this boundary state approach to the D-branes can be found, for example, in [13], while the standard description based on the use of open strings with Dirichlet boundary conditions can be found in the reviews in [6].

3.1 The boundary state with an external field

In the closed string operator formalism the supersymmetric Dp-branes of type II theories are described by means of boundary states $|B\rangle$ [8, 9, 20]. These are closed string states which insert a boundary on the world-sheet and enforce the appropriate boundary conditions on it. Both in the NS–NS and R–R sectors, there are two possible implementations for the boundary conditions of a Dp-brane which correspond to two boundary states $|B, \eta\rangle$, with $\eta = \pm 1$. However, only the combinations

$$|B\rangle_{\rm NS} = \frac{1}{2}[|B, +\rangle_{\rm NS} - |B, -\rangle_{\rm NS}] \tag{3.1}$$

and

$$|B\rangle_{\rm R} = \frac{1}{2}[|B, +\rangle_{\rm R} + |B, -\rangle_{\rm R}] \tag{3.2}$$

are selected by the GSO projection in the NS–NS and R–R sectors respectively. As discussed in [11], the boundary state $|B, \eta\rangle$ is the product of a matter part and a ghost part:

$$|B, \eta\rangle = \frac{1}{2} T_p |B_{\mathrm{mat}}, \eta\rangle |B_{\mathrm{g}}, \eta\rangle \tag{3.3}$$

where

$$|B_{\mathrm{mat}}, \eta\rangle = |B_X\rangle |B_\psi, \eta\rangle \qquad |B_{\mathrm{g}}, \eta\rangle = |B_{\mathrm{gh}}\rangle |B_{\mathrm{sgh}}, \eta\rangle. \tag{3.4}$$

The overall normalization T_p can be unambiguously fixed from the factorization of amplitudes of closed strings emitted from a disc [10, 21] and is the brane tension [6] in units of the ten-dimensional gravitational coupling constant (see equation (2.60)), namely

$$T_p = \sqrt{\pi}\left(2\pi\sqrt{\alpha'}\right)^{3-p}. \tag{3.5}$$

The explicit expressions of the various components of $|B\rangle$ have been given in [11] in the simplest case of a static D-brane. However, the operator structure of the boundary state does not change even when more general configurations are considered and is always of the form

$$|B_X\rangle = \exp\left[-\sum_{n=1}^{\infty} \frac{1}{n} \alpha_{-n} \cdot S \cdot \tilde{\alpha}_{-n}\right] |B_X\rangle^{(0)} \tag{3.6}$$

and

$$|B_\psi, \eta\rangle_{\mathrm{NS}} = -\mathrm{i} \exp\left[\mathrm{i}\eta \sum_{m=1/2}^{\infty} \psi_{-m} \cdot S \cdot \tilde{\psi}_{-m}\right] |0\rangle \tag{3.7}$$

for the NS–NS sector, and

$$|B_\psi, \eta\rangle_{\mathrm{R}} = -\exp\left[\mathrm{i}\eta \sum_{m=1}^{\infty} \psi_{-m} \cdot S \cdot \tilde{\psi}_{-m}\right] |B, \eta\rangle_{\mathrm{R}}^{(0)} \tag{3.8}$$

for the R–R sector. The matrix S and the zero-mode contributions $|B_X\rangle^{(0)}$ and $|B, \eta\rangle_{\mathrm{R}}^{(0)}$ encode all information about the overlap equations that the string coordinates have to satisfy, which in turn depend on the boundary conditions of the open strings ending on the Dp-brane. Since the ghost and superghost fields are not affected by the type of boundary conditions that are imposed, the ghost part of the boundary state is always the same. Its explicit expression can be found in [11] but we do not write it again here since it will not play any significant role in our present discussion. However, we would like to recall that the boundary state must be written in the $(-1, -1)$ superghost picture in the NS–NS sector and in the asymmetric $(-1/2, -3/2)$ picture in the R–R in order to saturate the superghost number anomaly of the disc [11,22].

When a constant gauge field F is present on the D-brane world-volume, the overlap conditions that the boundary state must satisfy are [8]

$$\begin{aligned}\{(\mathbb{1}+\hat{F})^a_b\alpha^b_n + (\mathbb{1}-\hat{F})^a_b\tilde{\alpha}^b_{-n}\}|B_X\rangle = 0\\ \{\alpha^i_n - \tilde{\alpha}^i_{-n}\}|B_X\rangle = 0\end{aligned} \tag{3.9}$$

for the bosonic part and

$$\begin{aligned}\{(\mathbb{1}+\hat{F})^a_b\psi^b_m - \mathrm{i}\eta(\mathbb{1}-\hat{F})^a_b\tilde{\psi}^b_{-m}\}|B_\psi,\eta\rangle = 0\\ \{\psi^i_m + \mathrm{i}\eta\tilde{\psi}^i_{-m}\}|B_\psi,\eta\rangle = 0\end{aligned} \tag{3.10}$$

for the fermionic part. In these equations, the indices $a, b, \ldots$ label the world-volume directions $0, 1, \ldots, p$ along which the Dp-brane extends, while the latin indices $i, j, \ldots$ label the transverse directions $p+1, \ldots, 9$; moreover $\hat{F} = 2\pi\alpha' F$. These equations are solved by the 'coherent states' (3.6)–(3.8) with a matrix S given by

$$S_{\mu\nu} = ([(\eta-\hat{F})(\eta+\hat{F})^{-1}]_{ab}; -\delta_{ij}) \tag{3.11}$$

and with the zero-mode parts given by

$$|B_X\rangle^{(0)} = \sqrt{-\det(\eta+\hat{F})}\;\delta^{(9-p)}(q^i - y^i)\prod_{\mu=0}^{9}|k^\mu = 0\rangle \tag{3.12}$$

for the bosonic sector and by

$$|B_\psi,\eta\rangle^{(0)}_{\mathrm{R}} = \left(C\Gamma^0\Gamma^1\ldots\Gamma^p\frac{1+\mathrm{i}\eta\Gamma_{11}}{1+\mathrm{i}\eta}U\right)_{AB}|A\rangle|\tilde{B}\rangle \tag{3.13}$$

for the R sector. In writing these formulae we have denoted by y^i the position of the D-brane, by C the charge conjugation matrix and by U the following matrix

$$U = \frac{1}{\sqrt{-\det(\eta+\hat{F})}}\,;\exp\left(-\frac{1}{2}\hat{F}_{ab}\Gamma^a\Gamma^b\right); \tag{3.14}$$

where the symbol ; ; means that one has to expand the exponential and then antisymmetrize the indices of the Γ-matrices. Finally, $|A\rangle|\tilde{B}\rangle$ stands for the spinor vacuum of the R–R sector.[1] We would like to remark that the overlap equations (3.9) and (3.10) do not allow us to determine the overall normalization of the boundary state, and not even to get the Born–Infeld prefactor of equation (3.12). The latter can be introduced by hand as in [8] but can also be derived by boosting the boundary state and then performing a T-duality as explicitly shown in [23].

[1] For our conventions on Γ-matrices, spinors etc see, for example, [10, 11].

We end this chapter with a few comments. If F is an external *magnetic* field, the corresponding boundary state describes a stable BPS bound state formed by a Dp-brane with other lower dimensional D-branes (like, for example, the Dp-D$(p-2)$ bound state). This case was explicitly considered in [10] where the long distance behaviour of the massless fields of these configurations was determined using the boundary state approach. In contrast, if F is an external *electric* field, then the boundary state describes a stable bound state between a fundamental string and a Dp-brane that preserves one-half of the spacetime supersymmetries. This kind of bound state denoted by (F, Dp) is a generalization of the dyonic string configurations of Schwarz [19] which has been studied from the supergravity point of view in [24] and from the operator formalism point of view in [12].

Chapter 4

The effective action of D-branes

We now show how the low-energy effective action of a D-brane is related to the boundary state we have just constructed. As we have mentioned before, the boundary state is the exact conformal description of a D-brane and therefore it contains the complete information about the interactions between a D-brane and the closed strings that propagate in the bulk. In particular, it encodes the couplings with the bulk massless fields which can be simply obtained by saturating the boundary state $|B\rangle$ with the massless states of the closed string spectrum. In order to find a non-vanishing result, it is necessary to soak up the superghost number anomaly of the disc and thus, as a consequence of the superghost charge of the boundary state, we have to use closed string states in the $(-1,-1)$ picture in the NS–NS sector and states in the asymmetric $(-\frac{1}{2},-\frac{3}{2})$ picture in the R–R sector.

In the NS–NS sector, the states that represent the graviton $h_{\mu\nu}$, the dilaton ϕ and the Kalb–Ramond antisymmetric tensor $A_{\mu\nu}$ are of the form

$$\epsilon_{\mu\nu}\tilde{\psi}^{\mu}_{-\frac{1}{2}}\psi^{\nu}_{-\frac{1}{2}}|k/2\rangle_{-1}|\widetilde{k/2}\rangle_{-1} \tag{4.1}$$

with

$$\epsilon_{\mu\nu} = h_{\mu\nu} \qquad h_{\mu\nu} = h_{\nu\mu} \qquad k^{\mu}h_{\mu\nu} = \eta^{\mu\nu}h_{\mu\nu} = 0 \tag{4.2}$$

for the graviton,

$$\epsilon_{\mu\nu} = \frac{\phi}{2\sqrt{2}}(\eta_{\mu\nu} - k_{\mu}\ell_{\nu} - k_{\nu}\ell_{\mu}) \qquad \ell^2 = 0 \qquad k\cdot\ell = 1 \tag{4.3}$$

for the dilaton and

$$\epsilon_{\mu\nu} = \frac{1}{\sqrt{2}}A_{\mu\nu} \qquad A_{\mu\nu} = -A_{\nu\mu} \qquad k^{\mu}A_{\mu\nu} = 0 \tag{4.4}$$

for the Kalb–Ramond field. In order to obtain their couplings with the boundary state it is useful first to compute the quantity

$$J^{\mu\nu} \equiv {}_{-1}\langle\widetilde{k/2}|\,{}_{-1}\langle k/2|\psi^{\nu}_{\frac{1}{2}}\tilde{\psi}^{\mu}_{\frac{1}{2}}|B\rangle_{\mathrm{NS}} = -\frac{T_p}{2}V_{p+1}\sqrt{-\det(\eta+\hat{F})}\,S^{\nu\mu} \tag{4.5}$$

where V_{p+1} is the (infinite) world-volume of the brane; and then to project it onto the various independent fields using their explicit polarizations. We thus obtain: for the graviton

$$J_h \equiv J^{\mu\nu} h_{\mu\nu} = -T_p V_{p+1} \sqrt{-\det(\eta + \hat{F})} [(\eta + \hat{F})^{-1}]^{ab} h_{ba} \qquad (4.6)$$

where we have used the tracelessness of $h_{\mu\nu}$; for the dilaton

$$\begin{aligned} J_\phi &\equiv \frac{1}{2\sqrt{2}} J^{\mu\nu} (\eta_{\mu\nu} - k_\mu \ell_\nu - k_\nu \ell_\mu) \phi \\ &= \frac{T_p}{2\sqrt{2}} V_{p+1} \sqrt{-\det(\eta + \hat{F})} [3 - p + \mathrm{Tr}(\hat{F}(\eta + \hat{F})^{-1})] \phi \end{aligned} \qquad (4.7)$$

and, finally, for the Kalb–Ramond field

$$\begin{aligned} J_A &\equiv \frac{1}{\sqrt{2}} J^{\mu\nu} A_{\mu\nu} \\ &= -\frac{T_p}{2\sqrt{2}} V_{p+1} \sqrt{-\det(\eta + \hat{F})} [(\eta - \hat{F})(\eta + \hat{F})^{-1}]^{ab} A_{ba} \\ &= -\frac{T_p}{\sqrt{2}} V_{p+1} \sqrt{-\det(\eta + \hat{F})} [(\eta + \hat{F})^{-1}]^{ab} A_{ba} \end{aligned} \qquad (4.8)$$

where in the last line we have used the antisymmetry of $A_{\mu\nu}$.

We now show that the couplings J_h, J_ϕ and J_A are precisely the ones that are produced by the Dirac–Born–Infeld action which governs the low-energy dynamics of the D-brane. In the string frame, this action reads as follows

$$S_{\mathrm{DBI}} = -\frac{T_p}{\kappa_{10}} \int_{V_{p+1}} \mathrm{d}^{p+1}\xi \, \mathrm{e}^{-\phi} \sqrt{-\det[G + \mathcal{A} + \hat{F}]} \qquad (4.9)$$

where κ_{10} is the gravitational coupling constant defined in equation (2.2), T_p is the brane tension defined in equation (3.5), and G_{ab} and $\mathcal{A}_{ab}$ are, respectively, the pullbacks of the spacetime metric and of the NS–NS antisymmetric tensor on the D-brane world-volume.

In order to compare the couplings described by this action with the ones obtained from the boundary state, it is first necessary to rewrite S_{DBI} in the Einstein frame. In fact, like any string amplitude computed with the operator formalism, also the couplings J_h, J_ϕ and J_A are written in the Einstein frame. Furthermore, it is also convenient to introduce canonically normalized fields. These two goals can be realized by means of the following field redefinitions

$$G_{\mu\nu} = \mathrm{e}^{\phi/2} g_{\mu\nu} \qquad \phi = \sqrt{2}\kappa_{10}\varphi \qquad \mathcal{A}_{\mu\nu} = \sqrt{2}\kappa_{10}\mathrm{e}^{\phi/2} A_{\mu\nu}. \qquad (4.10)$$

Using the new fields in equation (4.9), we easily get

$$S_{\mathrm{DBI}} = -\frac{T_p}{\kappa_{10}} \int_{V_{p+1}} \mathrm{d}^{p+1}\xi \, \mathrm{e}^{-\frac{\kappa_{10}(3-p)}{2\sqrt{2}}\varphi} \sqrt{-\det[g + \sqrt{2}\kappa_{10} A + \hat{F}\mathrm{e}^{-\frac{\kappa_{10}}{\sqrt{2}}\varphi}]}. \qquad (4.11)$$

By expanding the metric around the flat background

$$g_{\mu\nu} = \eta_{\mu\nu} + 2\kappa_{10} h_{\mu\nu} \tag{4.12}$$

and keeping only the terms which are linear in h, ϕ and A, the action (4.11) reduces to the following expression

$$\begin{aligned} S_{\mathrm{DBI}} \simeq & -T_p \int_{V_{p+1}} \mathrm{d}^{p+1}\xi \sqrt{-\det[\eta + \hat{F}]} \\ & \times \left\{ [(\eta+\hat{F})^{-1}]^{ab} h_{ba} - \frac{1}{2\sqrt{2}}[3 - p + \mathrm{Tr}(\hat{F}(\eta+\hat{F})^{-1})]\phi \right. \\ & \left. + \frac{1}{\sqrt{2}}[(\eta+\hat{F})^{-1}]^{ab} A_{ba} \right\}. \end{aligned} \tag{4.13}$$

It is now easy to see that the couplings with the graviton, the dilaton and the Kalb–Ramond field that can be obtained from this action are exactly the same as those obtained from the boundary state and given in equations (4.6), (4.7) and (4.8) respectively.

Let us now turn to the R–R sector. As we mentioned earlier, in this sector we have to use states in the asymmetric $(-\frac{1}{2}, -\frac{3}{2})$ picture in order to soak up the superghost number anomaly of the disc. In the more familiar symmetric $(-\frac{1}{2}, -\frac{1}{2})$ picture the massless states are associated to the field strengths of the R–R potentials. In contrast, in the $(-\frac{1}{2}, -\frac{3}{2})$ picture the massless states are associated directly to the R–R potentials which, in form notation, we denote by

$$C^{(n)} = \frac{1}{n!} C_{\mu_1 \ldots \mu_n} \, \mathrm{d}x^{\mu_1} \wedge \ldots \wedge \mathrm{d}x^{\mu_n} \tag{4.14}$$

with $n = 1, 3, 5, 7, 9$ in type IIA theory and $n = 0, 2, 4, 6, 8, 10$ in type IIB theory. The string states $|C^{(n)}\rangle$ representing these potentials have a rather non-trivial structure. In fact, as shown in [11], the natural expression

$$|C^{(n)}\rangle \simeq \frac{1}{n!} C_{\mu_1 \ldots \mu_n} \left(C\Gamma^{\mu_1 \ldots \mu_n} \frac{1 + \Gamma_{11}}{2} \right)_{AB} |A; k/2\rangle_{-1/2} |\widetilde{B}; \widetilde{k/2}\rangle_{-3/2} \tag{4.15}$$

is BRST invariant only if the potential is pure gauge. To avoid this restriction, in general it is necessary to add to equation (4.15) a whole series of terms with the same structure but with different contents of superghost zero-modes. However, in the present situation there exists a short-cut that considerably simplifies the analysis. In fact, one can use the incomplete states (4.15) and ignore the superghosts, whose contribution can then be recovered simply by changing at the end the overall normalizations of the amplitudes.[1] Keeping this in mind, the couplings between the R–R potentials (4.14) and the Dp-brane can therefore

[1] Note that this procedure is not allowed when the odd-spin structure contributes, see [11].

be obtained by computing the overlap between the states (4.15) and the R–R component of the boundary state, namely

$$J_{C^{(n)}} \equiv \langle C^{(n)} | B \rangle_{\mathrm{R}}. \tag{4.16}$$

The evaluation of $J_{C^{(n)}}$ is straightforward, even if a bit lengthy (for details see [12]); and the final result is

$$J_{C_{(n)}} = -\frac{T_p}{16\sqrt{2}n!} V_{p+1} C_{\mu_1 \ldots \mu_n} \operatorname{Tr}(\Gamma^{\mu_n \ldots \mu_1} \Gamma^0 \ldots \Gamma^p ; \mathrm{e}^{-\frac{1}{2}\hat{F}_{ab}\Gamma^a \Gamma^b} ;). \tag{4.17}$$

It is easy to realize that the trace in this equation is non-vanishing only if $n = p + 1 - 2\ell$, where ℓ denotes the power of $\hat{F}$ which is produced by expanding the exponential term. Due to the antisymmetrization ; ; prescription, the integer ℓ takes only a finite number of values up to a maximum $\ell_{\max}$ which is $p/2$ for the type IIA string and $(p + 1)/2$ for the type IIB string. The simplest term to compute, corresponding to $\ell = 0$, describes the coupling of the boundary state with a $(p + 1)$-form potential of the R–R sector and is given by

$$J_{C^{(p+1)}} = \frac{\sqrt{2}T_p}{(p+1)!} V_{p+1} C_{a_0 \ldots a_p} \epsilon^{a_0 \ldots a_p} \tag{4.18}$$

where $\epsilon^{a_0 \ldots a_p}$ is the completely antisymmetric tensor on the D-brane world-volume. From equation (4.18) we can immediately deduce that the charge μ_p of a Dp-brane with respect to the R–R potential $C^{(p+1)}$ is

$$\mu_p = \sqrt{2} T_p \tag{4.19}$$

in agreement with Polchinski's original calculation [5].

The next term in the expansion of the exponential of equation (4.17) corresponds to $\ell = 1$ and yields the coupling of the Dp-brane with a $(p-1)$-form potential which is given by

$$J_{C^{(p-1)}} = \frac{\mu_p}{2(p-1)!} V_{p+1} C_{a_0 \ldots a_{p-2}} \hat{F}_{a_{p-1} a_p} \epsilon^{a_0 \ldots a_p}. \tag{4.20}$$

By proceeding in the same way, one can also easily evaluate the higher-order terms generated by the exponential which describe the interactions of the D-brane with potential forms of lower degree. All these couplings can be encoded in the following Wess–Zumino-like term

$$S_{\mathrm{WZ}} = \mu_p \int_{V_{p+1}} \left[\sum_{\ell=0}^{\ell_{\max}} C^{(p+1-2\ell)} \wedge \mathrm{e}^{\hat{F}} \right]_{p+1} \tag{4.21}$$

where $\hat{F} = \frac{1}{2} \hat{F}_{ab}\, \mathrm{d}\xi^a \wedge \mathrm{d}\xi^b$, and $C^{(n)}$ is the pullback of the n-form potential (4.14) on the D-brane world-volume. The square bracket in equation (4.21) means that in expanding the exponential form one has to pick up only the terms of total degree $(p+1)$, which are then integrated over the $(p+1)$-dimensional world-volume.

In conclusion we have explicitly shown that by projecting the boundary state $|B\rangle$ with an external field onto the massless states of the closed string spectrum, one can reconstruct the linear part of the low-energy effective action of a Dp-brane. This is the sum of the Dirac–Born–Infeld part (4.13) and the (anomalous) Wess–Zumino term (4.21) which are produced, respectively, by the NS–NS and R–R components of the boundary state.

Chapter 5

Classical D-branes from the boundary state

In this section we are going to show that the boundary state is also a very efficient tool to obtain the classical solution corresponding to a Dp-brane at long distances.

For simplicity, from now on we will consider only the case of a pure Dp-brane, with no external gauge field on its world-volume, which is described by a boundary state like the one given in chapter 3 with $F = 0$ and a diagonal S matrix given by

$$S = (\eta_{ab}; -\delta_{ij}) \tag{5.1}$$

the procedure can, however, be applied to more general cases, as shown in [12].

To obtain the long-distance behaviour of the fields emitted by a Dp-brane, one simply adds a closed string propagator D to the boundary state B and then projects the resulting expression onto the various massless states of the closed string spectrum. According to this procedure, the long-distance fluctuation of a field Ψ is then given by

$$\delta\Psi \equiv \langle P^{(\Psi)} | D | B \rangle \tag{5.2}$$

where $\langle P^{(\Psi)}|$ denotes the projector associated to Ψ, i.e. the operator that, when applied to an arbitrary massless state of the closed string, selects the Ψ component contained in that state.

Before giving the details of this calculation, we would like to observe that, since we are not using explicitly the ghost and superghost degrees of freedom, we must take into account their contribution by shifting appropriately the zero-point energy and use for the closed string propagator the following expression

$$D = \frac{\alpha'}{4\pi} \int_{|z|\le 1} \frac{\mathrm{d}^2 z}{|z|^2} z^{L_0 - a} \bar{z}^{\tilde{L}_0 - a} \tag{5.3}$$

where the operators L_0 and $\tilde{L}_0$ depend only on the orbital oscillators and the intercept is $a = \frac{1}{2}$ in the NS–NS sector and $a = 0$ in the R–R sector. Let us now begin our analysis by studying the projection (5.2) in the NS–NS sector. The projector operators onto the states of the NS–NS sector can be easily obtained from equations (4.1)–(4.4) and are

$$\langle P^{(\phi)}| = {}_{-1}\langle\widetilde{k/2}|\,{}_{-1}\langle k/2|\psi^{\nu}_{\frac{1}{2}}\tilde{\psi}^{\mu}_{\frac{1}{2}}\frac{1}{\sqrt{8}}(\eta_{\mu\nu} - k_\mu\ell_\nu - k_\nu\ell_\mu) \tag{5.4}$$

$$\begin{aligned}\langle P^{(h)\mu\nu}| = {}& {}_{-1}\langle\widetilde{k/2}|\,{}_{-1}\langle k/2|\frac{1}{2}(\psi^{\nu}_{\frac{1}{2}}\tilde{\psi}^{\mu}_{\frac{1}{2}} + \psi^{\mu}_{\frac{1}{2}}\tilde{\psi}^{\nu}_{\frac{1}{2}}) \\ & - \langle P^{(\phi)}|\frac{1}{\sqrt{8}}(\eta^{\mu\nu} - k^\mu\ell^\nu - k^\nu\ell^\mu)\end{aligned} \tag{5.5}$$

$$\langle P^{(A)\mu\nu}| = {}_{-1}\langle\widetilde{k/2}|\,{}_{-1}\langle k/2|\frac{1}{\sqrt{2}}(\psi^{\nu}_{\frac{1}{2}}\tilde{\psi}^{\mu}_{\frac{1}{2}} - \psi^{\mu}_{\frac{1}{2}}\tilde{\psi}^{\nu}_{\frac{1}{2}}). \tag{5.6}$$

Since they all contain the following structure

$${}_{-1}\langle\widetilde{k/2}|\,{}_{-1}\langle k/2|\psi^{\nu}_{\frac{1}{2}}\tilde{\psi}^{\mu}_{\frac{1}{2}} \tag{5.7}$$

it is first convenient to compute the matrix element

$$T^{\mu\nu} \equiv {}_{-1}\langle\widetilde{k/2}|\,{}_{-1}\langle k/2|\psi^{\nu}_{\frac{1}{2}}\tilde{\psi}^{\mu}_{\frac{1}{2}}|D|B\rangle_{\mathrm{NS}} = -\frac{T_p}{2}\frac{V_{p+1}}{k_\perp^2}S^{\nu\mu} \tag{5.8}$$

where $k_\perp$ is the momentum in the transverse directions which is emitted by the brane. Note that the matrix $T^{\mu\nu}$ differs from the matrix $J^{\mu\nu}$ defined in equation (4.5) (computed for vanishing external field F) simply by the factor of $1/k_\perp^2$ coming from the insertion of the propagator.

Using this result and the explicit form of the dilaton projector (5.4), after some straighforward algebra, we find that the long-distance behaviour of the dilaton emitted by the Dp-brane is given by

$$\delta\phi \equiv \langle P^{(\phi)}|D|B\rangle_{\mathrm{NS}} = \frac{1}{2\sqrt{2}}(\eta^{\mu\nu} - k^\mu\ell^\nu - k^\nu\ell^\mu)T_{\mu\nu}. \tag{5.9}$$

Using the explicit expression for the matrix $T_{\mu\nu}$ we get

$$\delta\phi = \mu_p\frac{V_{p+1}}{k_\perp^2}\frac{3-p}{4} \tag{5.10}$$

where μ_p is the unit of R–R charge of a Dp-brane defined in equation (4.19). Similarly, using the projector (5.6) for the antisymmetric Kalb–Ramond field, we find

$$\delta A_{\mu\nu} \equiv \langle P^{(A)}_{\mu\nu}|D|B\rangle_{\mathrm{NS}} = \frac{1}{\sqrt{2}}(T_{\mu\nu} - T_{\nu\mu}). \tag{5.11}$$

Since, in our case, the matrix $T_{\mu\nu}$ is symmetric, we immediately conclude that the Kalb–Ramond field emitted by the pure Dp-brane is identically vanishing. Finally, using equation (5.5) we find that the components of the metric tensor are

$$\delta h_{\mu\nu} \equiv \langle P^{(h)}_{\mu\nu}|D|B\rangle_{\mathrm{NS}} = \frac{1}{2}(T_{\mu\nu} + T_{\nu\mu}) - \frac{\delta\phi}{2\sqrt{2}}\eta_{\mu\nu} \tag{5.12}$$

which explicitly read:

$$\delta h_{00} = -\delta h_{11} = \cdots = -\delta h_{pp} = \mu_p \frac{V_{p+1}}{k_\perp^2} \frac{7-p}{8\sqrt{2}}$$
$$\delta h_{p+1,p+1} = \cdots = \delta h_{99} = \mu_p \frac{V_{p+1}}{k_\perp^2} \frac{p+1}{8\sqrt{2}}.$$

Let us now turn to the R–R sector. In this case, after the insertion of the closed string propagator, we have to saturate the R–R boundary state (3.2) with the projector $\langle P^{(C)}|$ on the R–R massless field. Using equation (4.15) one can show that

$$\langle P^{(C)}_{\mu_1\ldots\mu_n}| = {}_{-1/2}\langle \widetilde{B}, \widetilde{k/2}|\,{}_{-3/2}\langle A, k/2| \left(C\Gamma_{\mu_1\ldots\mu_n} \frac{1-\Gamma_{11}}{2}\right)_{AB} \frac{(-1)^n}{2\sqrt{2}}. \quad (5.13)$$

This calculation is completely analogous to the one described in the previous section to obtain the couplings of a Dp-brane with the R–R potentials, the only new feature is the presence of the factor of $1/k_\perp^2$ produced by the closed string propagator. Due to the structure of the R–R component of the boundary state, it is not difficult to realize that the only projector of the form (5.13) that can give a non-vanishing result is the one corresponding to a $(p+1)$-form with all indices along the world-volume directions, and find that its long-distance behaviour is given by

$$\delta C_{01\ldots p} \equiv \langle P^{(C)}_{01\ldots p}|D|B\rangle_{\mathrm{R}} = -\mu_p \frac{V_{p+1}}{k_\perp^2}. \quad (5.14)$$

We can now rewrite the long-distance behaviour of the massless fields produced by the brane in a more suggestive way. First of all, we perform a Fourier transformation to work in configuration space. This is readily computed by observing that, for $p < 7$, one has

$$\int \mathrm{d}^{(p+1)}x\, \mathrm{d}^{(9-p)}y\, \frac{\mathrm{e}^{\mathrm{i}k_\perp\cdot y}}{(7-p)r^{7-p}\Omega_{8-p}} = \frac{V_{p+1}}{k_\perp^2} \quad (5.15)$$

where Ω_n is the area of a unit n-dimensional sphere defined in equation (2.40) and the radial coordinate r measures the distance from the branes. For later convenience, let us introduce the length L defined in equation (2.58) which we can also write as

$$L^{7-p} = \mu_p \frac{\sqrt{2}\kappa}{(7-p)\Omega_{8-p}}. \quad (5.16)$$

Then, using equation (5.10) and assuming that the dilaton has a vanishing vacuum expectation value, after some elementary steps, we obtain that the long-distance behaviour of the dilaton is

$$\phi = \sqrt{2}\kappa_{10}\varphi \simeq -\frac{p-3}{4}\frac{L^{7-p}}{r^{7-p}}. \quad (5.17)$$

Since we are going to compare our results with the standard supergravity description of D-branes described in chapter 2, we have reintroduced the field ϕ which differs from the canonically normalized dilaton φ by a factor of $\sqrt{2}\kappa_{10}$ (see also equation (4.10)). Similarly, recalling that $g_{\mu\nu} = \eta_{\mu\nu} + 2\kappa_{10} h_{\mu\nu}$, from equation (5.13) we find

$$g_{00} = -g_{11} = \cdots = -g_{pp} =\simeq -1 - \frac{(p-7)}{8}\frac{L^{7-p}}{r^{7-p}} \tag{5.18}$$
$$g_{p+1,p+1} = \cdots = g_{99} \simeq 1 + \frac{p+1}{8}\frac{L^{7-p}}{r^{7-p}}.$$

Finally, rescaling the R–R potential by a factor of $\sqrt{2}\kappa_{10}$ to obtain the standard supergravity normalization we easily get

$$C^{(p+1)} \simeq -\frac{L^{7-p}}{r^{7-p}}\, \mathrm{d}x^0 \wedge \ldots \wedge \mathrm{d}x^p. \tag{5.19}$$

Equations (5.17)–(5.19) represent the leading long-distance behaviour of the massless fields emitted by the Dp-brane. It is reasonable to expect that by inserting more boundaries on the closed string world-sheet, i.e. by introducing more boundary states, one can perturbatively reconstruct the exact brane solution of the supergravity field equations. Actually, this fact has been checked in [25], even if with a different formalism, and in [26] in the context of the non-BPS D-branes. We then assume that this is indeed what happens in general so that the exact solution can be written in terms of powers of the harmonic function

$$H(r) = 1 + \frac{L^{7-p}}{r^{7-p}}. \tag{5.20}$$

which exactly agrees with the one defined in equation (2.57). Under this assumption, from equations (5.17)–(5.19) we can infer that, in the exact solution corresponding to a Dp-brane, the dilaton is

$$\mathrm{e}^{\phi} = H^{(3-p)/4} \tag{5.21}$$

the metric is

$$\mathrm{d}s^2 = H^{(p-7)/8}[-(\mathrm{d}x^0)^2 + (\mathrm{d}x^1)^2 + \cdots + (\mathrm{d}x^p)^2] + H^{(p+1)/8}[(\mathrm{d}x^{p+1})^2 + \cdots + (\mathrm{d}x^9)^2] \tag{5.22}$$

and finally the R–R potential is

$$C_{(p+1)} = (H^{-1} - 1)\, \mathrm{d}x^0 \wedge \ldots \wedge \mathrm{d}x^p. \tag{5.23}$$

In writing this solution we have assumed that all fields except the metric have vanishing asymptotic values. This explains why we have subtracted the one in the last three equations. The solution obtained from the boundary state exactly agrees with the one derived in chapter 2 by solving the classical supergravity field equations (see equations (2.54)–(2.56)).

We can, therefore, conclude that the boundary state provides the complete conformal description of the D-branes of string theory; in fact it generates the correct D-brane effective action and reduces to the classical D-brane solution when it is projected onto the massless states of the closed string spectrum.

References

[1] For a review see, for example, Harvey J A Trieste HEP Cosmology 1995:66–125 (QCD161:W626:1995).

[2] Hull C H and Townsend P K 1995 *Nucl. Phys.* B **438** 109
Witten E 1995 *Nucl. Phys.* B **443** 85

[3] For a review see, for example, Schwarz J H 1997 *Nucl. Phys. Proc. Suppl.* B **55** 1
Townsend P K 1996 Four lectures on M theory *High Energy Physics and Cosmology (Trieste, Italy, 10 June–26 July 1996)* ed E Gava *et al* (Singapore: World Scientific) pp 385–438

[4] Duff M J, Khuri R R and Lu J X 1995 *Phys. Rep.* **259** 213

[5] Polchinski J 1995 *Phys. Rev. Lett.* **75** 4724

[6] Polchinski J, Chaudhuri S and Johnson C V 1996 Notes on D-branes *Preprint* hep-th/9602052
Polchinski J 1996 TASI lectures on D-branes *Preprint* hep-th/9611050
Bachas C 1999 Lectures on D-branes *Duality and Supersymmetric Theories* ed D Olive and P C West (Cambridge: Cambridge University Press) pp 414–73
Johnson C V 2000 D-brane primer *Preprint* hep-th/0007170

[7] Cremmer E and Scherk J 1972 *Nucl. Phys.* B **50** 222
Clavelli L and Shapiro J 1973 *Nucl. Phys.* B **57** 490
Ademollo M *et al* 1975 *Nucl. Phys.* B **94** 221

[8] Callan C G, Lovelace C, Nappi C R and Yost S A 1987 *Nucl. Phys.* B **288** 525
Callan C G, Lovelace C, Nappi C R and Yost S A 1987 *Nucl. Phys.* B **293** 83
Callan C G, Lovelace C, Nappi C R and Yost S A 1988 *Nucl. Phys.* B **308** 221

[9] Green M B 1992 *Nucl. Phys.* B **381** 201
Green M B and Wai P 1994 *Nucl. Phys.* B **431** 131
Green M B and Gutperle M 1996 *Nucl. Phys.* B **476** 484

[10] Di Vecchia P, Frau M, Pesando I, Sciuto S, Lerda A and Russo R 1997 *Nucl. Phys.* B **507** 259

[11] Billó M, Di Vecchia P, Frau M, Lerda A, Pesando I, Russo R and Sciuto S 1988 *Nucl. Phys.* B **526** 199
Billó M, Di Vecchia P, Frau M, Lerda A, Russo R and Sciuto S 1998 *Mod. Phys. Lett.* A **13** 2977

[12] Di Vecchia P, Frau M, Lerda A and Liccardo A 2000 *Nucl. Phys.* B **565** 397

[13] Di Vecchia P and Liccardo A 1999 D-branes in string theory. 1 *Preprint* hep-th/9912161
Di Vecchia P and Liccardo A 1999 D-branes in string theory. 2 *Preprint* hep-th/9912275

[14] Maldacena J 1998 *Adv. Theor. Math. Phys.* **2** 231
[15] Gubser S S, Klebanov I and Polyakov A M 1998 *Phys. Lett.* B **428** 105
Witten E 1998 *Adv. Theor. Math. Phys.* **2** 253
[16] For reviews on the AdS/CFT correspondence see, for example,
Petersen J L 1999 *Int. J. Mod. Phys.* A **14** 3597
Aharony O, Gubser S S, Maldacena J, Ooguri H and Oz Y 2000 *Phys. Rep.* **323** 183
Di Vecchia P 1999 *Preprint* hep-th/9908148
[17] Stelle K S 1998 BPS branes in supergravity *Preprint* hep-th/9803116
[18] Argurio R 1998 Brane physics in M theory *Preprint* hep-th/9807171
[19] Schwarz J H 1995 *Phys. Lett.* B **360** 13 (erratum 1995 *Phys. Lett.* B **364** 252)
[20] Polchinski J and Cai Y 1988 *Nucl. Phys.* B **296** 91
[21] Frau M, Pesando I, Sciuto S, Lerda A and Russo R 1997 *Phys. Lett.* B **400** 52
[22] Friedan D, Martinec E and Shenker S 1986 *Nucl. Phys.* B **271** 93
[23] Billó M, Cangemi D and Di Vecchia P 1997 *Phys. Lett.* B **400** 63
[24] Lu J X and Roy S 1999 *J. High Energy Phys.* **9908** 002
Lu J X and Roy S 1999 *Nucl. Phys.* B **560** 181
[25] Berenstein D and Leigh R G 1999 *Phys. Rev.* D **60** 106002
[26] Bertolini M, Di Vecchia P, Frau M, Lerda A, Marotta R and Russo R 2000 *Nucl. Phys.* B **590** 471

PART 2

PHYSICAL ASPECTS

Yassen Stanev, César Gómez and Pedro Resco

This page intentionally left blank

Chapter 6

Two-dimensional conformal field theory on open and unoriented surfaces

Yassen S Stanev
Dipartimento di Fisica, Università di Roma 'Tor Vergata'
INFN—Sezione di Roma 'Tor Vergata'
Via della Ricerca Scientifica 1, I-00133 Roma

6.1 Introduction

Two-dimensional conformal field theory (CFT) on open and unoriented surfaces is not a recent discovery. Its systematic study began in two seemingly different developments. On the one hand, the implications of the presence of a boundary in two-dimensional systems and the corresponding boundary conditions and boundary fields were first analysed by Cardy [1] and further in [2, 3]. On the other hand, a general prescription for the systematic construction of open and unoriented string models from a given closed oriented string model was proposed by Sagnotti [4] and further elaborated in [5, 6]. However, it was only after the discovery of D-branes [7] that the topic attracted so much attention and a huge number of different models have been explicitly constructed (any list will be incomplete). A parallel development was the study of the general consistency conditions for the models and, in particular, of the compatibility conditions between the Klein bottle projection and the annulus partition function embodied by the Möbius strip projection. As often in two-dimensional conformal theories a rational completely solved model like the $SU(2)$ Wess–Zumino–Witten model provided a good playground for such an analysis and exhibited three interesting properties:

- for the diagonal models there is a standard solution which extends the Cardy ansatz for the annulus to the unoriented case [8];
- there may be several different Klein bottle projections corresponding to different spectra in the unoriented sector; and

- the annulus partition function satisfies a completeness condition (i.e. satisfies the chiral fusion algebra) [9].

The last property also extends to all other explicitly solved examples but a better understanding of the physical principle underlying the completeness condition in the general case, in particular in the framework of string theory where so far open and closed string completeness conditions appear rather asymmetrically, is still absent. Another important open problem is whether there will be new constraints on the unoriented sector coming from higher genus surfaces.

Two-dimensional CFT on surfaces with boundaries and crosscaps is a large and rapidly developing subject. The aim of these lectures is to give an introduction to the topic, hence we have chosen to present a self-contained exposition based on one relatively simple and completely solved example, namely the $SU(2)$ Wess–Zumino–Witten (WZW) model. Even so some aspects like the explicit realization of the models in terms of D-branes and orientifolds [10] and their geometry are not covered. Other important developments which have to be mentioned are the relations of boundary conformal theory to graph theory (for a review see [11]) and to topological field theory [12].

The material is organized as follows. In section 6.2 we review some general properties of two-dimensional CFT. In section 6.3 we derive explicit expressions for the 4-point functions in the $SU(2)$ WZW model, the corresponding exchange operators and fusion matrix. Section 6.4 is devoted to the derivation of the sewing constraints for the correlation functions on open and unoriented surfaces. In section 6.5 we analyse the partition functions and the consistency conditions they satisfy.

6.2 General properties of two-dimensional CFT

6.2.1 The stress–energy tensor in two dimensions

Let us begin by recalling the particular properties of the stress–energy tensor in two-dimensional CFT. It is useful to introduce together with the flat Minkowski space light-cone coordinates $x_\pm = x^0 \pm x^1$ the coordinates on the cylindric space $\mathbb{S}^1 \times \mathbb{R}^1$ (on which the conformal transformations are well defined globally [13]) $t_\pm = \xi^0 \pm \xi^1$. Here ξ^0 is the non-compact time variable on the cylinder, while ξ^1 is the compact space variable ($\xi^1 + 2\pi$ is identified with ξ^1). We shall also use the analytic picture on the compact space $\mathbb{S}^1 \times \mathbb{S}^1$ with coordinates

$$z = \mathrm{e}^{\mathrm{i}t_-} \qquad \bar{z} = \mathrm{e}^{\mathrm{i}t_+} \tag{6.1}$$

where the complex variables z and $\bar{z}$ are obtained from the Minkowski light-cone coordinates by a Cayley transform

$$z = \frac{1 + \frac{\mathrm{i}}{2}x_-}{1 - \frac{\mathrm{i}}{2}x_-} \qquad \bar{z} = \frac{1 + \frac{\mathrm{i}}{2}x_+}{1 - \frac{\mathrm{i}}{2}x_+}. \tag{6.2}$$

Note that z and $\bar{z}$ are complex conjugate only if one starts from the Euclidean picture where ξ^0 is purely imaginary, while ξ^1 is real. Nonlinear transformations of the coordinates, like (6.2), require non-trivial accompanying changes of the field variables. To find the transformation law for the stress–energy tensor let us first write its components in the light-cone basis $x_\pm$

$$\Theta_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = \Theta_{++}\,\mathrm{d}x_+^2 + \Theta_{+-}\,\mathrm{d}x_+\,\mathrm{d}x_- + \Theta_{-+}\,\mathrm{d}x_-\,\mathrm{d}x_+ + \Theta_{--}\,\mathrm{d}x_-^2 \tag{6.3}$$

where

$$\begin{aligned}\Theta_{++} &= \tfrac{1}{4}(\Theta_{00} + \Theta_{10} + \Theta_{01} + \Theta_{11})\\ \Theta_{--} &= \tfrac{1}{4}(\Theta_{00} - \Theta_{10} - \Theta_{01} + \Theta_{11})\\ \Theta_{+-} &= \Theta_{-+} = \tfrac{1}{4}(\Theta_{00} - \Theta_{11}).\end{aligned}$$

The energy density with our choice of metric

$$\eta_{\mu\nu} = \mathrm{diag}(-,+) \tag{6.4}$$

is given by $\Theta_0^0 = -\Theta_{00}$, so let us choose the three independent components of $\Theta_{\mu\nu}$ as

$$\Theta = -\Theta_{--} \qquad \bar{\Theta} = -\Theta_{++} \qquad \Theta_0 = -\Theta_{+-} = \tfrac{1}{4}\,\mathrm{Tr}\,\Theta. \tag{6.5}$$

The conservation of the stress–energy tensor $\partial_\mu \Theta^{\mu\nu} = 0$ then implies

$$\partial_+\Theta = -\partial_-\Theta_0 \qquad \partial_-\bar{\Theta} = -\partial_+\Theta_0 \tag{6.6}$$

where $\partial_\pm = 1/2(\partial_0 \pm \partial_1)$. The corresponding fields in the analytic picture are

$$\begin{aligned}T(z,\bar{z}) &= 2\pi\left(\mathrm{i}\frac{\partial x_-}{\partial z}\right)^2 \Theta(x_+(\bar{z}), x_-(z))\\ \bar{T}(z,\bar{z}) &= 2\pi\left(\mathrm{i}\frac{\partial x_+}{\partial \bar{z}}\right)^2 \bar{\Theta}(x_+(\bar{z}), x_-(z))\\ T_0(z,\bar{z}) &= 2\pi\left(\mathrm{i}\frac{\partial x_-}{\partial z}\right)\left(\mathrm{i}\frac{\partial x_+}{\partial \bar{z}}\right) \Theta_0(x_+(\bar{z}), x_-(z)).\end{aligned} \tag{6.7}$$

The conservation of Θ leads to the equations

$$\bar{\partial}T = -\partial T_0 \qquad \partial\bar{T} = -\bar{\partial}T_0 \qquad \left(\partial = \frac{\partial}{\partial z}, \bar{\partial} = \frac{\partial}{\partial \bar{z}}\right). \tag{6.8}$$

Thus if the stress–energy tensor is traceless ($T_0 = 0$), each of the two components T and $\bar{T}$ depends on a single variable $T = T(z)$ and $\bar{T} = \bar{T}(\bar{z})$.

A similar separation in chiral and antichiral components is valid also for an Abelian current j_μ that is conserved together with its dual

$$\partial_\mu j^\mu = 0 = \partial^\mu \epsilon_{\mu\nu} j^\nu. \tag{6.9}$$

We shall call such fields which split into chiral and antichiral components local observables. In other words, one can define the two-dimensional CFT as a quantum field theory in which the observable algebra is a tensor product of two algebras

$$\mathcal{A} \otimes \bar{\mathcal{A}}. \tag{6.10}$$

The chiral (or analytic) algebra $\mathcal{A}$ and the antichiral (or antianalytic) algebra $\bar{\mathcal{A}}$ are related by space reflection. For the rest of these lectures we shall assume that $\mathcal{A}$ and $\bar{\mathcal{A}}$ are isomorphic. The algebra $\mathcal{A}$ is generated by a finite number of local fields $O_n(z)$. It should be stressed that this condition does not lead necessarily to a finite number of fields in the theory. Locality implies that all $O_n(z)$ mutually commute for different arguments, more precisely for any given n and m there exists an integer $N_0(n, m)$ such that for all $N \geq N_0$

$$(z_1 - z_2)^N [O_n(z_1), O_m(z_2)] = 0. \tag{6.11}$$

The general solution of this equation is given by a linear combination of the δ-function and its derivatives

$$[O_n(z_1), O_m(z_2)] = \sum_{\ell=0}^{N_0-1} C_\ell(z_2)\delta^{(\ell)}(z_{12}) \tag{6.12}$$

where δ on the unit circle can be defined as

$$\begin{aligned} \delta(z_{12}) &= \frac{1}{z_1} \sum_n \left(\frac{z_2}{z_1}\right)^n \\ &= \frac{1}{z_1} \sum_{n=0}^{\infty} \left(\frac{z_2}{z_1}\right)^n + \frac{1}{z_2} \sum_{n=0}^{\infty} \left(\frac{z_1}{z_2}\right)^n \end{aligned} \tag{6.13}$$

and satisfies

$$\oint \delta(z_{12}) f(z_2) \frac{\mathrm{d}z_2}{2\pi \mathrm{i}} = f(z_1). \tag{6.14}$$

For the currents (of scale dimension 1) and for the stress–energy tensor (of scale dimension 2) this leaves undetermined only one constant. In particular [14],

$$[T(z_1), T(z_2)] = -\frac{c}{12}\delta'''(z_{12}) - \delta'(z_{12})(T(z_1) + T(z_2)) \tag{6.15}$$

where the constant c is called the central charge. The same relation also holds for the antichiral component $\bar{T}$ with central charge $\bar{c}$ ($= c$ due to the assumption that $\mathcal{A}$ and $\bar{\mathcal{A}}$ are isomorphic). All fields from $\mathcal{A}$ commute with all fields from

$\bar{\mathcal{A}}$, hence $T(z)$ and $\bar{T}(\bar{z})$ commute. Under a general analytic reparametrization $z \to w(z)$ the stress–energy tensor transforms according to

$$T(z) \to T(w) = \left(\frac{\partial z}{\partial w}\right)^2 T(z(w)) + \frac{c}{12}\{w, z\} \tag{6.16}$$

where $\{w, z\}$ is the Schwartz derivative

$$\{w, z\} = \frac{w'''}{w'} - \frac{3}{2}\left(\frac{w''}{w'}\right)^2. \tag{6.17}$$

The central term in (6.15), (6.16) is related to the conformal anomaly. $T(z)$ has a Laurent expansion of the form

$$T(z) = \sum_n \frac{L_n}{z^{n+2}} \tag{6.18}$$

where the modes L_n are given by

$$L_n = \frac{1}{2\pi \mathrm{i}} \oint_{S^1} \mathrm{d}z\, T(z) z^{n+1}. \tag{6.19}$$

The commutator (6.15) for the chiral components of the stress–energy tensor implies for the modes L_n the commutation relations of the Virasoro algebra Vir [15], that

$$[L_n, L_m] = (n - m)L_{n+m} + \frac{c}{12} n(n^2 - 1)\delta_{n+m} \tag{6.20}$$

where δ_ℓ denotes the Kronecker symbol $\delta_{\ell,0}$. The central term in (6.20) vanishes for $n = 0, \pm 1$. The corresponding subalgebra generated by L_{-1}, L_0 and L_1 is $SL(2, \mathbb{R})$. The unique vacuum vector $|0\rangle$ is annihilated by L_{-1}, L_0 and L_1 (and by their antichiral counterparts):

$$L_{0,\pm 1}|0\rangle = 0 = \bar{L}_{0,\pm 1}|0\rangle. \tag{6.21}$$

The Hermiticity of the stress–energy tensor gives for the modes

$$L_n^\dagger = L_{-n}. \tag{6.22}$$

Not all the fields in the theory split into chiral and antichiral parts. In particular, there exist ‘primary’ conformal fields [14, 16], of conformal weights Δ and $\bar{\Delta}$ which, under reparametrizations $z \to w(z)$, $\bar{z} \to \bar{w}(\bar{z})$, transform as

$$\phi_{\Delta\bar{\Delta}}(z, \bar{z}) \to \phi_{\Delta\bar{\Delta}}(w, \bar{w}) = \left(\frac{\partial z}{\partial w}\right)^{\Delta} \left(\frac{\partial \bar{z}}{\partial \bar{w}}\right)^{\bar{\Delta}} \phi_{\Delta\bar{\Delta}}(z(w), \bar{z}(\bar{w})). \tag{6.23}$$

This transformation law implies the following commutation relations between the primary fields and the generators of the Virasoro algebra L_n

$$[L_n, \phi_{\Delta\bar{\Delta}}(z, \bar{z})] = z^n(z\partial_z + (n+1)\Delta)\phi_{\Delta\bar{\Delta}}(z, \bar{z}) \tag{6.24}$$
$$[\bar{L}_n, \phi_{\Delta\bar{\Delta}}(z, \bar{z})] = \bar{z}^n(\bar{z}\partial_{\bar{z}} + (n+1)\bar{\Delta})\phi_{\Delta\bar{\Delta}}(z, \bar{z}). \tag{6.25}$$

The corresponding states obtained by acting with the primary fields on the vacuum are also called primary:

$$|\Delta, \bar{\Delta}\rangle = \phi_{\Delta\bar{\Delta}}(0, 0)|0\rangle. \tag{6.26}$$

They are annihilated by all the generators L_n with $n > 0$:

$$L_n|\Delta, \bar{\Delta}\rangle = \bar{L}_n|\Delta, \bar{\Delta}\rangle = 0 \qquad \text{for } n > 0. \tag{6.27}$$

The conformal dimension of a primary field is equal to the sum of its two conformal weights, while its spin (or helicity) is equal to their difference:

$$d = \Delta + \bar{\Delta} \qquad s = \Delta - \bar{\Delta}. \tag{6.28}$$

There also exist fields that satisfy (6.24), (6.25) only for $n = 0, \pm 1$. Such fields are called quasiprimary (or conformal descendants). The corresponding quasiprimary states are obtained from the primary states (6.26) by the action of polynomials in L_n with negative n. All the properties of the quasiprimary fields follow from those of the underlying primary one.

6.2.2 Rational conformal field theories

One important class of theories are the rational conformal field theories (RCFTs). In an RCFT there are only a finite number of primary fields. For example, in the unitary minimal models [14, 16] corresponding to the central charge of the Virasoro algebra,

$$c = 1 - \frac{6}{m(m+1)} \qquad m \geq 3 \tag{6.29}$$

the primary fields have weights

$$\Delta_{r,s} = \frac{[r(m+1) - sm]^2 - 1}{4m(m+1)} \qquad 1 \leq r \leq m-1,\ 1 \leq s \leq m. \tag{6.30}$$

Another important example are the superconformal models. The supersymmetry generator $G(z)$ has conformal weight $\frac{3}{2}$ and hence a Laurent expansion

$$G(z) = \sum_r \frac{G_r}{z^{r+\frac{3}{2}}}. \tag{6.31}$$

Since $G(z)$ has half-integer spin, it can be chosen to be either periodic (Ramond sector) or antiperiodic (Neveu-Schwarz sector) [17]. In the Ramond sector the sum in (6.31) is over r integer, while in the Neveu-Schwarz sector it is over r half-integer. The (anti)commutation relations between L_n and G_r are

$$[L_n, G_r] = \left(\frac{n}{2} - r\right) G_{n+r} \tag{6.32}$$

$$\{G_r, G_s\} = 2L_{r+s} + \frac{c}{3}\left(r^2 - \frac{1}{4}\right)\delta_{r+s}. \tag{6.33}$$

The unitary $N = 1$ superconformal models have central charge

$$c = \frac{3}{2}\left[1 - \frac{8}{m(m+2)}\right] \qquad m \geq 3 \tag{6.34}$$

while the conformal weights of the primary fields are [18, 19]

$$\Delta_{r,s} = \frac{[r(m+2) - sm]^2 - 4}{8m(m+2)} + \frac{1}{32}[1 - (-1)^{r-s}] \tag{6.35}$$

where $1 \leq r \leq m - 1$ and $1 \leq s \leq m$. The Neveu-Schwarz sector contains the fields with $r - s$ even, while the Ramond sector contains the fields with $r - s$ odd.

In order to describe the $N = 2$ superconformal models [20] it is convenient to study first the simplest example of a conformal current algebra, namely the Abelian $U(1)$ case. The chiral part of the $U(1)$ current satisfying (6.9) has the following expansion in Laurent modes:

$$J(z) = \sum_n \frac{J_n}{z^{n+1}} \qquad J_n^\dagger = J_{-n}. \tag{6.36}$$

Since the $U(1)$ current is a primary field of the Virasoro algebra of weight one, its commutation relations with the modes of the stress–energy tensor are

$$[L_n, J_m] = -mJ_{m+n}. \tag{6.37}$$

The locality condition (6.12) also completely determines the commutation relations between two currents

$$[J(z_1), J(z_2)] = -\delta'(z_{12}) \qquad \text{or} \qquad [J_n, J_m] = n\delta_{n+m} \tag{6.38}$$

where, for convenience, we have chosen to normalize the central term to one. The same relations also hold for the antichiral components. The primary fields of the $U(1)$ conformal current algebra are characterized by their charges q and $\bar{q}$ and satisfy the following commutation relations with the current components:

$$[J(z_1), \phi_{q\bar{q}}(z_2, \bar{z}_2)] = -q\phi_{q\bar{q}}(z_2, \bar{z}_2)\delta(z_{12}) \tag{6.39}$$

$$[\bar{J}(\bar{z}_1), \phi_{q\bar{q}}(z_2, \bar{z}_2)] = -\bar{q}\phi_{q\bar{q}}(z_2, \bar{z}_2)\delta(\bar{z}_{12}). \tag{6.40}$$

The stress–energy tensor can be expressed in terms of the currents by the Sugawara formula [21] and the central charge of the Virasoro algebra is equal to one

$$T(z) = \tfrac{1}{2} : J^2(z) : \quad \Rightarrow \quad c(u(1)) = 1 \tag{6.41}$$

which, for the Laurent modes, gives

$$L_n = \frac{1}{2}\left(\sum_{m \geq 1} + \sum_{m \geq -n}\right) J_{-m} J_{m+n}. \tag{6.42}$$

The consistency of equations (6.39), (6.42) and (6.24) implies a relation between the $U(1)$ charges and the conformal weights:

$$\Delta = \tfrac{1}{2} q^2 \qquad \bar{\Delta} = \tfrac{1}{2} \bar{q}^2 \tag{6.43}$$

as well as the following equations for the primary fields [22, 23]:

$$\partial_z \phi_{q\bar{q}}(z, \bar{z}) + q : J(z) \phi_{q\bar{q}}(z, \bar{z}) := 0 \tag{6.44}$$

$$\partial_{\bar{z}} \phi_{q\bar{q}}(z, \bar{z}) + \bar{q} : \bar{J}(\bar{z}) \phi_{q\bar{q}}(z, \bar{z}) := 0. \tag{6.45}$$

The $N = 2$ superconformal algebra contains two supersymmetry generators $G^\alpha(z)$, $\alpha = 1, 2$, with Laurent expansions (6.31) and a $U(1)$ current $J(z)$ with expansion (6.36). The new (anti)commutation relations are

$$\{G_r^\alpha, G_s^\beta\} = 2\delta^{\alpha\beta} L_{r+s} + \mathrm{i}(r - s)\epsilon^{\alpha\beta} J_{r+s} + \frac{c}{3}\left(r^2 - \frac{1}{4}\right)\delta^{\alpha\beta}\delta_{r+s} \tag{6.46}$$

$$[J_m, G_r^\alpha] = \mathrm{i}\epsilon^{\alpha\beta} G_r^\beta \tag{6.47}$$

where $\epsilon^{\alpha\beta}$ is antisymmetric and $\epsilon^{12} = 1$. There are three sectors: in the Neveu-Schwarz and Ramond sectors the $U(1)$ current has integer modes, while in the twisted sector the $U(1)$ current has half-integer modes [24]. The unitary minimal $N = 2$ superconformal models correspond to central charges

$$c = 3\left(1 - \frac{2}{m}\right) \qquad m \geq 3. \tag{6.48}$$

6.2.3 Non-Abelian conformal current algebras

The non-Abelian generalization of the $U(1)$ conformal current algebra (6.38) known also as the Wess–Zumino–Witten (WZW) model is one of the few cases of two-dimensional CFT for which one can also write an explicit action [25]. Alternatively, one can use the following definition. Let G be a compact semi-simple Lie group and $\mathcal{G}$ be its Lie algebra of dimension d_G. The chiral conformal current algebra $\mathcal{A}(\mathcal{G})$ is the algebra generated by the d_G chiral currents in the adjoint representation of $\mathcal{G}$. The currents are primary fields of the Virasoro algebra of conformal weight one and have the Laurent expansion

$$J^a(z) = \sum_n \frac{J_n^a}{z^{n+1}} \qquad J_n^{a^*} = J_{-n}^a. \tag{6.49}$$

The commutation relations for their modes are

$$[J_n^a, J_m^b] = \mathrm{i} \sum_c f_{abc}\, J_{n+m}^c + \frac{k}{2} n \delta_{ab} \delta_{n+m} \tag{6.50}$$

where f_{abc} are the structure constants of $\mathcal{G}$ and the level k is a non-negative integer. These relations define an affine Kac–Moody algebra [26].

The stress–energy tensor can be expressed in terms of the currents (6.49) by the Sugawara formula

$$2h\ T(z) = \sum_{a=1}^{d_G} : J_a^2(z) : \tag{6.51}$$

where the height h is the sum of the level k and the dual Coxeter number of $\mathcal{G}$, $h = k + \check{g}\,(= k + N$ for $SU(N))$. In terms of the Laurent modes, (6.51) becomes

$$2hL_n = \left(\sum_{\ell=1}^{\infty} + \sum_{\ell=-n}^{\infty} \right) \sum_{a=1}^{d_G} J_{-\ell}^a J_{n+\ell}^a \tag{6.52}$$

while the central charge of the Virasoro algebra is

$$c = \frac{k}{h} d_G. \tag{6.53}$$

The primary fields of $\mathcal{A}(\mathcal{G})$ are in one-to-one correspondence with the irreducible representations $\mathcal{G}$, hence we can label them by highest weight vectors $\Lambda = (\lambda_1, \ldots, \lambda_r)$ of $\mathcal{G}$. We shall denote the primary fields by $V_\Lambda(z)$. They satisfy the following commutation relations with the currents (for brevity we omit the dependence on $\bar{z}$ and write only the relations in the chiral sector):

$$[J^a(z_1), V_\Lambda(z_2)] = \delta(z_{12}) V_\Lambda(z_2) t_\Lambda^a \tag{6.54}$$

or, in terms of the modes (6.49),

$$[J_n^a, V_\Lambda(z)] = z^n V_\Lambda(z) t_\Lambda^a \tag{6.55}$$

where t_Λ^a are the matrices of J_0^a in the representation Λ. The consistency of equations (6.52) and (6.55) with (6.24) implies the relation

$$2h\Delta_\Lambda = C_2(\Lambda) \tag{6.56}$$

between the conformal weight of the primary field and the eigenvalue of the second-order Casimir operator in the representation Λ, as well as the operator form of the Knizhnik–Zamolodchikov (KZ) equation [22, 23]

$$h\frac{\mathrm{d}}{\mathrm{d}z}V_\Lambda(z) = \sum_{a=1}^{d_G} : V_\Lambda(z)t^a_\Lambda J^a(z) : . \tag{6.57}$$

The primary fields in a two-dimensional conformal theory transforming as in (6.23) in general do not split in a sum of chiral and antichiral components. Rather they are given by a (finite in the case of a rational conformal theory) sum of products of chiral and antichiral vertex operators [27, 28]. In order to define a chiral vertex operator properly we have to specify a triple of weights $\begin{pmatrix} & \Lambda_f & \\ \Lambda & & \Lambda_i \end{pmatrix}$ where Λ_i is the weight on which V_Λ acts, while Λ_f is the weight to which V_Λ maps. In other words, the chiral vertex operators can be represented as

$$V_{\Lambda\ \Lambda_i}^{\ \Lambda_f}(z) = \Pi_{\Lambda_f} V_\Lambda(z) \Pi_{\Lambda_i} \tag{6.58}$$

where Π_Λ are orthogonal projectors and, in general, are multi-valued functions of z:

$$V_{\Lambda\ \Lambda_i}^{\ \Lambda_f}(\mathrm{e}^{2\pi\mathrm{i}}z) = \mathrm{e}^{2\pi\mathrm{i}(\Delta_{\Lambda_f}-\Delta_\Lambda-\Delta_{\Lambda_i})} V_{\Lambda\ \Lambda_i}^{\ \Lambda_f}(z). \tag{6.59}$$

The correlation functions of the chiral vertex operators are called chiral conformal blocks and, due to (6.59), are also multivalued functions of the coordinates. The two-dimensional primary fields $\phi(z, \bar{z})$ can be written in terms of the chiral vertex operators (6.58):

$$\phi_{\Lambda\bar{\Lambda}}(z, \bar{z}) = \sum_{\substack{\Lambda_i\ \bar{\Lambda}_i \\ \Lambda_f\ \bar{\Lambda}_f}} V_{\Lambda\ \Lambda_i}^{\ \Lambda_f}(z) \bar{V}_{\bar{\Lambda}\ \bar{\Lambda}_i}^{\ \bar{\Lambda}_f}(\bar{z}). \tag{6.60}$$

Locality and (6.59) imply that the spin of all fields $\Delta_\Lambda - \Delta_{\bar{\Lambda}}$ has to be integer. Note that this selection rule must also be respected by the pairs of weights $(\Lambda_i, \bar{\Lambda}_i)$ and $(\Lambda_f, \bar{\Lambda}_f)$. One large class of theories which trivially satisfy this requirement are the diagonal theories with $\Lambda = \bar{\Lambda}$.

6.2.4 Partition function, modular invariance

Due to the factorization of the observable algebra (6.10) we can analyse independently the chiral and antichiral sectors, but in order to reconstruct the whole two-dimensional theory we also need the pairings between the fields from the two sectors. They can be found by requiring the modular invariance of the partition function on the torus. From the viewpoint of string theory the modular invariance condition is very natural, since it ensures that one can define the theory on surfaces of arbitrary genus [30,31]. In statistical mechanics models its physical meaning is more subtle, since the modular transformations relate the low- and the high-temperature behaviour of the theory [32].

Let us briefly recall the construction of the partition function. To every primary field φ_i of $\mathcal{A}$ corresponds a character of the Virasoro algebra [26]:

$$\chi_i(\tau) = \mathrm{Tr}_{\mathcal{H}_i}\, \mathrm{e}^{2\pi \mathrm{i}\tau(L_0 - \frac{c}{24})} \tag{6.61}$$

where the trace is over the space of all quasiprimary descendants of φ_i. Note that the energy operator L_0 on the torus is modified according to (6.16). In this notation the torus partition function

$$Z_T = \mathrm{Tr}(\mathrm{e}^{2\pi \mathrm{i}\tau(L_0 - \frac{c}{24})}\mathrm{e}^{2\pi \mathrm{i}\bar{\tau}(\bar{L}_0 - \frac{\bar{c}}{24})}) \tag{6.62}$$

can be rewritten as (we recall that $\bar{c} = c$)

$$Z_T = \sum_{i,j} \chi_i X_{ij} \bar{\chi}_j \tag{6.63}$$

where X_{ij} are non-negative integers which give the multiplicities of the two-dimensional fields. For the rational theories the sum in (6.63) is over a finite set of characters.

Not all values of τ in (6.62) correspond to inequivalent tori. In particular, the transformations

$$S : \tau \longrightarrow -\frac{1}{\tau} \tag{6.64}$$

$$T : \tau \longrightarrow \tau + 1 \tag{6.65}$$

are just redefinitions of the fundamental cell of the torus. They generate the modular group $PSL(2, \mathbb{Z})$ under which τ transforms as

$$\tau \longrightarrow \tau' = \frac{a\tau + b}{c\tau + d} \qquad ad - bc = 1 \tag{6.66}$$

with integer a, b, c and d. These transformations act linearly on the characters (6.61)

$$\chi_i\left(-\frac{1}{\tau}\right) = \sum_j S_{ij}\chi_j(\tau) \qquad \chi_i(\tau+1) = \sum_j T_{ij}\chi_j(\tau) \tag{6.67}$$

where T is a diagonal matrix, while S is a symmetric matrix. Both S and T are unitary and satisfy $S^2 = (ST)^3 = C$, where the matrix C is called the charge conjugation matrix and satisfies $C^2 = 1$.

The modular invariance of the torus partition function implies

$$SXS^\dagger = X \qquad TXT^\dagger = X. \tag{6.68}$$

The solutions to these equations are of two distinct types [33]. The first ones are called permutation (or automorphism) invariants, for which

$$X_{ij} = \delta_{i\sigma(j)} \tag{6.69}$$

where $\sigma(j)$ is a permutation of the labels j. The second ones correspond to extensions of the observable algebra and can always be rewritten as a permutation invariant (6.69) in terms of the characters of the maximally extended observable algebra (that are linear combinations of the characters of the unextended one).

Let us denote by $[\varphi_i]$ the conformal family of the primary field φ_i, i.e. the collection of all the conformal descendants of φ_i. The product of two conformal families is determined by the fusion algebra

$$[\varphi_i] \times [\varphi_j] = \sum_k N_{ij}{}^k [\varphi_k]. \tag{6.70}$$

The non-negative integers $N_{ij}{}^k$, called fusion rules, can be expressed in terms of the modular matrix S by the Verlinde formula:

$$N_{ij}{}^k = \sum_\ell \frac{S_{i\ell} S_{j\ell} S^\dagger_{k\ell}}{S_{1\ell}} \tag{6.71}$$

and, as matrices, $(N_i)_j{}^k$ satisfy the commutative and associative fusion algebra [34]

$$(N_i)(N_j) = \sum_k N_{ij}{}^k (N_k). \tag{6.72}$$

There are several known classifications of modular-invariant partition functions, e.g. [35–38], but the problem is still not solved in general. We shall often refer to the A–D–E classification of Cappelli, Itzykson and Zuber [35] of the modular invariants of the $SU(2)$ conformal current algebra. In this classification, the diagonal A and the D_{odd} series are permutation invariants, the D_{even} series, E_6 and E_8 are diagonal invariants of an extended algebra, while E_7 is a non-trivial permutation invariant of an extended algebra.

There is also an alternative method to compute the allowed pairings between the fields of the two sectors that makes no use of higher-genus partition functions. In two-dimensional CFT the product of two primary fields can be expressed as a sum of primary fields and their conformal descendants using the Operator Product Expansion (OPE):

$$\begin{aligned}
&\phi_{\Delta_i,\bar{\Delta}_i}(z,\bar{z})\phi_{\Delta_j,\bar{\Delta}_j}(w,\bar{w}) \\
&\quad = \sum_{k,\bar{k}} \frac{C^{(k,\bar{k})}_{(i,\bar{\imath})(j,\bar{\jmath})}}{(z-w)^{\Delta_i+\Delta_j-\Delta_k}(\bar{z}-\bar{w})^{\bar{\Delta}_i+\bar{\Delta}_j-\bar{\Delta}_k}} \phi_{\Delta_k,\bar{\Delta}_k}(w,\bar{w}) + \cdots
\end{aligned} \tag{6.73}$$

where the dots stand for the descendants. The two-dimensional structure constants $C_{(i,\bar{i})(j,\bar{j})}^{(k,\bar{k})}$ vanish whenever the corresponding fusion rules $N_{ij}{}^{k}$ or $N_{\bar{i}\bar{j}}{}^{\bar{k}}$ are zero and completely define the theory. In particular, they determine also the allowed pairings between the fields of the two sectors. Moreover they permit to reconstruct all the Green functions of the two-dimensional fields. In RCFTs the structure constants can, in principle, be computed imposing the locality (or crossing symmetry) of the 4-point Green functions. Indeed for a generic 4-point function

$$\langle \phi_{\Delta_1,\bar{\Delta}_1}(z_1,\bar{z}_1)\phi_{\Delta_2,\bar{\Delta}_2}(z_2,\bar{z}_2)\phi_{\Delta_3,\bar{\Delta}_3}(z_3,\bar{z}_3)\phi_{\Delta_4,\bar{\Delta}_4}(z_4,\bar{z}_4)\rangle \tag{6.74}$$

we can apply the OPE (6.73) in three different ways which schematically can be denoted as (12)(34), (13)(24) and (14)(23). This gives two duality relations between the structure constants and determines them up to global rescalings of the two-dimensional fields. In practice, this procedure is very complicated and the closed expressions for the two-dimensional structure constants are known only in a very limited number of cases (in particular for the $SU(2)$ current algebra models and for the unitary minimal models [39, 40]).

Let us stress that while the crossing symmetry relations are also satisfied for any subset of primary fields closed under OPE, e.g. for the identity operator alone to give a trivial example, the modular invariance condition is satisfied only by the maximal (or complete) set of fields.

In fact, these two approaches are complementary, since, as demonstrated in [28, 41], both the condition of crossing symmetry of the 4-point functions and the modular invariance of the torus partition function are necessary and sufficient for the consistency of the theory on a surface of arbitrary genus.

6.3 Correlation functions in current algebra models

In the conformal current algebra models the operator Knizhnik–Zamolodchikov equation (6.57) implies a system of first-order partial differential equations for the n-point chiral conformal blocks. This allows one to reformulate all the properties of the primary conformal fields as conditions on their chiral correlators. Moreover, for the $SU(2)$ models that we shall review in some detail this also allows us to obtain explicit expressions for the chiral conformal blocks and to compute the structure constants that enter the two-dimensional operator product expansion (6.73).

6.3.1 Properties of the chiral conformal blocks

Let G be a simply connected compact Lie group with Lie algebra $\mathcal{G}$ and let $V_i = V(\Lambda_i)$, $i = 1, 2, \ldots, n$ be chiral vertex operators of highest weight Λ_i such that the space $\mathcal{J}_n = \mathcal{J}(\Lambda_1, \ldots, \Lambda_n)$ of G invariant tensors is non-trivial ($d_{\mathcal{J}} = \dim \mathcal{J}_n > 0$). Consider the $d_{\mathcal{J}}$-dimensional vector space $\mathcal{L}_n$

of holomorphic functions $w_n = w(z_1, \Lambda_1; \ldots; z_n, \Lambda_n)$ called chiral conformal blocks [14] with values in $\mathcal{J}_n$.

Möbius invariance of the vacuum implies that the functions w_n are covariant under local Möbius transformations. In particular, they are translation invariant (hence depend only on the differences z_{ij}), they transform covariantly under uniform dilations $z_i \to \rho z_i$, $\rho > 0$

$$\rho^{\Delta_1+\cdots+\Delta_n} w(\rho z_1, \Lambda_1; \ldots; \rho z_n, \Lambda_n) = w(z_1, \Lambda_1; \ldots; z_n, \Lambda_n) \tag{6.75}$$

where $\Delta_i = \Delta(\Lambda_i)$ are the conformal weights (6.56). Finally, w_n are covariant under infinitesimal special conformal transformations $z \to z/(1+\varepsilon z)$ with $\varepsilon \to 0$, thus satisfy the differential equation

$$\sum_{i=1}^{n} z_i \left(z_i \frac{\partial}{\partial z_i} + 2\Delta_i \right) w_n = 0. \tag{6.76}$$

The operator form of the Knizhnik–Zamolodchikov equation (6.57) implies that all elements in $\mathcal{L}_n$ satisfy the system of partial differential equations [22]

$$\left(\frac{\partial}{\partial z_i} + \frac{1}{h} \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{\sum_a t^a_{\Lambda_i} t^a_{\Lambda_j}}{z_{ij}} \right) w_n = 0 \tag{6.77}$$

for $i = 1, \ldots, n$, where h is the height defined after equation (6.51).

Every function w_n of $\mathcal{L}_n$ admits a path-dependent multivalued analytic continuation in the product of complex planes minus the diagonal $\{z_i \in \mathbb{C}, z_i \neq z_j \text{ for } i \neq j\}$. Let us choose a basis $\{w_n^\nu, \nu = 1, \ldots, d_{\mathcal{J}}\}$ in $\mathcal{L}_n$ and consider the analytic continuation of w_n^ν along a pair of paths $\mathcal{C}_i^\pm$ that exchange two neighbouring arguments z_i, z_{i+1} in positive/negative directions:

$$\mathcal{C}_i^\pm : \begin{pmatrix} z_i \\ z_{i+1} \end{pmatrix} \to \frac{1}{2}(z_i + z_{i+1}) + \frac{1}{2}\begin{pmatrix} z_{ii+1} \\ -z_{ii+1} \end{pmatrix} \mathrm{e}^{\pm \mathrm{i}\pi t} \tag{6.78}$$

where $0 \le t \le 1$. This operation followed by the permutation of the two weights Λ_i and Λ_{i+1} defines the action of two exchange operators B_i and $\bar{B}_i$ [27, 28, 42]. The exchange operator B_i transforms the basis $\{w_n^\nu\}$ in $\mathcal{L}(\Lambda_1, \ldots, \Lambda_i, \Lambda_{i+1}, \ldots, \Lambda_n)$ in a basis $\{w_n^\mu\}$ in $\mathcal{L}(\Lambda_1, \ldots, \Lambda_{i+1}, \Lambda_i, \ldots, \Lambda_n)$.

$$\begin{aligned} B_i = B_i^{\Lambda_1 \ldots \Lambda_n} : \mathcal{L}(\Lambda_1, \ldots, \Lambda_i,\ \Lambda_{i+1}, \ldots, \Lambda_n) \\ \to \mathcal{L}(\Lambda_1, \ldots, \Lambda_{i+1}, \Lambda_i, \ldots \Lambda_n). \end{aligned} \tag{6.79}$$

The exchange operator $\bar{B}_i$ is the inverse to B_i. More precisely,

$$\bar{B}_i^{\Lambda_1 \ldots \Lambda_{i+1} \Lambda_i \ldots \Lambda_n} B_i^{\Lambda_1 \ldots \Lambda_i \Lambda_{i+1} \ldots \Lambda_n} = \mathbf{1}. \tag{6.80}$$

For real analytic w_n^ν the matrix $\bar{B}_i$ is complex conjugate to B_i. The operators B_i, $i = 1, \ldots, n-1$ with various order of the weights $(\Lambda_1, \ldots, \Lambda_n)$ generate a representation of the exchange (called also braid [43]) algebra $\mathcal{B}_n$.

The two-dimensional n-point Green functions G_n can be written as a finite sum of products of n-point chiral and antichiral blocks

$$\begin{aligned} G_n &= \langle 0|\phi_1(z_1, \bar{z}_1)\ldots\phi_n(z_n, \bar{z}_n)|0\rangle \\ &= \bar{w}_n^\mu Q_{\mu\nu}^{\Lambda_1\ldots\Lambda_n} w_n^\nu. \end{aligned} \tag{6.81}$$

Local commutativity of the two-dimensional fields is equivalent to the invariance of the Green functions G_n under the combined action of the two exchange algebras which implies a braid invariance condition for the matrices $Q^{\Lambda_1\ldots\Lambda_n}$ [42]:

$$(B_i^{\Lambda_1\ldots\Lambda_i\Lambda_{i+1}\ldots\Lambda_n})^\dagger Q^{\Lambda_1\ldots\Lambda_i\Lambda_{i+1}\ldots\Lambda_n} B_i^{\Lambda_1\ldots\Lambda_{i+1}\Lambda_i\ldots\Lambda_n} = Q^{\Lambda_1\ldots\Lambda_{i+1}\Lambda_i\ldots\Lambda_n}. \tag{6.82}$$

The relative normalization of G_n for different n and different sets of weights are constrained by the factorization properties implied by the two-dimensional operator product expansion (6.73).

6.3.2 Regular basis of 4-point functions in the $SU(2)$ model

We shall consider in some detail only the simplest non-trivial case of 4-point functions for $G = SU(2)$. Note that there is an infinite series of such models corresponding to integer height $h = k + 2$ and Virasoro central charge $c = \frac{3k}{k+2}$. The primary fields can be labelled by their isospin I which has to satisfy the integrability condition $I \leq k/2$ [44] and have conformal dimension $\Delta(I) = \frac{I(I+1)}{(k+2)}$. The fusion rules can be computed from the Verlinde formula (6.71) and in terms of the isospins of the fields are

$$[I_1] \times [I_2] = \sum_{I=|I_1-I_2|}^{\min(I_1+I_2, k-I_1-I_2)} [I]. \tag{6.83}$$

Exploiting Möbius invariance one can reduce the KZ equation (6.77) to a system of ordinary differential equations. In order to write more compact formulae we shall make use of the polynomial realization of the irreducible $SU(2)$ modules [45] and introduce an auxiliary variable ζ to keep track of the third isospin projection m of the operators. In particular, we shall set

$$V_I(z, \zeta) = \sum_{m=-I}^{m=I} \frac{\zeta^{I+m}}{(I+m)!} V_I^m(z). \tag{6.84}$$

The $SU(2)$ generators act on $V_I(z, \zeta)$ as first-order differential operators in ζ, while the correlation functions are polynomials in ζ. We shall also assume that the isospins of the four fields satisfy the inequalities ($I_{ij} = I_i - I_j$)

$$I (= \min I_i) = I_4 \qquad |I_{12}| \leq I_{34} \qquad |I_{23}| \leq I_{14}. \tag{6.85}$$

The other cases can be treated in exactly the same way.

Möbius and $SU(2)$ invariance imply that the 4-point chiral conformal blocks have the form

$$w(z_1, \zeta_1, I_1; \ldots; z_4, \zeta_4, I_4) = g(z_{ij}, \Delta) p(\zeta_{ij}, I_{ij}) F(\eta, \xi_1, \xi_2). \tag{6.86}$$

Here $g(z_{ij}, \Delta)$ is a scale prefactor:

$$g(z_{ij}, \Delta) = \frac{z_{13}^{\Delta_2+\Delta_4} z_{24}^{\Delta_1+\Delta_3} \eta^{\Delta_s} (1-\eta)^{\Delta_u}}{z_{12}^{\Delta_1+\Delta_2} z_{23}^{\Delta_2+\Delta_3} z_{34}^{\Delta_2+\Delta_4} z_{14}^{\Delta_1+\Delta_4}} \tag{6.87}$$

η is the Möbius invariant cross ratio:

$$\eta = \frac{z_{12} z_{34}}{z_{13} z_{24}} \left(= 1 - \frac{z_{14} z_{23}}{z_{13} z_{24}} \right) \tag{6.88}$$

while Δ_s and Δ_u are the threshold dimensions in the s- $(12)(34)$ and u- $(23)(14)$ channels. For isospins constrained by (6.85), they are given by

$$\Delta_s = \Delta(I_{34}) = \frac{1}{h} I_{34}(I_{34}+1) \qquad \Delta_u = \Delta(I_{14}) = \frac{1}{h} I_{14}(I_{14}+1). \tag{6.89}$$

The polynomial

$$p(\zeta_{ij}, I_{ij}) = \zeta_{12}^{I_{14}+I_{23}} \zeta_{23}^{I_{34}-I_{12}} \zeta_{13}^{I_{12}+I_{34}}. \tag{6.90}$$

Finally, the Möbius invariant function F is a homogeneous polynomial:

$$F(\eta; \xi_1, \xi_2) = \sum_{\ell=0}^{2I} (\xi_2 \eta)^\ell [\xi_1 (1-\eta)]^{2I-\ell} f_\ell(\eta) \tag{6.91}$$

in the combinations

$$\xi_1 = \zeta_{12}\zeta_{34} \qquad \xi_2 = \zeta_{14}\zeta_{23} \qquad (\xi_1 + \xi_2 = \zeta_{13}\zeta_{24}). \tag{6.92}$$

Inserting these formulae into the KZ equation (6.77) for $n = 4$, after some algebra we obtain a system of first-order ordinary differential equations for the functions $f_\ell(\eta)$:

$$\begin{aligned} \frac{\mathrm{d} f_\ell}{\mathrm{d}\eta} &= \left\{ \frac{\ell}{\eta} [\alpha + \gamma - 1 + (\ell - 1)\delta] - \frac{2I - \ell}{1-\eta} [\beta + \gamma - 1 + (2I - \ell - 1)\delta] \right\} f_\ell \\ &\quad + \frac{\ell + 1}{1-\eta} (\alpha + \ell\delta) f_{\ell+1} - \frac{2I - \ell + 1}{\eta} [\beta + (2I - \ell)\delta] f_{\ell-1} \end{aligned} \tag{6.93}$$

where

$$\begin{gathered} h\alpha = 1 + I_{34} - I_{12} \qquad h\beta = 1 + I_{14} + I_{23} \qquad h\gamma = 1 + I_{34} + I_{12} \\ h\delta = 1 \qquad (h = k + 2). \end{gathered} \tag{6.94}$$

The system (6.93) has $2I+1$ linearly independent solutions $f_{\lambda\ell}$, $\lambda = 0, 1, \ldots, 2I$ which, for $0 < \eta < 1$, are given by the integral representations [46, 47]

$$f_{\lambda\ell}(\eta) = \int_0^\eta \mathrm{d}t_1 \int_0^{t_1} \mathrm{d}t_2 \ldots \int_0^{t_{\lambda-1}} \mathrm{d}t_\lambda \int_\eta^1 \mathrm{d}t_{\lambda+1} \times \int_{t_{\lambda+1}}^1 \mathrm{d}t_{\lambda+2} \ldots \int_{t_{2I-1}}^1 \mathrm{d}t_{2I}\, P_{\lambda\ell}(t_i;\eta;\alpha,\beta,\gamma,\delta) \tag{6.95}$$

where

$$P_{\lambda\ell} = \prod_{i=1}^{2I} t_i^\alpha (1-t_i)^\beta \prod_{i=1}^{\lambda} (\eta - t_i)^{\gamma-1} \prod_{j=\lambda+1}^{2I} (t_j - \eta)^{\gamma-1} \prod_{i<j} (\varepsilon_{\lambda j} t_{ij})^{2\delta} \times \sum_\sigma \frac{1}{\ell!(2I-\ell)!} \prod_{s=1}^{\ell} t_{i_s}^{-1} \prod_{r=\ell+1}^{2I} (1-t_{i_r})^{-1} \tag{6.96}$$

$$\varepsilon_{\lambda j} = \begin{cases} 1 & \text{for } \lambda \geq j \\ -1 & \text{for } \lambda < j \end{cases} \qquad t_{ij} = t_i - t_j.$$

The sum in (6.96) extends over all $(2I)!$ permutations $\sigma : (1, \ldots, 2I) \to (i_1, \ldots, i_{2I})$. Note that the integration contours in (6.95) never go to infinity. This is an important difference with respect to the commonly used integral representations [39, 40, 45] which correspond to tree expansions. Our choice has the advantage that the solutions are linearly independent and non-singular (if all four external isospins satisfy the integrability condition $I_i \leq k/2$). In particular, the exchange operators are also well defined.

6.3.3 Matrix representation of the exchange algebra

Each basis of solutions $\{w^\lambda, \lambda = 0, \ldots, 2I\}$ of the (4-point) KZ equation gives rise to a matrix representation of the algebra of exchange operators B_1, B_2 and B_3 [42, 48]. We shall work out only the action of B_1 and B_2 on the 4-point blocks (6.86) since B_3 is proportional to B_1 (see equation (6.109)). According to (6.78) B_i act on the cross ratio η (6.88) as follows

$$B_1 : \eta \to \frac{\eta \mathrm{e}^{\mathrm{i}\pi}}{1-\eta} \qquad \left(= \lim_{t\to 1} \frac{\eta \mathrm{e}^{\mathrm{i}\pi t}}{1 + \mathrm{i}\eta \mathrm{e}^{\mathrm{i}\frac{\pi}{2}t} \sin \frac{\pi}{2}t} \right) \tag{6.97}$$

$$B_2 : \eta \to \frac{1}{\eta} \qquad \left(= \lim_{t\to 1} \frac{\eta \cos\frac{\pi}{2}t - \mathrm{i}\sin\frac{\pi}{2}t}{\cos\frac{\pi}{2}t - \mathrm{i}\eta \sin\frac{\pi}{2}t} \right). \tag{6.98}$$

The expressions within parentheses indicate the analytic continuation path in the η-plane, hence B_1 carries η around 0 from above, while B_2 carries η around 1 from below. Note that in order to specify the domain and the target space of the exchange operators B_i, one actually has to indicate all four isospins. We shall use the notation

$$B_1^{I_1 I_2 I_3 I_4} : \mathcal{L}(I_1 I_2 I_3 I_4) \to \mathcal{L}(I_2 I_1 I_3 I_4) \tag{6.99}$$

$$B_2^{I_1 I_2 I_3 I_4} : \mathcal{L}(I_1 I_2 I_3 I_4) \to \mathcal{L}(I_1 I_3 I_2 I_4). \tag{6.100}$$

The action of the exchange operators on the basis constructed in the previous section, $B_i : w^\lambda \to (B_i)^\lambda{}_\mu w^\mu$, can be obtained by analytic continuation of the integral representations (6.95). Note that B_i not only transforms the integrand (6.96) but also reorders the integration contours in (6.95). The explicit expressions can be written in a more compact form, if one introduces q-deformed numbers

$$[\lambda] = \frac{q^\lambda - q^{-\lambda}}{q - q^{-1}} \tag{6.101}$$

where

$$q = \mathrm{e}^{\mathrm{i}\pi\delta} = \mathrm{e}^{\mathrm{i}\frac{\pi}{h}} (\Rightarrow q^h = -1) \qquad \bar{q} = q^{-1} \tag{6.102}$$

and q-deformed binomial coefficients

$$\begin{bmatrix} \mu \\ \lambda \end{bmatrix} = \frac{[\mu]!}{[\lambda]![\mu - \lambda]!} \qquad [\lambda]! = [\lambda][\lambda - 1]! \qquad [0]! = 1. \tag{6.103}$$

The exchange matrix B_1 is upper triangular in our basis

$$\begin{aligned}(B_1^{I_1 I_2 I_3 I_4})^\lambda{}_\mu &= (-1)^{I_1 + I_2 - I_{34} - \mu} \\ &\quad \times q^{(I_{34}+\mu)(I_{34}+\lambda+1)+I_{12}(\mu-\lambda)-I_1(I_1+1)-I_2(I_2+1)} \begin{bmatrix} \mu \\ \lambda \end{bmatrix}\end{aligned} \tag{6.104}$$

while the exchange matrix B_2 is lower triangular and is related to B_1 by a similarity transformation:

$$B_2^{I_1 I_2 I_3 I_4} = F^{I_2 I_3 I_1 I_4} B_1^{I_3 I_2 I_1 I_4} F^{I_1 I_2 I_3 I_4}. \tag{6.105}$$

The matrix $F^{I_1 I_2 I_3 I_4} : \mathcal{L}(I_1 I_2 I_3 I_4) \to \mathcal{L}(I_3 I_2 I_1 I_4)$, called a fusion matrix [28], is involutive:

$$F^{I_3 I_2 I_1 I_4} F^{I_1 I_2 I_3 I_4} = 1 \tag{6.106}$$

and in the basis (6.95) is represented by an antidiagonal matrix whose elements are independent of the order of the isospins

$$(F^{I_1 \ldots I_4})^\lambda{}_\mu = \delta_\mu^{2I-\lambda}. \tag{6.107}$$

Using the expressions (6.104) and (6.105) one can verify that the exchange operators B_i satisfy the parameter free Yang–Baxter equation [49]

$$\begin{aligned}B_1^{I_2 I_3 I_1 I_4} B_2^{I_2 I_1 I_3 I_4} B_1^{I_1 I_2 I_3 I_4} &= B_2^{I_3 I_1 I_2 I_4} B_1^{I_1 I_3 I_2 I_4} B_2^{I_1 I_2 I_3 I_4} \\ &= (-1)^{I_1 + I_2 + I_{34}} q^{I_4(I_4+1) - I_1(I_1+1) - I_2(I_2+1) - I_3(I_3+1)} F.\end{aligned} \tag{6.108}$$

Let us note also that, for the 4-point functions, $B_3^{I_1I_2I_3I_4}$ and $B_1^{I_1I_2I_3I_4}$ are proportional

$$B_3^{I_1I_2I_3I_4} = (-1)^{I_3+I_4-I_1-I_2} q^{\{I_1(I_1+1)+I_2(I_2+1)-I_3(I_3+1)-I_4(I_4+1)\}} B_1^{I_1I_2I_3I_4}. \tag{6.109}$$

The exchange operators for the 3-point functions are just phases, since the space of $SU(2)$ invariants is one-dimensional in this case. They can be obtained as a special case (for $I_4 = 0$) from the general expressions (6.104), (6.105):

$$B_1^{I_1I_2I_3} = (-1)^{I_1+I_2-I_3} q^{I_3(I_3+1)-I_1(I_1+1)-I_2(I_2+1)} \tag{6.110}$$

$$B_2^{I_1I_2I_3} = (-1)^{I_2+I_3-I_1} q^{I_1(I_1+1)-I_2(I_2+1)-I_3(I_3+1)}. \tag{6.111}$$

The exchange operator for the 2-point function which exists only for $I_2 = I_1$ is given by

$$B^{I_1I_2} = (-1)^{2I_1} q^{-2I_1(I_1+1)}. \tag{6.112}$$

6.3.4 Two-dimensional braid invariant Green functions

So far we have computed only the exchange operators in the chiral sector of the theory. To compute the two-dimensional Green functions (6.81) we also need the expressions for the antichiral sector. To derive them let us recall that the corresponding current algebras are isomorphic, while the orientation of the analytic continuation contours (6.78) are opposite in the two sectors. Thus the exchange operators in the antichiral sector are complex conjugates of the corresponding chiral ones and can be obtained from them by the substitution $q \to q^{-1}$ $(= \bar{q})$ (see equation (6.102)).

The locality condition for the two-dimensional Green functions (6.81) implies the braid invariance constraints (6.82) for the matrices Q. For a generic value of q on the unit circle (or, equivalently, for a generic value of the level k) the solution of (6.82) is unique and corresponds to a diagonal pairing of the two sectors (a diagonal modular invariant). For special values of the level k there also exist other solutions which correspond to non-diagonal modular invariants. Let us first consider the generic diagonal case. The solution of the braid invariance condition (6.82) is [47]

$$Q_{\mu\nu}(I_1, I_2, I_3, I_4) = (-1)^{\mu+\nu} \frac{[\mu]![\nu]![\mu - I_{12} + I_{34}]![\nu - I_{12} + I_{34}]!}{[2I_1]![2I_2]![2I_3]![2I_4]!} \times \sum_{\rho=0}^{\min(\mu,\nu,k(I))} T_\rho(\mu, \nu; I_i) \tag{6.113}$$

where $\mu, \nu = 0, \ldots, 2I_4$,

$$k(I) = k - I_1 - I_2 - I_3 + I_4$$

and

$$T_\rho(\mu, \nu; I_i) = [2I_{34} + 2\rho + 1] \times \frac{[I_1 + I_2 + I_{34} + \rho + 1]![I_1 + I_2 - I_{34} - \rho]![2I_4 - \rho]!}{[2I_{34} + \mu + \rho + 1]![2I_{34} + \nu + \rho + 1]!} \times \frac{[2I_{34} + \rho]![I_{12} + I_{34} + \rho]![2I_3 + \rho + 1]!}{[\mu - \rho]![\nu - \rho]![\rho]![I_{34} - I_{12} + \rho]!}.$$

It is straightforward but rather lengthy to check that (6.113) satisfies (6.82).

In order to study the factorization properties of the two-dimensional Green functions let us rewrite the 4-point chiral conformal blocks in the tree bases. In the s-channel, which exhibits the singularities of the solutions for small z_{12} (hence small η), we find

$$\begin{aligned} S^{(I_1,I_2,I_3,I_4)}_{I_{34}+\lambda}(z, \zeta) &= \sum_{\nu=0}^{2I_4} w^{(I_1,I_2,I_3,I_4)}_\nu(z, \zeta)\sigma^{-1}_{\nu\lambda}(I_1, I_2, I_3, I_4) \\ &= \sum_{\nu=\lambda}^{2I_4} \frac{(-1)^{\nu-\lambda}[\nu]![\nu - I_{12} + I_{34}]![2I_{34} + 2\lambda + 1]!}{[\nu - \lambda]![\lambda]![\lambda - I_{12} + I_{34}]![2I_{34} + \nu + \lambda + 1]!} \\ &\quad \times w^{(I_1,I_2,I_3,I_4)}_\nu(z, \zeta). \end{aligned} \tag{6.114}$$

Let us stress that for q a root of unity (note that $q^{k+2} = -1$) the matrix elements of the matrix σ^{-1} are well defined only if

$$I_1 + I_2 + I_{34} + \lambda \leq k. \tag{6.115}$$

In other words, the s-channel conformal blocks (6.114) are well defined only for intermediate fields that respect the fusion rules (6.83). In the rest of the chapter, we shall use (6.114) and all other tree bases formulae only for such intermediate fields. Having this in mind, we can also introduce the matrix formally inverse to σ^{-1}:

$$\sigma_{\lambda\mu}(I_1, I_2, I_3, I_4) = \frac{[2I_{34} + \lambda + \mu]![I_{34} - I_{12} + \lambda]!}{[2I_{34} + 2\lambda]![I_{34} - I_{12} + \mu]!} \begin{bmatrix} \lambda \\ \mu \end{bmatrix}. \tag{6.116}$$

In the s-channel basis (6.114), the exchange operator B_1 has a simple diagonal form:

$$((B^s_1)^{I_1I_2I_3I_4})^\lambda{}_\mu = \delta^\lambda_\mu(-1)^{I_1+I_2-I_{34}-\mu}q^{(I_{34}+\mu)(I_{34}+\mu+1)-I_1(I_1+1)-I_2(I_2+1)} \tag{6.117}$$

while the exchange operator B_2 and the fusion matrix F are given by complicated expressions. In particular, for F one finds

$$((F^s)^{I_1I_2I_3I_4})_{\mu\nu} = \sum_{\lambda=0}^{2I} \sigma_{\nu\lambda}(I_1, I_2, I_3, I_4)\sigma^{-1}_{2I-\lambda,\mu}(I_3, I_2, I_1, I_4) \tag{6.118}$$

where $I = I_4$ (see equation (6.85)). Inserting the expressions for σ and σ^{-1} we obtain

$$((F^s)^{I_1 I_2 I_3 I_4})_{\mu\nu}$$
$$= \sum_{\lambda=0}^{\min(\nu,2I-\mu)} (-1)^{2I-\lambda-\mu} \frac{[2I_{34}+\nu+\lambda]![I_{34}-I_{12}+\nu]![\nu]!}{[2I_{34}+2\nu]![I_{34}-I_{12}+\lambda]![\nu-\lambda]![\lambda]!}$$
$$\times \frac{[2I-\lambda]![2I-\lambda-I_{32}+I_{14}]![2I_{14}+2\mu+1]!}{[2I-\lambda-\mu]![\mu]![\mu-I_{32}+I_{14}]![2I_{14}+2I-\lambda+\mu+1]!}. \tag{6.119}$$

The other tree basis, the u-channel, exhibits the singularities of the solutions for small z_{23} (hence small $1-\eta$). To construct it let us note that the KZ equation as a differential equation in $1-\eta$ for the conformal blocks with isospin order $I_3 I_2 I_1 I_4$ coincides with the KZ equation in η for the conformal blocks with isospin order $I_1 I_2 I_3 I_4$. Thus we can define the u-channel blocks which diagonalize the exchange operator B_2 as

$$U_{I_{14}+\lambda}^{(I_1,I_2,I_3,I_4)}(\eta) = S_{I_{14}+\lambda}^{(I_2,I_3,I_4,I_1)}(1-\eta) = (-1)^{I_2+I_3-I_1-I_4}$$
$$\times q^{I_1(I_1+1)+I_4(I_4+1)-I_2(I_2+1)-I_3(I_3+1)} S_{I_{14}+\lambda}^{(I_3,I_2,I_1,I_4)}(1-\eta) \tag{6.120}$$

where the second equation follows from the diagonal form of the exchange operator B_1 (6.117) (and hence also of B_3 and $B_1 B_3^{-1}$) in the s-channel basis. The u-channel blocks (6.120) are related to the s-channel blocks (6.114) by the fusion matrix F^s (6.119):

$$U_{I_{14}+\mu}^{(I_1,I_2,I_3,I_4)}(\eta) = \sum_{\nu} ((F^s)^{I_1 I_2 I_3 I_4})_{\mu\nu} S_{I_{34}+\nu}^{(I_1,I_2,I_3,I_4)}(\eta). \tag{6.121}$$

The two-dimensional Green functions can be expressed in terms of the tree conformal blocks as

$$G_4^{(I_i)}(z,\bar{z};\zeta,\bar{\zeta}) = \sum_{\nu=0}^{\min(2I,k(I))} [g_s]_\nu^{(I_i)} \overline{S}_{I_{34}+\nu}(\bar{z},\bar{\zeta}) S_{I_{34}+\nu}(z,\zeta)$$
$$= \sum_{\mu=0}^{\min(2I,k(I))} [g_u]_\mu^{(I_i)} \overline{U}_{I_{14}+\mu}(\bar{z},\bar{\zeta}) U_{I_{14}+\mu}(z,\zeta). \tag{6.122}$$

The normalization constants $[g_s]$ and $[g_u]$ can be written as

$$[g_s]_\nu^{(I_1 I_2 I_3 I_4)} = \frac{C_{I_1 I_2 I_{34}+\nu} C_{I_3 I_4 I_{34}+\nu}}{N_{I_{34}+\nu}} \tag{6.123}$$

$$[g_u]_\mu^{(I_1 I_2 I_3 I_4)} = \frac{C_{I_2 I_3 I_{14}+\mu} C_{I_1 I_4 I_{14}+\mu}}{N_{I_{14}+\mu}} \tag{6.124}$$

where $C_{I_1 I_2 I_3}$ are:

$$C_{I_1 I_2 I_3} = [I_1 + I_2 + I_3 + 1]! \times \frac{[I_1 + I_2 - I_3]![I_2 + I_3 - I_1]![I_1 + I_3 - I_2]!}{[2I_1]![2I_2]![2I_3]!} \tag{6.125}$$

while

$$N_I = C_{II0} = [2I + 1]. \tag{6.126}$$

Now we can impose the factorization property in both the s- and u-channels. Comparison of (6.123) and (6.124) with the two-dimensional OPE (6.73) shows that the two-dimensional structure constants in the diagonal model (in which the two-dimensional fields have equal chiral and antichiral labels) are given by

$$C^{(KK)}_{(II)(JJ)} = \frac{C_{IJK}}{N_K}. \tag{6.127}$$

Moreover, if we choose the normalizations of the 2-point functions to be equal to N_I (6.126), the normalizations of the 3-point functions are equal to $C_{I_1 I_2 I_3}$ (6.125).

This construction can also be extended to the non-diagonal $SU(2)$ current algebra models. For the D_{odd} series of models, which exist for values of the level $k = 4p - 2$, the structure constants are

$$C^{(K\bar{K})}_{(I\bar{I})(J\bar{J})} = \epsilon_{(I\bar{I})(J\bar{J})(K\bar{K})} \sqrt{\frac{C_{IJK} C_{\bar{I}\bar{J}\bar{K}}}{N_K N_{\bar{K}}}} \tag{6.128}$$

where the signs ϵ are symmetric in all three pairs of indices and differ from $+1$ only if two pairs of the isospins, say I, $\bar{I}$, J, $\bar{J}$, are half-integers, in which case they are equal to $(-1)^K$ $(= (-1)^{\bar{K}})$.

For the other $SU(2)$ current algebra models denoted by D_{even} and E_6, E_7, E_8 one can also compute the structure constants [50]. The resulting expressions are not as simply related to the diagonal ones. This can be explained by the fact that these models correspond to extensions of the observable algebra, so their structure is determined by this extension, rather than by the underlying $SU(2)$ current algebra.

6.4 CFT on surfaces with holes and crosscaps

Conformal field theories in presence of boundaries have been introduced by Cardy to describe critical phenomena in statistical mechanics and solid state physics [1,2,29]. An alternative approach, called open and unoriented descendant construction, was proposed by Sagnotti in the framework of string theory to unify open strings with closed oriented and unoriented strings in a consistent way [4]. In this section we shall review some general properties of boundary CFT. The

$SU(2)$ conformal current algebra models will again be used as an example. On one hand, they are relatively simple and all the necessary data (chiral conformal blocks, structure constants, exchange operators) are explicitly known. On the other hand, the $SU(2)$ models exhibit many features of the general case (like infinite series of non-diagonal models and non-Abelian fusion rules).

6.4.1 Open sector, sewing constraints

The presence of a boundary breaks the two-dimensional conformal symmetry, since the boundary cannot be invariant under all the transformations of $Vir \otimes \overline{Vir}$. If the central charges of the two chiral algebras are equal ($\bar{c} = c$), it is possible to introduce boundaries which are preserved at most by the diagonal subalgebra Vir_{diag}. We shall call such boundaries conformal boundaries. In the rest we shall assume that all boundaries are conformal. The introduction of non-conformal boundaries is also possible, but one can no longer use the tools of conformal field theory for their study.

Assume that the conformal boundary coincides with the line $x^1 = 0$. The conformal invariance condition means that there is no energy transfer across the boundary, hence the stress–energy tensor satisfies [1]

$$\Theta(x_-) = \bar{\Theta}(x_+) \qquad \text{for } x_- = x_+ \Leftrightarrow x^1 = 0 \tag{6.129}$$

since $x_\pm = x^0 \pm x^1$. So one can define the stress–energy tensor in the theory with conformal boundaries as

$$\Theta_d(x) = \begin{cases} \Theta(x_-) & \text{for } x^1 \geq 0 \\ \bar{\Theta}(x_+) & \text{for } x^1 < 0. \end{cases} \tag{6.130}$$

In a similar way, if the two-dimensional theory is invariant under the product of two isomorphic conformal current algebras $\mathcal{A} \otimes \bar{\mathcal{A}}$ with equal levels $\bar{k} = k$, the boundary can be preserved at most by the diagonal subalgebra $\mathcal{A}_{\text{diag}}$. Such boundaries are called symmetry-preserving, the currents in this case being defined as

$$j^a_d(x) = \begin{cases} j^a(x_-) & \text{for } x^1 \geq 0 \\ \bar{j}^a(x_+) & \text{for } x^1 < 0. \end{cases} \tag{6.131}$$

One can also introduce conformal boundaries that are preserved only by a proper subalgebra $\mathcal{A}' \subset \mathcal{A}_{\text{diag}}$ (such that the boundary is still invariant under Vir_{diag}). Such boundaries are called symmetry-breaking (or symmetry non-preserving) boundaries and have also been studied [51]. In these lectures we shall restrict our attention only to the simpler case of symmetry-preserving boundaries.

We can pass to the analytic picture by mapping the boundary onto the unit circle by a Cayley transform (6.2). The stress–energy tensor becomes

$$T_d(z) = \begin{cases} T(z) & \text{for } |z| \leq 1 \\ \dfrac{1}{z^4}\bar{T}\left(\dfrac{1}{z}\right) & \text{for } |z| > 1 \end{cases} \tag{6.132}$$

(where we used $\bar{z} \leftrightarrow 1/z$ in this picture), while the currents are

$$J_d^a(z) = \begin{cases} J^a(z) & \text{for } |z| \leq 1 \\ -\dfrac{1}{z^2}\bar{J}^a\left(\dfrac{1}{z}\right) & \text{for } |z| > 1. \end{cases} \tag{6.133}$$

The sign change with respect to (6.132) comes from the prefactor in the Cayley transform.

Let us also introduce the following combinations of the Laurent modes of the stress–energy tensor T and the currents J^a

$$\mathcal{L}_n = L_n - \bar{L}_{-n} \tag{6.134}$$

and

$$\mathcal{J}^a{}_n = J_n^a + \bar{J}_{-n}^a. \tag{6.135}$$

Since the left and right central charges and levels are equal ($\bar{c} = c$, $\bar{k} = k$) the modes (6.134) satisfy the commutation relations of the Virasoro algebra with the central charge being equal to zero:

$$[\mathcal{L}_n, \mathcal{L}_m] = (n-m)\mathcal{L}_{n+m} \tag{6.136}$$

while the modes (6.135) satisfy the commutation relations of the current algebra with a level equal to zero:

$$[\mathcal{J}^a{}_n, \mathcal{J}^b{}_m] = \mathrm{i} f^{abc}\mathcal{J}^c{}_{n+m}. \tag{6.137}$$

These two algebras have no non-trivial representations, hence the modes (6.134), (6.135) annihilate all the boundary states $|B\rangle$ in the theory

$$\mathcal{L}_n|B\rangle = (L_n - \bar{L}_{-n})|B\rangle = 0 \tag{6.138}$$

and

$$\mathcal{J}^a{}_n|B\rangle = (J_n^a + \bar{J}_{-n}^a)|B\rangle = 0. \tag{6.139}$$

For rational models a basis of states that satisfy (6.138) (called the Ishibashi states) has been constructed in [52] as infinite sums of products of left and right states:

$$|\mathcal{I}_\Lambda\rangle = \sum_m |\Lambda, m\rangle \otimes \overline{|\Lambda, m\rangle} \tag{6.140}$$

where the sum extends over all the quasiprimary descendants of the primary state $|\Lambda\rangle$. Note that the Ishibashi states are not eigenvalues of the energy $L_0 + \bar{L}_0$ and are not normalizable in the usual sense.

One important consequence of equations (6.132), (6.133) is that in the presence of boundaries the n-point functions of two-dimensional primary fields $\phi_{\Lambda\bar{\Lambda}}(z, \bar{z})$ and the chiral conformal blocks of $2n$-chiral vertex operators with the same weights satisfy the same equations as functions of the $2n$ variables

$(z_1, \bar{z}_1, \ldots, z_n, \bar{z}_n)$ [1]. Indeed, since the chiral and antichiral parts of the stress–energy tensor and of the currents act independently on the chiral and antichiral vertex operators in the decomposition of the two-dimensional primary fields (see also equation (6.60)),

$$\phi_{\Lambda\bar{\Lambda}}(z,\bar{z}) = \sum_{\substack{\Lambda_i \ \bar{\Lambda}_i \\ \Lambda_f \ \bar{\Lambda}_f}} V_{\Lambda\ \Lambda_i}^{\ \Lambda_f}(z)\, \bar{V}_{\bar{\Lambda}\ \bar{\Lambda}_i}^{\ \bar{\Lambda}_f}(\bar{z}) n_{i\bar{i}}^{f\bar{f}} \tag{6.141}$$

the n-point functions of the two-dimensional fields in the theory with boundaries are linear combinations of $2n$-point chiral conformal blocks. Note the difference with respect to the case without boundaries reviewed in the previous section, where the two-dimensional functions are sesquilinear combinations of n-point conformal blocks.

Another important property of the boundary is the existence of one-dimensional fields ψ called boundary fields [1]. They are defined only on the boundary (on the unit circle in the analytic picture). Equations (6.132), (6.133) imply that the Virasoro algebra and the conformal current algebra which act on the boundary have the same central charge and the same level as the chiral ones. Hence, the primary boundary fields can be labelled by the same set of weights Λ. There can be different boundary conditions on different portions of the boundary, which we denote by labels a, b, c. The boundary fields carry two boundary condition labels $\psi_{\Lambda}^{ab}(x)$ and change the boundary condition from b to a. In general, a degeneracy label accounting for the multiplicity of the boundary fields may also be necessary. For simplicity we shall omit the degeneracy labels. For a more accurate analysis of this point see, e.g., [53]. We shall denote the argument of the boundary fields by x which takes values only on the unit circle, to distinguish it from z $(\bar{z})$ which take values inside (outside) the unit circle.

In general, the boundary fields do not locally commute, rather they behave much like the chiral vertex operators under the exchange algebra. In other words, in correlation functions the ordering of their arguments on the circle cannot be changed arbitrarily. In particular, this implies that the 4-point functions of boundary fields satisfy only one crossing symmetry relation, called planar duality, in contrast with the two-dimensional case where the crossing symmetry of the 4-point functions implies two duality relations.

To simplify the notation in this section we shall consider only the $SU(2)$ current algebra models and label the fields by $i = 2I_i + 1$ and $\bar{i} = 2\bar{I}_i + 1$ rather than by their weights Λ_i and $\bar{\Lambda}_i$. Hence, the identity operators carry the label 1. When this is not ambiguous, we shall also omit the spacetime (z and x) and $SU(2)$ (ζ) variables.

The operator product expansion for the boundary fields schematically has the form (note the continuity of the boundary indices)

$$\psi_i{}^{ab}\psi_j{}^{bc} \sim \sum_l C_{ijl}^{abc}\, \psi_l^{ac} \tag{6.142}$$

where the sum is over all the values allowed by the fusion rules (6.83). The boundary structure constants C_{ijl}^{abc} are, in general, not symmetric. Other important data are the normalizations of the 2-point functions of the boundary fields, since they cannot be chosen arbitrarily [2]. To define them one also has to specify the order of the arguments, since the boundary fields do not commute. Both variables are on the unit circle so we can order them by their phase

$$\langle \psi_i{}^{ab}(x_1;\zeta_1)\psi_i{}^{ba}(x_2;\zeta_2)\rangle = \frac{\alpha_i{}^{ab}(\zeta_{12})^{2I_i}}{(x_{12})^{2\Delta_i}} \qquad \text{for } \mathrm{Arg}(x_2) < \mathrm{Arg}(x_1) \quad (6.143)$$

where I_i is the isospin of ψ_i. The normalizations of the fields with exchanged boundary labels are related. For example, for the $SU(2)$ current algebra models one finds

$$\alpha_i{}^{ab} = \alpha_i{}^{ba}(-1)^{2I_i}. \qquad (6.144)$$

Let us stress that even if we consider in detail only the $SU(2)$ conformal current algebra case, most of the formulae are also valid in more general cases (with minor modifications in the numerical factors). For instance, in the unitary minimal models case one just has to omit all the isospin dependence.

Using the boundary OPE (6.142) we can compute the three-point functions of the boundary fields $\langle \psi_i{}^{ab}\psi_j{}^{bc}\psi_l{}^{ca}\rangle$ and $\langle \psi_j{}^{bc}\psi_l{}^{ca}\psi_i{}^{ab}\rangle$ in two different ways. This gives the following consistency conditions:

$$C_{ijl}^{abc}\alpha_l{}^{ac} = C_{jli}^{bca}\alpha_i{}^{ab} \qquad \text{and} \qquad C_{jli}^{bca}\alpha_i{}^{ba} = C_{lij}^{cab}\alpha_j{}^{bc} \qquad (6.145)$$

that together with (6.144) imply also

$$C_{ijl}^{abc}\alpha_l{}^{ac} = (-1)^{2I_i}C_{lij}^{cab}\alpha_j{}^{bc}. \qquad (6.146)$$

The natural normalization of the boundary identity operator is

$$C_{i\mathbf{1}i}^{abb} = 1 \qquad \langle \mathbf{1}^{aa}\rangle = \alpha_{\mathbf{1}}^{aa} \qquad (6.147)$$

while all other 1-point functions of the boundary fields vanish.

The planar duality constraint for the 4-point functions $\langle \psi_i^{ab}\psi_j^{bc}\psi_k^{cd}\psi_l^{da}\rangle$ reads:

$$\sum_p C_{ijp}^{abc}C_{klp}^{cda}\alpha_p^{ac}S_p(i,j,k,l) = \sum_q C_{jkq}^{bcd}C_{liq}^{dab}\alpha_q^{bd}U_q(i,j,k,l) \qquad (6.148)$$

and after expressing the u-channel blocks (6.120) in terms of the s-channel blocks (6.114) by the fusion matrix F^s (6.119) as (see also (6.121)):

$$U_q(i,j,k,l) = \sum_p F_{qp}(i,j,k,l)S_p(i,j,k,l) \qquad (6.149)$$

we obtain a quadratic relation for the boundary structure constants C_{ijk}^{abc} and the 2-point normalizations α_i^{ab}:

$$C_{ijp}^{abc} C_{klp}^{cda} \alpha_p^{ac} = \sum_q C_{jkq}^{bcd} C_{liq}^{dab} \alpha_q^{bd} F_{qp}(i, j, k, l). \tag{6.150}$$

These relations do not determine the boundary structure constants completely. In other words, the boundary theory cannot be considered independently but only as a part of the two-dimensional conformal theory.

The relation between the bulk and boundary fields is encoded into the bulk-to-boundary expansion

$$\phi_{i,\bar{i}}|_a \sim \sum_j C_{(i,\bar{i})j}^{a} \psi_j{}^{aa} \tag{6.151}$$

that expresses the two-dimensional fields in front of a portion of boundary with given boundary condition a in terms of the corresponding boundary fields. The sum is again over all the values allowed by the fusion rules. The proper normalization of the identity operator gives

$$C_{(\mathbf{1},\mathbf{1})\mathbf{1}}^{a} = 1 \tag{6.152}$$

for all boundary conditions a.

The consistency of the operator product expansions (6.73), (6.142) and (6.151) have been studied by Lewellen [3], who has shown that the complete set of relations (also called sewing constraints) which guarantee the consistency of the theory includes two more equations, the first one involving 4-point functions and the second one involving 5-point functions. The first relation arises from the correlation functions of one two-dimensional bulk field and two boundary fields. As already stressed, the boundary fields have a fixed order of the arguments but the two-dimensional fields also have to be local in the presence of boundary fields, which implies

$$\langle \phi_{(i,\bar{i})} \psi_j^{ba} \psi_k^{ab} \rangle = \langle \psi_j^{ba} \phi_{(i,\bar{i})} \psi_k^{ab} \rangle. \tag{6.153}$$

Note that the bulk field is expanded in front of portions of the boundary with different boundary conditions in the left- and in the right-hand sides of this equation. Using also (6.142), (6.151) one obtains

$$\sum_l C_{(i,\bar{i})l}^{b} C_{ljk}^{bba} \alpha_k^{ba} S_l(i, \bar{i}, j, k) = \sum_n C_{(i,\bar{i})n}^{a} C_{jnk}^{baa} \alpha_k^{ba} U_n(j, i, \bar{i}, k). \tag{6.154}$$

To derive the constraint on the structure constants we have to relate the U- and the S-blocks. A convenient way to do this is to use repeatedly the fusion matrix (and its inverse) in such a way that the exchange operators always act diagonally (see (6.117)). In other words before applying B_1 or B_3 we change to the s-channel basis, while before applying B_2 we change to the u-channel basis. The resulting composite exchange operator is

$$U_n(j,i,\bar{i},k) = \sum_{m,r,s,p,l} F_{nm}(j,i,\bar{i},k)(B_1)_{mr}(i,j,\bar{i},k)F^{-1}_{rs}(i,j,\bar{i},k) \times (B_2)^{-1}_{sp}(i,\bar{i},j,k)F_{pl}(i,\bar{i},j,k)S_l(i,\bar{i},j,k). \tag{6.155}$$

Inserting this in (6.154) and using the explicit expressions for the exchange operators B_1 and B_2 in the $SU(2)$ model, we obtain the constraint

$$C^b_{(i,\bar{i})l} C^{bab}_{jkl} \alpha_l{}^{bb} = \sum_{m,n,p} (-1)^{(I_i - I_{\bar{i}} + 2I_j + I_p - I_m)} \mathrm{e}^{-\mathrm{i}\pi(\Delta_i - \Delta_{\bar{i}} - \Delta_m + \Delta_p)} \times C^a_{(i,\bar{i})n} C^{aba}_{kjn} \alpha_n{}^{aa} F_{nm}(j,i,\bar{i},k) \times F^{-1}{}_{mp}(i,j,\bar{i},k) F_{pl}(i,\bar{i},j,k). \tag{6.156}$$

The other independent relation can be derived from the 5-point functions of two bulk fields and one boundary field of the form

$$\langle \phi_{(i,\bar{i})} \phi_{(j,\bar{j})} \psi_k{}^{aa} \rangle. \tag{6.157}$$

This function can again be computed in two different ways. We can first use the two-dimensional OPE (6.73) followed by a bulk-to-boundary OPE (6.151), alternatively we can use the bulk-to-boundary OPE (6.151) twice followed by a boundary OPE (6.142). Before proceeding we have to define a basis in the space of 5-point functions. We shall use a tree representation which decomposes the 5-point functions into products of a 4-point function and a 3-point function (denoted by $g(1,2,3)$) with one common external leg:

$$\begin{aligned} X_{pq}(1,2,3,4,5) &= S_p(1,2,q,5) g(q,3,4) \\ &= g(1,2,p) U_q(p,3,4,5). \end{aligned} \tag{6.158}$$

In this notation the equivalence of the two ways of computing the function (6.157) implies

$$\sum_{p,q} C^{(p,\bar{q})}_{(i,\bar{i})(j,\bar{j})} C^a_{(p,\bar{q})k} \alpha^{aa}_k X_{p\bar{q}}(j,i,\bar{i},\bar{j},k) = \sum_{p,q} C^a_{(i,\bar{i})p} C^a_{(j,\bar{j})q} C^{aaa}_{pqk} \alpha^{aa}_k X_{pq}(i,\bar{i},j,\bar{j},k). \tag{6.159}$$

The two expressions can again be related by the exchange operators, for example by $F_{[2]} B_1^{-1} B_2 B_1 B_2 F^{-1}_{[2]}$, where the label in brackets indicates on which 4-point subtree acts the fusion matrix F. We can use the Yang–Baxter equations (6.108) to express $B_2 B_1 B_2$ in terms of $F_{[1]}$ obtaining (if $\alpha^{aa}_k \neq 0$)

$$C^{(p,\bar{q})}_{(i,\bar{i})(j,\bar{j})} C^a_{(p,\bar{q})k} = \sum_{r,s,t} (-1)^{(I_j - I_t + I_r)} \mathrm{e}^{-\mathrm{i}\pi(\Delta_j - \Delta_t + \Delta_r)} C^{aaa}_{rsk} \times C^a_{(i,\bar{i})r} C^a_{(j,\bar{j})s} F_{st}(r,j,\bar{j},k) F_{rp}(j,i,\bar{i},t) F^{-1}_{t\bar{q}}(p,\bar{i},\bar{j},k). \tag{6.160}$$

Both equations (6.156) and (6.160) can be written in several different equivalent forms, since the exchange operators satisfy duality relations (like the Yang–Baxter equation (6.108)) [28]. Our derivation follows [9]. For alternative ones see also [3, 53, 54]. In particular, [54] contains the general solution of the sewing constraints for the unitary minimal models with a detailed analysis of the residual normalization freedom. Here we shall address only a simpler problem, namely we shall try to count the allowed boundary conditions. Note that in all the sewing constraints the boundary fields enter as external insertions, so one can always start with only one type of boundary labels, say a, and solve only the corresponding subsystem. There is, however, a systematic way to determine the whole set of allowed boundary conditions. In other words, by only analysing the sewing constraints one can find all boundary states $|a\rangle$. To illustrate this point, let us consider one particular case of the function (6.157), namely $\langle \phi_{(i,\bar{i})} \phi_{(j,\bar{j})} \mathbf{1}^{aa} \rangle$. Then the condition (6.160) becomes

$$C^{(q,\bar{q})}_{(i,\bar{i})(j,\bar{j})} C^{a}_{(q,\bar{q})\mathbf{1}} \alpha^{aa}_{\mathbf{1}} = \sum_{p} (-1)^{(I_j - I_{\bar{j}} + I_p)} \mathrm{e}^{-\mathrm{i}\pi(\Delta_j - \Delta_{\bar{j}} + \Delta_p)} \times \alpha^{aa}_{p} C^{a}_{(i,\bar{i})p} C^{a}_{(j,\bar{j})p} F_{pq}(j, i, \bar{i}, \bar{j}). \qquad (6.161)$$

Multiplying by $F^{-1}_{qr}(j, i, \bar{i}, \bar{j})$, summing on q and keeping only the equation corresponding to $r = 1$ we find a system of equations for the bulk-to-boundary coefficients in front of the boundary identity operators

$$B^{a}_{i} = C^{a}_{(i,\bar{i})\mathbf{1}} \qquad (6.162)$$

where to simplify notation we have used the fact that for a permutation modular invariant the antichiral label of a field $\bar{i}$ is determined by its chiral label i. The resulting relation has the form

$$B^{a}_{i} B^{a}_{j} = \sum_{l} X_{ij}{}^{l} B^{a}_{l} \qquad (6.163)$$

for all a with a-independent structure constants $X_{ij}{}^{l}$ that vanish if the fusion rules $N_{ij}{}^{l}$ are zero. The number of different solutions of these equations also determines the number of allowed boundary conditions. In order to compute the values of the structure constants $X_{ij}{}^{l}$, one needs to know the two-dimensional structure constants and the expressions for the fusion matrix in the model. As already stressed, these data are known only in a very restricted number of cases. In order to bypass this difficulty, in [55] an alternative approach was proposed. One can postulate that (6.163) holds and that the structure constants $X_{ij}{}^{l}$ form a commutative and associative algebra, called a classifying algebra. Then the reflection coefficients B^{a}_{i} are given by the representations of this algebra which, in some cases, can be explicitly found.

For the $SU(2)$ case from the explicit expressions of the fusion matrix (6.119) and the two-dimensional structure constants (6.128), we can compute the values

Table 6.1. Reflection coefficients for the diagonal $SU(2)$ level $k = 6$ model.

a	B_1	B_3	B_5	B_7	B_2	B_4	B_6
1	1	$1+\sqrt{2}$	$1+\sqrt{2}$	1	$\sqrt{2+\sqrt{2}}$	$\sqrt{2(2+\sqrt{2})}$	$\sqrt{2+\sqrt{2}}$
2	1	1	-1	-1	$\sqrt{2}$	0	$-\sqrt{2}$
3	1	$1-\sqrt{2}$	$1-\sqrt{2}$	1	$\sqrt{2-\sqrt{2}}$	$-\sqrt{2(2-\sqrt{2})}$	$\sqrt{2-\sqrt{2}}$
4	1	-1	1	-1	0	0	0
5	1	$1-\sqrt{2}$	$1-\sqrt{2}$	1	$-\sqrt{2-\sqrt{2}}$	$\sqrt{2(2-\sqrt{2})}$	$-\sqrt{2-\sqrt{2}}$
6	1	1	-1	-1	$-\sqrt{2}$	0	$\sqrt{2}$
7	1	$1+\sqrt{2}$	$1+\sqrt{2}$	1	$-\sqrt{2+\sqrt{2}}$	$-\sqrt{2(2+\sqrt{2})}$	$-\sqrt{2+\sqrt{2}}$

of $X_{ij}{}^l$ both in the diagonal A models and in the non-diagonal D_{odd} models, obtaining

$$B_i^a B_j^a = \sum_l \epsilon_{ijl} N_{ij}{}^l B_l^a \tag{6.164}$$

where the signs ϵ_{ijl}, present only for the D_{odd} models, are defined after equation (6.128) (they are symmetric in all three indices and are equal to (-1) only if two of the isospins are half-integer, while the third isospin is an odd integer).

As an illustration we shall write down the solutions of the system (6.164) in the two $SU(2)$ models of level $k = 6$. In the diagonal A model there are seven different solutions for the reflection coefficients B_i, which are reported in table 6.1. Note that in the diagonal models the number of boundary conditions is always equal to the number of two-dimensional fields.

In the non-diagonal D_5 model, with torus partition function

$$Z_T^{D_5} = |\chi_1|^2 + |\chi_3|^2 + |\chi_5|^2 + |\chi_7|^2 + |\chi_4|^2 + \chi_2\bar{\chi}_6 + \chi_6\bar{\chi}_2 \tag{6.165}$$

two of the coefficients (B_2 and B_6) vanish, since the corresponding two-dimensional fields are non-diagonal, while the presence of the signs ϵ_{ijl} modifies the equations for B_4 as follows:

$$\begin{aligned} B_4 B_{2I+1} &= (-1)^I B_4 \\ B_4 B_4 &= B_1 - B_3 + B_5 - B_7. \end{aligned} \tag{6.166}$$

Hence there are only five different solutions for the reflection coefficients B_i, which are reported in table 6.2. Note that the number of different boundary conditions is again equal to the number of two-dimensional fields with charge conjugate chiral and antichiral labels (or, equivalently, to the number of different Ishibashi states (6.140)). This, in fact, is a general property of two-dimensional conformal theories with boundaries [55,56].

Table 6.2. Reflection coefficients for the non-diagonal $SU(2)$ level $k = 6$ model.

a	B_1	B_3	B_5	B_7	B_4
1	1	1	-1	-1	0
2	1	$1+\sqrt{2}$	$1+\sqrt{2}$	1	0
3	1	-1	1	-1	2
4	1	-1	1	-1	-2
5	1	$1-\sqrt{2}$	$1-\sqrt{2}$	1	0

6.4.2 Closed unoriented sector, crosscap constraint

To study the behaviour of the two-dimensional fields on non-oriented surfaces let us first introduce the crosscap. The crosscap is the projective plane and can be represented as a unit disc with diametrically opposite points identified. Two-dimensional surfaces with crosscaps cannot be oriented. For example the Klein bottle is topologically equivalent to a cylinder terminating at two crosscaps.

Our analysis will follow closely the one in the boundary case. Like the boundaries, the crosscap breaks the two-dimensional conformal symmetry since it is not invariant under all transformations of $Vir \otimes \overline{Vir}$. If the central charges of the two algebras are equal ($\bar{c} = c$) there exist crosscaps that are preserved at most by the diagonal subalgebra Vir_{diag}. Let us again pass to the analytic picture mapping the boundary of the crosscap onto the unit circle. Then the crosscap implies the identification $\bar{z} \leftrightarrow -1/z$. Similarly to the case of a boundary, the absence of energy flux through the crosscap allows us to define the stress–energy tensor as

$$T_d(z) = \begin{cases} T(z) & \text{for } |z| \le 1 \\ \dfrac{1}{z^4}\bar{T}\left(-\dfrac{1}{z}\right) & \text{for } |z| > 1 \end{cases} \tag{6.167}$$

while the currents are

$$J_d^a(z) = \begin{cases} J^a(z) & \text{for } |z| \le 1 \\ -\dfrac{1}{z^2}\bar{J}^a\left(-\dfrac{1}{z}\right) & \text{for } |z| > 1 \end{cases} \tag{6.168}$$

The combinations of the Laurent modes of the stress–energy tensor and of the currents that satisfy the Virasoro algebra with vanishing central charge (6.136) and the current algebra of zero level (6.137) are, in this case,

$$\mathcal{L}_n = L_n - (-1)^n \bar{L}_{-n} \tag{6.169}$$

and

$$\mathcal{J}^a{}_n = J_n^a + (-1)^n \bar{J}_{-n}^a. \tag{6.170}$$

The crosscap states $|C\rangle$ [57] in the theory are annihilated by the modes (6.169) and (6.170):

$$\mathcal{L}_n|C\rangle = (L_n - (-1)^n \bar{L}_{-n})|C\rangle = 0 \qquad (6.171)$$

and

$$\mathcal{J}^a{}_n|C\rangle = (J^a_n + (-1)^n \bar{J}^a_{-n})|C\rangle = 0. \qquad (6.172)$$

These equation have the same number of solutions as the corresponding equations (6.138) and (6.139) for the boundary states and one can explicitly construct the Ishibashi-type crosscap states as in (6.140). There is, however, an important difference with the boundary case, since the consistency conditions imply the crosscap constraint [58, 59], which singles out one crosscap state $|C\rangle$. Let us stress that, in general, there may be several different crosscap states corresponding to different actions of the involution $\Omega : z \leftrightarrow -1/z$ on the fields. The crosscap constraint tells us only that two different crosscap states cannot exist simultaneously in the same theory.

Just as in the boundary case, the presence of a crosscap implies that the n-point functions of the two-dimensional primary fields are linear combinations of the $2n$-point chiral conformal blocks. However, in contrast with the boundary case one cannot introduce non-trivial crosscap operators, since the involution $\Omega : z \leftrightarrow -1/z$ has no fixed points. In particular, only the identity operator (which has no z dependence and hence is the only invariant under Ω one) can contribute to the expansion of a two-dimensional primary field in front of a crosscap:

$$\phi_{\Lambda\bar{\Lambda}}(z, \bar{z})|_{\text{crosscap}} \sim \Gamma_{\Lambda\bar{\Lambda}}\delta_{\Lambda\bar{\Lambda}^C}1. \qquad (6.173)$$

Here $\Gamma_{\Lambda\bar{\Lambda}}$ is a normalization constant and $\bar{\Lambda}^C$ is the charge conjugate of $\bar{\Lambda}$. Let us stress that the expansion (6.173) can be used only for the computation of the 1-point functions of the fields in front of a crosscap. The reason is that the operator product expansions are valid only if the arguments can be connected without encountering other singularities; but in all $n \geq 2$ point functions in front of a crosscap z and $\bar{z} = -1/z$ are always separated by the arguments of the other fields.

The involution Ω acts on the two-dimensional primary fields (6.141) transforming the chiral vertex operators into antichiral ones and vice versa; and thus relating the two-dimensional field (6.141) to the field with weights and arguments exchanged:

$$\phi_{\bar{\Lambda}\Lambda}(\bar{z}, z) = \sum_{\substack{\Lambda_i \ \bar{\Lambda}_i \\ \Lambda_f \ \bar{\Lambda}_f}} V_{\bar{\Lambda}}{}^{\bar{\Lambda}_f}{}_{\bar{\Lambda}_i}(\bar{z})\bar{V}_{\Lambda}{}^{\Lambda_f}{}_{\Lambda_i}(z) n^{f\bar{f}}_{i\bar{i}}. \qquad (6.174)$$

To simplify the notation, let us denote the two weights of the field by a single label (this is unambiguous for a permutation modular invariant) setting $\phi_i = \phi_{\Lambda_i\bar{\Lambda}_i}$ and $\phi_{\bar{i}} = \phi_{\bar{\Lambda}_i\Lambda_i}$. The action of Ω is [8]

$$\Omega\phi_i(z, \bar{z}) = \epsilon_i\phi_{\bar{i}}(\bar{z}, z). \qquad (6.175)$$

Since Ω is an involution, the ϵ_i are just signs

$$\epsilon_i = \epsilon_{\bar{i}} = \pm 1 \tag{6.176}$$

which have to respect the fusion rules (6.71), hence

$$\epsilon_i \epsilon_j \epsilon_k = 1 \qquad \text{if } N_{ijk} \neq 0. \tag{6.177}$$

As an example let us again take the $SU(2)$ current algebra. In this case the equations (6.177) have only two different solutions: $\epsilon_i = +1$ for all integer isospin fields and $\epsilon_i = \epsilon = \pm 1$ for all half-integer isospin fields

One convenient way to compute the n-point functions of the two-dimensional fields in the front of a crosscap is to introduce the crosscap operator [59]

$$\hat{C} = \sum_l \Gamma_l \frac{|\Delta_l\rangle\langle\bar{\Delta}_l|}{\sqrt{N_l}} \tag{6.178}$$

where N_l is the normalization constant of the two-dimensional 2-point function (6.126). the operator $\hat{C}$ allows us to explicitly correlate the n-point functions of the two-dimensional fields in presence of a crosscap with the $2n$-point chiral conformal blocks:

$$\begin{aligned}\langle\phi_{1,\bar{1}}\dots\phi_{n,\bar{n}}\rangle_C &= \langle 0|\hat{C}\phi_{1,\bar{1}}\dots\phi_{n,\bar{n}}|0\rangle \\ &= \sum_l \frac{\Gamma_l}{\sqrt{N_l}}\langle 0|V_{\Delta_1}(z_1)\dots V_{\Delta_n}(z_n)|\Delta_l\rangle\langle\bar{\Delta}_l|\bar{V}_{\bar{\Delta}_1}(\bar{z}_1)\dots\bar{V}_{\bar{\Delta}_n}(\bar{z}_n)|0\rangle.\end{aligned} \tag{6.179}$$

The relation (6.175) for the two-dimensional fields implies for their functions in the presence of a crosscap:

$$\langle\phi_{i,\bar{i}}(z_i,\bar{z}_i)X\rangle_C = \epsilon_{(i,\bar{i})}\langle\phi_{\bar{i},i}(\bar{z}_i,z_i)X\rangle_C \tag{6.180}$$

where X is an arbitrary polynomial in the fields. These equations determine the coefficients Γ_n. In particular, for the 1-point functions which satisfy

$$\begin{aligned}\langle\phi_{i,\bar{i}}(z,\bar{z})\rangle_C &= \sum_l \frac{\Gamma_l}{\sqrt{N_l}}\langle 0|V_i(z)|\Delta_l\rangle\langle\bar{\Delta}_l|\bar{V}_{\bar{i}}(\bar{z})|0\rangle \\ &= \frac{\Gamma_i}{\sqrt{N_i}}\delta_{i\bar{i}}\langle 0|V_i(z)V_{\bar{i}}(\bar{z})|0\rangle = \langle\phi_{\bar{i},i}(\bar{z},z)\rangle_C\end{aligned} \tag{6.181}$$

equation (6.180) implies the vanishing of Γ_ℓ for all fields on which Ω acts non-trivially ($\epsilon_\ell = -1$). Note that the factor $\sqrt{N_i}$ in (6.181) is compensated by the normalization of the chiral function $\langle 0|V_i\ V_{\bar{i}}|0\rangle$ in accord with (6.173).

To derive the crosscap constraint, let us apply (6.180) for the 2-point functions in the presence of a crosscap. The left-hand side is

$$\begin{aligned}
&\langle \phi_{i,\bar{i}}(z_1,\bar{z}_1)\phi_{j,\bar{j}}(z_2,\bar{z}_2)\rangle_C \\
&\quad = \sum_l \frac{\Gamma_l}{\sqrt{N_l}}\langle 0|V_i(z_1)\ V_j(z_2)|\Delta_l\rangle\langle\bar{\Delta}_l|\bar{V}_{\bar{i}}(\bar{z}_1)\ \bar{V}_{\bar{j}}(\bar{z}_2)|0\rangle \\
&\quad = \sum_l \Gamma_l \tilde{C}^{(l,l)}_{(i,\bar{i})(j,\bar{j})} S_l(z_1,z_2,\bar{z}_1,\bar{z}_2)
\end{aligned} \tag{6.182}$$

where S_l are the normalized s-channel chiral conformal blocks (6.114) (note the order of the arguments z_i). The constants $\tilde{C}$ are proportional to the two-dimensional structure constants (6.128):

$$\tilde{C}^{(l,l)}_{(i,\bar{i})(j,\bar{j})} = \sqrt{N_l}C^{(l,l)}_{(i,\bar{i})(j,\bar{j})}. \tag{6.183}$$

In the same way for the right-hand side we obtain

$$\begin{aligned}
&\langle \phi_{\bar{i},i}(\bar{z}_1,z_1)\phi_{j,\bar{j}}(z_2,\bar{z}_2)\rangle_C \\
&\quad = \sum_l \frac{\Gamma_l}{\sqrt{N_l}}\langle 0|V_{\bar{i}}(\bar{z}_1)V_j(z_2)|\Delta_l\rangle\langle\bar{\Delta}_l|\bar{V}_i(z_1)\bar{V}_{\bar{j}}(\bar{z}_2)|0\rangle \\
&\quad = \sum_l \Gamma_l \tilde{C}^{(l,l)}_{(\bar{i},i)(j,\bar{j})} S_l(\bar{z}_1,z_2,z_1,\bar{z}_2).
\end{aligned} \tag{6.184}$$

The s-channel blocks $S_l(\bar{z}_1,z_2,z_1,\bar{z}_2)$ are proportional to the u-channel blocks $U_l(z_1,z_2,\bar{z}_1,\bar{z}_2)$ (see equation (6.120)) and can be related to the conformal blocks in (6.182) by the exchange operator $B_1(B_3)^{-1}F$. Using also the explicit form of B_1 and B_3 (6.117), we find

$$S_l(\bar{i},j,i,\bar{j}) = (-1)^{\Delta_i-\bar{\Delta}_i+\Delta_j-\bar{\Delta}_j}\sum_n F_{ln}(i,j,\bar{i},\bar{j})S_n(i,j,\bar{i},\bar{j}). \tag{6.185}$$

Inserting (6.182), (6.184), (6.185) into equation (6.180) we obtain the final form of the crosscap constraint [59]:

$$\epsilon_{(i,\bar{i})}(-1)^{\Delta_i-\bar{\Delta}_i+\Delta_j-\bar{\Delta}_j}\Gamma_n\tilde{C}^{(n,n)}_{(i,\bar{i})(j,\bar{j})} = \sum_l \Gamma_l\tilde{C}^{(l,l)}_{(\bar{i},i)(j,\bar{j})}F_{ln}(i,j,\bar{i},\bar{j}) \tag{6.186}$$

for all n. Applying Ω to the second field in the 2-point function leads to the same equation. Note that the crosscap constraint is linear in Γ, hence it only determines the ratios Γ_l/Γ_1. The remaining freedom is only in the normalization of the two-dimensional identity operator in front of the crosscap Γ_1. The simplest way to determine Γ_1 is to impose the integrality condition on the partition functions which we shall describe in the next section. An alternative approach would be to use the topological equivalence of three crosscaps to a handle and one crosscap that is expected to give a nonlinear relation for Γ_l. The explicit form of this relation is, however, still not known.

6.5 Partition functions

The two-dimensional structure constants are explicitly known only in a very limited number of cases. This does not allow us, in general, to compute the n-point functions in the presence of boundaries or crosscaps and to solve the sewing constraints. Here we shall describe an alternative approach, proposed in [4] in the framework of string theory. It gives less detailed information about the theory but is applicable in all cases when the modular matrices S and T (6.67) are known. Just as modular-invariant torus partition functions are classified in many cases when the structure constants are not known, the partition function on the annulus and the Klein bottle and Möbius strip projections can be explicitly computed in many cases when we cannot obtain detailed information about the corresponding n-point functions in the presence of boundaries or crosscaps. The method is particularly powerful if the completeness of the boundary conditions [9] is used. Note that modular invariance of the torus partition function also plays the role of completeness condition for the two-dimensional fields.

One starts with a general (not necessary rational) two-dimensional theory with isomorphic chiral and antichiral observable algebras $\mathcal{A}$ and $\bar{\mathcal{A}}$, corresponding to a symmetric $X_{ij} = X_{ji}$ torus modular invariant (6.63).

To simplify the formulae we shall assume that the theory is rational and that the modular invariant is of the permutation type (6.69). This has the advantage that one can write all expressions using only chiral labels, while in the general case additional degeneracy labels may be needed to distinguish fields with multiplicities larger than one.

6.5.1 Klein bottle projection

Let us first construct the non-oriented sector. The simplest non-orientable surface, the Klein bottle, can be represented as a cylinder terminating at two crosscaps. The Klein bottle contribution to the partition function is a linear combination of the Virasoro characters [4–6], hence, in general, it is not a modular invariant. In fact, there are two distinct expressions for the Klein bottle contribution, called the direct and transverse channels which are related by the modular S transformation (6.64). In string theory language they correspond to inequivalent choices of time on the world-sheet. In the direct channel the Klein bottle contribution is a projection of the torus partition function that describes the (anti)symmetry properties of the two-dimensional fields under the involution Ω (6.175):

$$K = \sum_i \chi_i K^i \tag{6.187}$$

where the integers K^i satisfy

$$|K^i| \leq X_{ii} \qquad K^i = X_{ii} \pmod 2. \tag{6.188}$$

Hence for permutation invariants, K^i can take only the values 0 or ± 1. The K_i are related (but not necessary equal) to the signs ϵ_i in (6.175).

The modular S transformation turns (6.187) into the transverse channel, which describes the reflection of the two-dimensional fields from the two crosscaps at the ends of the cylinder. It has the form

$$\tilde{K} = \sum_i \chi_i \Gamma_i{}^2 \tag{6.189}$$

where the reflection coefficients Γ_i are the normalizations of the 1-point functions of the two-dimensional fields in front of the crosscap (see equation (6.173)), so they vanish if $X_{iiC} = 0$.

The complete partition function in the unoriented case is given by the half sum of the torus and direct channel Klein bottle contributions

$$Z_{\text{unoriented}} = \tfrac{1}{2}(Z_T + K). \tag{6.190}$$

The multiplicity of a field $\phi_{ij} (= \phi_{ji})$ can be read from the partition function (6.190) as follows (for a permutation invariant, if there are multiplicities the argument applies for each copy of the fields).

- If $i \neq j$ it is equal to $1/2(X_{ij} + X_{ji})$ and is non-negative integer due to the assumption that the torus invariant is symmetric. Only one combination of the two fields ϕ_{ij} and ϕ_{ji} remains in the spectrum, the other is projected out.
- If $i = j$ it is equal to $1/2(X_{ii} + K_i)$ and is a non-negative integer since K_i satisfy (6.188). In particular, if $X_{ii} = 1$ the fields with $K_i = 1$ remain in the spectrum, while the ones with $K_i = -1$ are projected out.

If the ground state is degenerate, the Klein bottle projects out the part antisymmetric under the left-right exchange rather that the whole field.

We shall illustrate the construction on the example of the non-diagonal D_5 model of the $SU(2)$ current algebra with level $k = 6$ with torus partition function (6.165). There are two different Klein bottle projections, corresponding to the two choices for the signs ϵ_i in (6.175) [59]. For reasons that will become clear in the next section, we shall distinguish them by the subscripts ‘r’ and ‘c’ (for ‘real’ and ‘complex’)

$$K_{\mathrm{r}}^{D_5} = \chi_1 + \chi_3 + \chi_5 + \chi_7 - \chi_4 \tag{6.191}$$
$$K_{\mathrm{c}}^{D_5} = \chi_1 + \chi_3 + \chi_5 + \chi_7 + \chi_4. \tag{6.192}$$

Comparison with the values of ϵ_i given by

- $\epsilon_i = 1$ for all i in the real case
- $\epsilon_i = (-1)^{i-1}$ in the complex case

shows that there is a relative factor $(-1)^{2I}$ between K_i and the signs ϵ_i which comes from the $SU(2)$ structure of the fields. Indeed the singlet is in the symmetric (antisymmetric) part of the tensor product of integer (half-integer) isospins. So the real Klein bottle projection corresponds to keeping all singlets, while the complex one projects out the singlet corresponding to χ_4.

As an application of these ideas to string theory, let us mention that in [60] the first tachyon free non-supersymmetric string model has been constructed by a non-standard Klein bottle projection.

6.5.2 Annulus partition function

The spectrum of the boundary fields is described by the annulus (or cylinder) partition function with all possible boundary conditions at the two ends. Again the partition function is linear in the characters, hence not a modular invariant, so there are two distinct expressions for the annulus contribution [1]. They are called the direct and transverse channels and are related by the modular S transformation (6.64).

In the direct channel the annulus partition function counts the number of operators that intertwine the boundary conditions at the two ends and can be represented as

$$A = \sum_{i,a,b} \chi^i A_i^{ab} n_a n_b \tag{6.193}$$

where the non-negative integers A_i^{ab} give the multiplicities of the boundary fields ψ_i^{ab}. The auxiliary multiplicities n^a associated with the boundaries in open string models correspond to the introduction of Chan–Paton gauge groups [61], which can be $U(n)$, $O(n)$ or $USp(2n)$ [62]. In the case of $U(n)$ groups, the boundaries can be oriented, since there are two inequivalent choices of the fundamental representation, hence the Chan–Paton charges come in numerically equal pairs $\bar{n} = n$. We shall call such charges complex. The other two cases, $USp(2n)$ and $O(2n)$, do not lead to similar identifications and we shall call the corresponding charges real. The labels 'r' and 'c' on the partition functions originate from this interpretation. In applications to Statistical Mechanics one may regard (6.193) as a generating function for the multiplicities of the allowed boundary fields.

The transverse channel, related to (6.193) by a modular S transformation, has a very different interpretation. It describes the reflection of a two-dimensional field from the two boundaries and can be represented as

$$\tilde{A} = \sum_i \chi^i \left[\sum_a \mathcal{B}_{ia}\, n^a \right]^2 . \tag{6.194}$$

Since only fields with charge conjugate chiral and antichiral labels can couple to the boundaries, it is again sufficient to specify only the chiral label. The reflection coefficient $\mathcal{B}_{ia}$ for the field i $(\bar{i})$ from a boundary a is proportional to the coefficient of the identity operator in the bulk-to-boundary expansion of the two-dimensional field in front of the boundary (6.151):

$$\mathcal{B}_{ia} = \frac{C^{a}_{(i,\bar{i})1}\alpha^{aa}_{1}}{\sqrt{N_{i\bar{i}}}}. \tag{6.195}$$

One can define charge conjugation on the boundary labels. It is non-trivial only if the boundaries are oriented (that corresponds to complex charges) and is given by the involutive matrix $(A_1)_{ab} = (A_1)^{ab}$, such that

$$A_{ia}{}^{b} = \sum_{c} A_{1ac} A_{i}{}^{cb} \tag{6.196}$$

hence $(A_1)^b_a = \delta^b_a$. Let us also assume that the boundaries are a complete set. To justify this assumption let us recall that the modular invariance condition of the torus partition function also plays the role of completeness condition for the two-dimensional fields. The completeness condition for the boundaries has two equivalent formulations. The first one [9] is to require that the coefficients $A_{ia}{}^{b}$ satisfy the fusion algebra

$$\sum_{b} A_{ia}{}^{b} A_{jb}{}^{c} = \sum_{k} N^{k}_{ij} A_{ka}{}^{c}. \tag{6.197}$$

Intuitively this relation corresponds to two different ways of counting the boundary fields. The second one [56] is to require that the boundary states are related to the Ishibashi states (6.140) by a unitary transformation which, in particular, implies that they are the same number.

Equation (6.197) contains only chiral information, so it cannot determine completely the multiplicities A^{ab}_{i}. The two-dimensional input is provided by the torus modular invariant (6.62). In particular, if for some j the torus coefficient $X_{jj^C} = 0$ (where j^C is the charge conjugate of j) then there is no two-dimensional field with these labels, so the coefficients $C^{a}_{(jj^C)1}$ are zero for all a. Hence, due to (6.195) all the $\mathcal{B}_{ja}$'s also vanish and χ_j will not contribute to (6.194). After a modular transformation this implies

$$\sum_{i} A^{ab}_{i} S^{i}_{j} = 0 \tag{6.198}$$

for all a and b and this particular j.

Hence we can reformulate the problem of finding the annulus partition function in the following way: solve over the non-negative integers the two equations (6.197) and (6.198). In general, this system may have several solutions but in all known cases fixing the boundary charge conjugation matrix $(A_1)_{ab}$ also completely determines all A^{ab}_{i}, and thus the only freedom is in choosing the orientation on pairs of boundaries. The proof of this fact in the general case is, however, still a challenging open problem.

As an illustration let us again consider the D_5 model with torus partition function (6.165). In the real charge case $(A_1)_{ab} = \delta_{ab}$ and the solution is (the labels of the charges correspond to the first column in table 6.2)

$$\begin{aligned}
A_{\mathrm{r}}^{D_5} = {} & \chi_1(n_1^2 + n_2^2 + n_3^2 + n_4^2 + n_5^2) \\
& + (\chi_2 + \chi_6)(2n_1n_2 + 2n_1n_5 + 2n_3n_5 + 2n_4n_5) \\
& + \chi_3(n_1^2 + 2n_1n_3 + 2n_1n_4 + 2n_3n_4 + 2n_2n_5 + 2n_5^2) \\
& + \chi_4(4n_1n_5 + 2n_2n_3 + 2n_3n_5 + 2n_2n_4 + 2n_4n_5) \\
& + \chi_5(n_1^2 + n_3^2 + n_4^2 + 2n_5^2 + 2n_1n_3 + 2n_1n_4 + 2n_2n_5) \\
& + \chi_7(n_1^2 + n_2^2 + n_5^2 + 2n_3n_4).
\end{aligned} \tag{6.199}$$

In the complex case the two charges n_3 and n_4 become a complex pair $\bar{n} = n$ and the solution is

$$\begin{aligned}
A_{\mathrm{c}}^{D_5} = {} & \chi_1(n_1^2 + n_2^2 + 2n\bar{n} + n_5^2) \\
& + (\chi_2 + \chi_6)(2n_1n_2 + 2n_1n_5 + 2nn_5 + 2\bar{n}n_5) \\
& + \chi_3(n_1^2 + n^2 + \bar{n}^2 + 2n_1n + 2n_1\bar{n} + 2n_2n_5 + 2n_5^2) \\
& + \chi_4(4n_1n_5 + 2n_2n + 2n_2\bar{n} + 2nn_5 + 2\bar{n}n_5) \\
& + \chi_5(n_1^2 + 2n_5^2 + 2n_1n + 2n_1\bar{n} + 2n_2n_5 + 2n\bar{n}) \\
& + \chi_7(n_1^2 + n_2^2 + n_5^2 + n^2 + \bar{n}^2).
\end{aligned} \tag{6.200}$$

Note that in both cases some boundary fields (corresponding to the n_5 charge) have multiplicities equal to two.

6.5.3 Möbius strip projection

The consistency of the theory in presence of both boundaries and crosscaps is determined by the Möbius strip contribution [4–6]. The Möbius strip can be represented as a cylinder terminating at one boundary and at one crosscap. Hence, in the transverse channel the two-dimensional field reflects from the boundary and the crosscap with the same reflection coefficients $\mathcal{B}_{ia}$ and Γ_i which enter equations (6.189), (6.194)

$$\tilde{M} = \sum_i \hat{\chi}^i \Gamma_i \left[\sum_a \mathcal{B}_{ia} n^a \right]. \tag{6.201}$$

As we have seen there are, in general, more than one solution for both $\mathcal{B}_{ia}$ and Γ_i, so we also have to specify which of these solutions we shall use in equation (6.201). To determine this we can pass to the direct channel (by a P transformation, see equation (6.205))

$$M = \sum_i \hat{\chi}^i M_i^a n_a \tag{6.202}$$

and compare this expression with the annulus partition function (6.193). The integer coefficients M_i^a can be interpreted as twists (or projections) of the open spectrum and thus have to satisfy

$$M_i^a = A_i^{aa} \pmod 2 \qquad |M_i^a| \leq A_i^{aa}. \tag{6.203}$$

These equations choose consistent pairs of annulus and Klein bottle partition functions.

The natural modular parameter in the direct channel for the Möbius strip is $(\mathrm{i}\tau + 1)/2$, while in the transverse channel it is $(\mathrm{i} + \tau)/2\tau$. The non-vanishing real part of the direct channel modular parameter implies that the natural basis of characters for the Möbius strip is

$$\hat{\chi}_j = \mathrm{e}^{-\mathrm{i}\pi(\Delta_j - c/24)} \chi_j \left(\frac{\mathrm{i}\tau + 1}{2} \right) \tag{6.204}$$

hence the transformation which relates the direct and transverse channels is given by [5]

$$P = T^{1/2} S T^2 S T^{1/2} \tag{6.205}$$

and satisfies $P^2 = C$. The square root of T in (6.205) denotes the diagonal matrix whose eigenvalues are square roots of the eigenvalues of T.

By a formula similar to the Verlinde formula (6.71) one can define the coefficients ${Y_{ij}}^k$ [8]:

$${Y_{ij}}^k = \sum_\ell \frac{S_{i\ell} P_{j\ell} P^\dagger_{k\ell}}{S_{1\ell}} \tag{6.206}$$

which are integers [63, 64] and satisfy the fusion algebra

$$\sum_l {Y_{im}}^l {Y_{jl}}^n = \sum_\ell {N_{ij}}^\ell {Y_{\ell m}}^n \tag{6.207}$$

$$\sum_i Y_{ijk} {Y^i}_{lm} = \sum_i Y_{ijm} {Y^i}_{lk}. \tag{6.208}$$

The complete partition function in the unoriented open sector is

$$Z_{\text{open}} = \tfrac{1}{2}(A \pm M). \tag{6.209}$$

Its integrality is guaranteed by the conditions (6.203). Note that the overall sign of the Möbius strip projection is not determined by conformal theory. In open string models this sign is fixed by the tadpole cancellation conditions [65] and determines the gauge group.

The completeness condition (6.197) implies two relations between the integer coefficients in the direct channel partition functions A, M and K and the

Y tensor (6.206):

$$\sum_b A_i{}^{ab} M_{jb} = \sum_l Y_{ij}{}^l M_l{}^a \tag{6.210}$$

$$\sum_b M_i{}^b M_{jb} = \sum_l Y^l{}_{ij} K_l \tag{6.211}$$

that put very strong constraints on K_i and $M_i{}^a$ for given $A_i{}^{ab}$ (in all known cases they completely determine them).

Coming back to our example, in the non-diagonal D_5 model there are two consistent choices for the annulus and Klein bottle partition functions, namely the pairs with the same subscript (r or c). The two Möbius strip projections are correspondingly

$$\begin{aligned} M_{\mathrm{r}}^{D_5} = &\ \hat{\chi}_1(n_1 - n_2 + n_3 + n_4 - n_5) \\ &+ \hat{\chi}_3(-n_1 + 2n_5) \\ &+ \hat{\chi}_5(n_1 + n_3 + n_4) \\ &+ \hat{\chi}_7(n_1 + n_2 + n_5) \end{aligned} \tag{6.212}$$

and

$$\begin{aligned} M_{\mathrm{c}}^{D_5} = &\ \hat{\chi}_1(-n_1 + n_2 + n_5) \\ &+ \hat{\chi}_3(n_1 + n + \bar{n}) \\ &+ \hat{\chi}_5(n_1 + 2n_5) \\ &+ \hat{\chi}_7(n_1 + n_2 + n + \bar{n} + n_5). \end{aligned} \tag{6.213}$$

It is instructive to verify that these indeed satisfy the polynomial equations and to determine the open spectrum of the models. Note that when the annulus coefficient is equal to $2n_5^2$, there are two possibilities for the Möbius strip coefficient. It can be either $2n_5 = n_5 + n_5$ or $0 = n_5 - n_5$. This corresponds to two operators with equal or opposite symmetrization properties.

6.5.4 Solutions for the partition functions

If the torus modular invariant is given by the charge conjugation matrix $X = C$ then the number of boundary conditions coincides with the number of chiral representations, so we can label both by the same label. In this case the standard solution for the annulus was found in [1], while the expressions for the Klein bottle and Möbius strip were found in [8]:

$$A_{ijk} = N_{ijk} \tag{6.214}$$

$$M_{ij} = Y_{ji1} \tag{6.215}$$

$$K_i = Y_{i11}. \tag{6.216}$$

Using the properties of N_{ijk} and Y_{ijk} it is straightforward to verify that these solutions satisfy all the consistency requirements. Moreover, the standard Klein bottle projection (6.216) is equal to the Frobenius–Schur indicator [64] and corresponds to keeping all the singlets in the spectrum. A modular transformation to the transverse channel gives

$$\tilde{A} = \sum_i \left(\sum_j \frac{S_{ij} n^j}{\sqrt{S_{1i}}} \right)^2 \chi_i \tag{6.217}$$

$$\tilde{M} = \sum_i \left(\sum_j \frac{P_{1i} S_{ij} n^j}{S_{1i}} \right) \hat{\chi}_i \tag{6.218}$$

$$\tilde{K} = \sum_i \left(\frac{P_{1i}}{\sqrt{S_{1i}}} \right)^2 \chi_i . \tag{6.219}$$

Before these general formulae were known, in [66] the standard Klein bottle and Möbius partition functions for the diagonal case of the unitary minimal models had been explicitly constructed.

As a simple example of a non-standard solution we shall also list the expressions for the second possible solution in the diagonal $SU(2)$ current algebra models of level k denoted by A_{k+1}. The modular matrices S and T (we label the fields by $j = 2I + 1$) are

$$S_{jl} = \sqrt{\frac{2}{k+2}} \sin\left(\frac{\pi j l}{k+2} \right) \tag{6.220}$$

$$T_{jl} = \delta_{jl} \mathrm{e}^{\mathrm{i}\pi \left(\frac{j^2}{2(k+2)} - \frac{1}{4} \right)} . \tag{6.221}$$

The charge conjugation matrix is the identity $C = S^2 = (ST)^3 = 1$. The modular matrix $P = T^{1/2} S T^2 S T^{1/2}$ which satisfies $P^2 = C = 1$ is

$$P_{jl} = \frac{2}{\sqrt{k+2}} \sin\left(\frac{\pi j l}{2(k+2)} \right) (E_k E_{j+l} + O_k O_{j+l}) \tag{6.222}$$

where E_n and O_n are projectors on n even and odd correspondingly.

The standard solution in the diagonal model has $k + 1$ real charges and is given by (6.214), (6.219). The explicit expression for the direct-channel Klein bottle projection is

$$K_{\mathrm{r}}^{\{A_{k+1}\}} = \sum_{j=1}^{k+1} Y^j{}_{11} \chi_j = \sum_{j=1}^{k+1} (-1)^{j-1} \chi_j \tag{6.223}$$

hence indeed all singlets are kept in the unoriented spectrum.

The second solution has also $k+1$ charges (most are in complex pairs) and in the direct channel is given by [8]

$$K_{\rm c}^{\{A_{k+1}\}} = \sum_{j=1}^{k+1} Y^j{}_{k+1,k+1}\chi_j = \sum_{j=1}^{k+1}\chi_j \tag{6.224}$$

$$A_{\rm c}^{\{A_{k+1}\}} = \sum_{j,l,m=1}^{k+1} N_{lm}{}^j\,\chi_{k+2-j}n^l n^m \tag{6.225}$$

$$M_{\rm c}^{\{A_{k+1}\}} = \sum_{j,l=1}^{k+1} Y_{l,k+1}{}^j\,\hat\chi_j n^l. \tag{6.226}$$

Note that the Klein bottle projects out the singlets for all the half-integer isospin fields, so they cannot couple to the identity on the boundaries or the crosscap, hence the corresponding reflection coefficient should vanish. After a modular transformation we find in the transverse channel

$$\tilde K_{\rm c}^{\{A_{k+1}\}} = \sum_i \left(\frac{P_{k+1,i}}{\sqrt{S_{1i}}}\right)^2 \chi_i \tag{6.227}$$

$$\tilde A_{\rm c}^{\{A_{k+1}\}} = \sum_i (-1)^{i-1}\left(\sum_j \frac{S_{ij}n^j}{\sqrt{S_{1i}}}\right)^2 \chi_i \tag{6.228}$$

$$\tilde M_{\rm c}^{\{A_{k+1}\}} = \sum_i \left(\sum_j \frac{P_{k+1,i}S_{ij}n^j}{S_{1i}}\right)\hat\chi_i. \tag{6.229}$$

The vanishing of the reflection coefficients of the fields with half-integer isospin in (6.228) implies the complex charge identifications $n_{k+2-i} = \bar n_i = n_i$ for all i.

In the $D_{\rm odd}$ models there are again two different choices for the Klein bottle projection (which generalize equations (6.191), (6.192) for D_5). Both lead to $k/2+2$ charges. The corresponding annulus and Möbius strip partition functions are rather involved [9, 59]. The solutions for $D_{\rm even}$, E_6 and E_8 (with charge conjugation modular invariants if considered as models with extended symmetry) are given by the general formulae (6.214), (6.219). The solution for the exceptional case E_7 is given in [59]. In the $D_{\rm even}$ and E models one can study also boundary conditions that do not respect the extended symmetry of the bulk model, but only the $SU(2)$ symmetry. The corresponding solutions are given in [53].

Many other solutions have been found. Let us note only the general formulae in [67] where the partition functions for all simple currents modular invariants are given.

Acknowledgments

It is a pleasure to thank the organizers of the Fourth SIGRAV Graduate School and 2001 School on Algebraic Geometry and Physics 'Geometry and Physics of Branes' at Como for giving me the opportunity to present these lectures. This work was supported in part by the EEC contracts HPRN-CT-2000-00122 and HPRN-CT-2000-00148 and by the INTAS contract 99-1-590.

References

[1] Cardy J L 1989 *Nucl. Phys.* B **324** 581
[2] Cardy J L and Lewellen D C 1991 *Phys. Lett.* B **259** 274
[3] Lewellen D 1992 *Nucl. Phys.* B **372** 654
[4] Sagnotti A 1988 *Nonperturbative Quantum Field Theory* ed G Mack *et al* (Oxford: Pergamon) p 521
[5] Pradisi G and Sagnotti A 1989 *Phys. Lett.* B **216** 59
[6] Bianchi M and Sagnotti A 1990 *Phys. Lett.* B **247** 517
Bianchi M and Sagnotti A 1991 *Nucl. Phys.* B **361** 519
[7] Polchinski J 1995 *Phys. Rev. Lett.* **75** 4724
[8] Pradisi G, Sagnotti A and Stanev Ya S 1995 *Phys. Lett.* B **354** 279
[9] Pradisi G, Sagnotti A and Stanev Ya S 1996 *Phys. Lett.* B **381** 97
[10] Alekseev A Yu and Schomerus V 1999 *Phys. Rev.* D **60** 061901
Felder G, Frohlich J, Fuchs J and Schweigert C 2000 *J. Geom. Phys.* **34** 162–90
Huiszoon L R, Schalm K and Schellekens A N 2002 *Nucl. Phys.* B **624** 219
Bachas C, Couchoud N and Windey P 2002 *J. High Energy Phys.* **0206** 027
[11] Petkova V and Zuber J-B 2000 Conformal boundary conditions and what they teach us *Preprint* hep-th/0103007
[12] Felder G, Frohlich J , Fuchs J and Schweigert C 2000 *Phys. Rev. Lett.* **84** 1659
[13] Mack G 1977 *Nucl. Phys.* B **118** 445
[14] Belavin A A, Polyakov A M and Zamolodchikov A B 1984 *Nucl. Phys.* B **241** 333
[15] Virasoro M A 1970 *Phys. Rev.* D **1** 2933
[16] Friedan D, Qiu Z and Shenker S 1984 *Phys. Rev. Lett.* **52** 1575
Friedan D, Qiu Z and Shenker S 1986 *Commun. Math. Phys.* **107** 535
[17] Ramond P 1971 *Phys. Rev.* D **3** 2415
Neveu A and Schwarz J 1971 *Phys. Rev.* D **4** 1109
[18] Friedan D, Qiu Z and Shenker S 1985 *Phys. Lett.* B **151** 37
Bershadsky M, Knizhnik V and Teitelman M 1985 *Phys. Lett.* B **151** 31
[19] Meurman A and Rocha-Caridi A 1986 *Commun. Math. Phys.* **107** 263
[20] Zamolodchikov A B and Fateev V A 1986 *Zh. Eksp. Theor. Fis.* **90** 1553
Di Vecchia P, Petersen J L and Yu M 1986 *Phys. Lett.* B **172** 211
Friedan D, Martinec E and Shenker S 1986 *Nucl. Phys.* B **271** 93
[21] Sugawara H 1968 *Phys. Rev.* **170** 1659
[22] Knizhnik V G and Zamolodchikov A B 1984 *Nucl. Phys.* B **247** 83
[23] Todorov I T 1985 *Phys. Lett.* B **153** 77
[24] Boucher W, Friedan D and Kent A 1986 *Phys. Lett.* B **172** 316
[25] Witten E 1984 *Commun. Math. Phys.* **92** 455

[26] Kac V G 1985 *Infinite Dimensional Lie Algebras* (Cambridge: Cambridge University Press)
[27] Tsuchiya A and Kanie Y 1987 *Lett. Math. Phys.* **13** 303
[28] Moore G and Seiberg N 1989 *Commun. Math. Phys.* **123** 177
Moore G and Seiberg N 1990 *Superstrings 89, Proceedings of the Trieste 1989 Spring School* (Singapore: World Scientific) pp 1–129.
[29] Cardy J L 1986 *Nucl. Phys.* B **270** 186
Cardy J L 1986 *Nucl. Phys.* B **275** 200
[30] Seiberg N and Witten E 1986 *Nucl. Phys.* B **276** 272
[31] Nahm W 1991 *Int. J. Mod. Phys.* A **6** 2837
[32] Buchholz D, Mack G and Todorov I T 1988 *Nucl. Phys. B (Proc. Suppl.)* **5** 20
[33] Moore G and Seiberg N 1989 *Nucl. Phys.* B **313** 16
[34] Verlinde E 1988 *Nucl. Phys.* B **300** 360
Dijkgraaf R and Verlinde E 1988 *Nucl. Phys. B (Proc. Suppl.)* **5** 87
[35] Cappelli A, Itzykson C and Zuber J B 1987 *Nucl. Phys.* B **280** 445
Cappelli A, Itzykson C and Zuber J B 1987 *Commun. Math. Phys.* **113** 1
[36] Kastor D 1987 *Nucl. Phys.* B **280** 304
Cappelli A 1987 *Phys. Lett.* B **185** 82
[37] Gato-Rivera B and Schellekens A N 1992 *Commun. Math. Phys.* **145** 85
Kreuzer M and Schellekens A N 1994 *Nucl. Phys.* B **411** 97
[38] Gannon T 1993 *Nucl. Phys.* B **402** 729
Gannon T 1994 *Commun. Math. Phys.* **161** 233
[39] Christe P and Flüme R 1987 *Nucl. Phys.* B **282** 466
[40] Dotsenko Vl S and Fateev V A 1984 *Nucl. Phys.* B **240** 312
Dotsenko Vl S and Fateev V A 1985 *Phys. Lett.* B **154** 291
Dotsenko Vl S and Fateev V A 1985 *Nucl. Phys.* B **251** 691
[41] Sonoda H 1988 *Nucl. Phys.* B **311** 417
[42] Fredenhagen K, Rehren K H and Schroer B 1989 *Commun. Math. Phys.* **125** 201
[43] Artin E 1947 *Ann. Math.* **48** 101
Birman J S 1974 *Braids, Links and Mapping Class Groups* (Princeton, NJ: Princeton University Press)
[44] Gepner D and Witten E 1986 *Nucl. Phys.* B **278** 493
[45] Zamolodchikov A B and Fateev V A 1986 *Sov. J. Nucl. Phys.* **43** 657
[46] Stanev Ya S, Todorov I T and Hadjiivanov L K 1992 *Phys. Lett.* B **276** 87
[47] Stanev Ya S and Todorov I T 1992 *Chiral Current Algebras and 2-Dimensional Conformal Models* (Université de Lausanne) pp 1–133
[48] Felder G 1989 *Nucl. Phys.* B **317** 215
Felder G, Fröhlich J and Keller G 1989 *Commun. Math. Phys.* **124** 647
[49] Kohno T 1987 *Ann. Inst. Fourier* **37** 139
Babelon O 1988 *Phys. Lett.* B **215** 523
[50] Petkova V B and Zuber J B 1995 *Nucl. Phys.* B **438** 347
[51] Fuchs J and Schweigert C 1999 *Phys. Lett.* B **447** 266
Fuchs J and Schweigert C 1999 *Nucl. Phys.* B **558** 419
Fuchs J and Schweigert C 2000 *Nucl. Phys.* B **568** 163
[52] Ishibashi N 1989 *Mod. Phys. Lett.* A **4** 251
[53] Behrend R, Pierce P, Petkova V and Zuber J-B 2000 *Nucl. Phys.* B **579** 707
[54] Runkel I 1999 *Nucl. Phys.* B **549** 563
Runkel I 2000 *Nucl. Phys.* B **579** 561

[55] Fuchs J and Schweigert C 1997 *Phys. Lett.* B **414** 251

[56] Behrend R, Pierce P, Petkova V and Zuber J-B 1998 *Phys. Lett.* B **444** 163

[57] Callan C G, Lovelace C, Nappi C R and Yost S A 1987 *Nucl. Phys.* B **293** 83

[58] Fioravanti D, Pradisi G and Sagnotti A 1994 *Phys. Lett.* B **321** 349

[59] Pradisi G, Sagnotti A and Stanev Ya S 1995 *Phys. Lett.* B **356** 230

[60] Sagnotti A 1995 *Susy 95* (Gif-sur-Yvette: Éditions Frontières) p 473
Sagnotti A 1997 *Nucl. Phys. B (Proc. Suppl.)* **56** 332

[61] Paton J E and Chan H M 1969 *Nucl. Phys.* B **10** 516

[62] Schwarz J H 1982 *Proc. 6th Johns Hopkins Workshop on Current Problems in High-Energy Particle Theory, Florence, Italy* p 233
Marcus N and Sagnotti A 1982 *Phys. Lett.* B **119** 97

[63] Gannon T 2000 *Phys. Lett.* B **473** 80
Borisov L, Halpern M B and Schweigert C 1998 *Int. J. Mod. Phys.* A **13** 125
Bantay P 1998 *Phys. Lett.* B **419** 175

[64] Bantay P 1997 *Phys. Lett.* B **394** 87
Huiszoon L R, Schellekens A N and Sousa N 1999 *Phys. Lett.* B **470** 95

[65] Green M and Schwarz J 1985 *Phys. Lett.* B **151** 21

[66] Bianchi M, Pradisi G and Sagnotti A 1991 *Phys. Lett.* B **273** 389

[67] Fuchs J, Huiszoon L R, Schellekens A N, Schweigert C and Walcher J 2000 *Phys. Lett.* B **495** 427

Chapter 7

Topics in string tachyon dynamics

César Gómez and Pedro Resco
Instituto de Física Teórica, C-XVI, Universidad Autónoma de Madrid
E-28049-Madrid, Spain[1]

7.1 Introduction

In recent years a new understanding of the dynamic role of tachyons in string theory has started to emerge ([1–17], see [18] for early work on tachyon condensation). For the simplest open bosonic string much evidence on tachyon condensation already exists [1, 6–8, 11, 12]. The tachyon vacuum expectation value characterizing this condensate exactly cancels out the open-string one-loop contribution to the cosmological constant, what we now understand as the D_{25} filling brane tension. The vacuum defined by this condensate is naturally identified with the closed-string vacua. Precise computations of the tachyon potential supporting this picture has been carried out both in open-string field theory [7–10, 12–15] and in background independent open-string field theory [19–24]. At this level of understanding two main problems remain open. First of all, we have the problem of the closed tachyon that survives as an instability of the closed-string vacua defined by the open-string tachyon condensate. Second, we lack a precise understanding of the dynamical mechanism by which the $U(1)$ gauge open degrees of freedom are decoupled from the closed-string spectrum.

Concerning the problem of the closed-string tachyon, the σ-model beta-functions [25,26] indicate that closed tachyon condensation creates a contribution to the cosmological constant of the same type generated by working with non-critical dimensions. The well-known result on the $c = 1$ barrier in the context of linear dilaton backgrounds [27] could indicate a sort of instability that drastically reduces the spacetime dimensions until the safe $D = 2$ is reached.

[1] Unidad de Investigación Asociada al Centro de Física Miguel Catalán (CSIC).

With respect to the problem of the fate of $U(1)$ gauge degrees of freedom after open tachyon condensation—a sort of confinement of open degrees of freedom into a closed spectrum—there are two formal hints. One is the suggestion of a trivial nilpotent BRST charge of type ac_0, for c the ghost field, around the background defined by the tachyon condensate [28]. The other hint comes from observing that the open-string effective Born–Infeld Lagrangian is multiplied by a factor e^{-T} with $T = \infty$ defining the open tachyon condensate [24, 29–31].

In the context of more healthy superstrings without tachyons, the phenomenon of tachyon condensation sheds some new light on the solitonic interpretation of the D-branes. We have two main examples corresponding to pairs D_p–$D_{\bar{p}}$-brane–antibrane which will support an open tachyon on the world-volume spectrum and the case of configurations of unstable non-BPS D-branes. In both cases tachyon condensation will allow us to interpret stable BPS D-branes as topologically stable extended objects, or solitons, of the auxiliary gauge theory defined on the world-volume of the original configuration of unstable D-branes.

The mechanism for decay into closed-string vacua by tachyon condensation can be used to define a new algebraic structure to characterize D-brane stability and D-brane charges, namely K-theory [32–36]. The main ingredient in order to go to K-theory is the use of the stability equivalence with respect to the creation–annihilation of branes. In type IIB D_p-branes of space codimension $2k$ are related to $K(B^{2k}, S^{2k-1})$ and for type IIA D_p-branes of space codimension $2k + 1$ are related to $K^{-1}(B^{2k+1}, S^{2k})$. The characterization of $K(X, Y)$ in terms of triplets [37] (E, F, α) with E, F vector bundles on X and α an isomorphism $\alpha : E|_Y \to F|_Y$ makes the mathematical meaning of the open tachyon field as defining the isomorphism α particularly clear. A similar construction in terms of pairs (E, α) with α an automorphism of E can be carried out for the definition of the higher K^{-1}-group [33].

Finally we would like to point out some striking similarities between the topological characterization of stable D_p-branes in type IIA string and gauge-fixing singularities for unitary gauges [38] of the type of 't Hooft's Abelian projection [41]. Can we learn something of dynamical relevance from this analogy? After the discovery of asymptotic freedom, the Holy Grail of high-energy physics is the solution of the confinement problem. The Abelian projection gauge was originally suggested in [41] as a first step towards a quantitative approach to confinement, i.e. to the computation of the magnetic monopole condensate. The analogy between stable D_p-branes ($p \leq 6$) in type IIA and the magnetic monopoles associated with the Abelian projection gauge singularities seems to indicate, as the stringy analogue of confinement, the decay of the gauge vacua associated with a configuration of unstable D_9-filling branes into a closed-string vacua populated of stable D_p-branes. Another interesting lesson we learn from the analogy is that as for magnetic monopoles in the Abelian projection, which should be considered as physical degrees of freedom independently of whatever the phase, confinement, Higgs or Coulomb of the underlying gauge theory is, the same should be true concerning type IIA D_p-branes, independently of the concrete form of the open tachyon potential. What is relevant to characterize the 'confinement' closed-string phase is the 'dualization' of the original open gauge string degrees of freedom into R–R closed-string fields whose sources are stable D_p-branes. Finally and from a different point of view, another hint suggested by this analogy is the potential relevance of the higher K-group K^{-1} to describe gauge-fixing singularities in ordinary gauge theories. Maybe the answer to the natural question of why the higher K-group K^{-1} is pointing out to some hidden 'M-theoretical' meaning of the gauge θ-parameter.

The present review is not intended to be complete in any sense. It simply covers the material presented by one of us (CG) during the Fourth SIGRAV School on Contemporary Relativity and Gravitational Physics and 2001 School on Algebraic Geometry and Physics[2].

7.2 Why tachyons?

In quantizing string theory in flat Minkowski spacetime there are two constants that should be fixed by consistency, namely the normal ordering constant appearing in the mass formula:

$$M^2 = \frac{4}{\alpha'}(N - a) \tag{7.1}$$

and the dimension D of the spacetime. These two constants determines the Virasoro anomaly

$$[L_m, L_n] = (m - n)L_{m+n} + A(D, a, m)\delta_{m+n} \tag{7.2}$$

with

$$A(D, a, m) = \frac{D}{12}(m^3 - m) + \frac{1}{6}(m - 13m^3) + 2am \tag{7.3}$$

and

$$L_m = L_m^{(\mathrm{matter})} + L_m^{(\mathrm{ghosts})} - a\delta_m. \tag{7.4}$$

Imposing $A(D, a, m) = 0$ implies the standard constraints on the bosonic string, namely $D = 26$ and $a = 1$.

The first consequence of the non-vanishing normal ordering constant a is that the $(mass)^2$ of the ground state ($N = 0$) is negative, i.e. it is a tachyon. In spite of this there is an advantage in this normal ordering value, namely the existence, at the first level, of a massless vector boson in the open case and a massless graviton in the closed case.

A priori, the only consistency requirement we should impose is the absence of negative norm ghost states in the physical Hilbert space. This will allow us to relax the condition on D and a to $D \leq 26$ and $a \leq 1$.

2 Parts of these lectures were also presented at the Second Workshop on Non-commutative Geometry, String Theory, and Particle Physics, Rabat, May 2001 and in the Workshop on New Interfaces between Geometry and Physics, Miraflores, June 2001.

Although in these conditions the open-string theory is perfectly healthy at tree level, we will find unitarity problems for higher-order corrections, more precisely singularity cuts for one-loop non-planar diagrams. In the closed-string case the problems at one loop will show up as a lack of modular invariance. Thus we will limit ourselves to critical dimension $D = 26$ and $a = 1$.

One important place where the normal ordering constant appears in string theory is in the definition of the BRST operator:

$$Q = \sum_m \left(L_m c_{-m} - \frac{1}{2} \sum_n (m-n) c_{-m} c_{-n} b_{m+n} \right) \tag{7.5}$$

with b, c the usual ghost system for the bosonic string. The charge Q can be written in a more compact way as

$$Q = \sum_m (L_m^{\text{(matter)}} + \tfrac{1}{2} L_m^{\text{(ghosts)}} - a\delta_m) c_{-m}. \tag{7.6}$$

Note that the contribution of the normal ordering constant to Q is simply ac_0. This quantity by itself defines a BRST charge—since it is trivially nilpotent $c_0^2 = 0$—with a trivial cohomology.[3]

In standard quantum field theory, a tachyon is not such an unfamiliar object. A good example is, for instance, the Higgs field if we perturb around the wrong vacua $\langle \phi \rangle = 0$. In this sense the presence of a tachyon usually means that we are perturbing around an unstable vacua. In a physically sensible situation we expect the system to roll down to some stable vacua where the tachyon will disappear automatically. In the bosonic string it is not at all clear whether this is the case since we still lack a powerful tool with which to study string theory off-shell. The only real procedure to address this issue is, of course, string field theory.

In superstring theories with spacetime supersymmetry, i.e. type I, type II or heterotic, the tachyons are projected out by imposing GSO. However, even in these cases open-string tachyons can appear if we consider non-BPS D-branes. In these cases the open tachyon is associated with the instabilities of these non-BPS D-branes.

7.3 Tachyons in AdS: The $c = 1$ barrier

A simple way to see the instabilities induced by tachyonic fields with negative $(mass)^2$ is to compute their contribution to the energy in flat Minkowski spacetime. Generically the energy is defined by

$$E = \int \mathrm{d}^{n-1}x \, \mathrm{d}r \, \sqrt{g} [g^{\mu\nu} \partial_\mu \phi^* \partial_\nu \phi + m^2 \phi^* \phi] \tag{7.7}$$

[3] This is the BRST operator recently suggested in [28] to describe the cohomology around the open tachyon condensate.

where n is the spacetime dimension. The condition of finite energy requires an exponential falloff $\phi \sim e^{-\lambda r}$ at infinity with $\lambda > 0$. The energy of a field fluctuation with this falloff at infinity follows $E \sim (\lambda^2 + m^2)$. Thus if $m^2 < 0$, this energy can become negative for small enough λ, which means instability. This is not necessarily the case if we consider curved spacetime.

For AdS_n the metric can be written as

$$ds^2 = e^{2ky}\, dx_{n-1}^2 + dy^2 \tag{7.8}$$

with the curvature radius being

$$R = \frac{1}{k}. \tag{7.9}$$

For simplicity let us consider fluctuations of the field depending only on the y coordinate. The condition of finite energy now requires an exponential falloff $\phi \sim e^{-\lambda y}$ for $y \to \infty$ with

$$\lambda > \frac{k(n-1)}{2}. \tag{7.10}$$

As before the contribution to the energy will follow $E \sim (\lambda^2 + m^2)$ and therefore we have positive energy for tachyon fields with $m^2 = -a$ if

$$a \leq \frac{(n-1)^2}{4R^2}. \tag{7.11}$$

This bound on the tachyon mass in AdS_n is known as the Breitenlohner–Freedmann bound [42].

In the case of string theory the contribution to the energy of closed-string tachyons is as follows:

$$E = \int d^{d-1}x\, dr\, \sqrt{g} e^{-2\Phi}[g^{\mu\nu}\partial_\mu T \partial_\nu T + m^2 T^2] \tag{7.12}$$

with $m^2 = -\frac{4}{\alpha'}$. The field Φ in (7.12) is the dilaton field. We will be interested in working in flat Minkowski spacetime of dimension n. The dilaton σ-model beta-function equation

$$\frac{n-26}{6\alpha'} + (\nabla\Phi)^2 - \frac{1}{2}(\nabla^2\Phi) = 0 \tag{7.13}$$

implies a linear dilaton behaviour:

$$\Phi = y\sqrt{\frac{n-26}{6\alpha'}} \tag{7.14}$$

for some arbitrary coordinate y. Let us now consider tachyon fluctuations on this background depending only on coordinate y. Using the same argument as that for AdS_n we get the bound on the tachyon mass $m^2 = -a$:

$$a \leq \frac{n-26}{6\alpha'}. \tag{7.15}$$

Thus in order to saturate this bound for the closed-string tachyon $a = \frac{4}{\alpha'}$ we need $n = 2$. This is the celebrated $c = 1$ barrier, namely only for a spacetime dimension equal to two or smaller does the closed-string tachyon not induce any instability.

Note that from the point of view of the tachyon mass bound, linear dilaton for dimension n behaves as AdS_n with its curvature radius given by[4]

$$R^2 = \frac{3(n-1)^2\alpha'}{2(n-26)}. \tag{7.16}$$

7.4 Tachyon σ-model beta-functions

The partition function for the bosonic string in a closed tachyon background is given by

$$Z(T) = \int \mathrm{D}x \, \exp\left(\frac{-1}{2\pi\alpha'}\int \mathrm{d}^2\sigma \, \sqrt{h}(h^{\alpha\beta}\partial_\alpha x^\mu \partial_\beta x^\nu \eta_{\mu\nu} + T(x))\right). \tag{7.17}$$

The first thing we notice is that the tachyon term $\int \sqrt{h} T(x)$ is clearly non-invariant with respect to Weyl rescalings of the world-sheet metric. The strategy we will follow would be to fix $h_{\alpha\beta} = \mathrm{e}^{2\phi}\eta_{\alpha\beta}$ in (7.17) and to impose invariance with respect to changes in ϕ for the quantum corrected σ-model. We will use a background field x_0^μ with $x^\mu = x_0^\mu + \xi^\mu$ such that $\partial_\mu T(x_0) = 0$. In these conditions we get at one loop in the σ-model:

$$\frac{1}{2\pi\alpha'}\int \mathrm{d}^2\sigma \, \mathrm{e}^{2\phi}\left(T(x_0) + \frac{\alpha'}{2}\partial_\mu\partial_\nu T(x_0)\langle\xi^\mu\xi^\nu\rangle + \cdots\right) \tag{7.18}$$

$$= \frac{1}{2\pi\alpha'}\int \mathrm{d}^2\sigma \, \mathrm{e}^{2\phi}\left(T(x_0) + \frac{\alpha'}{2}\partial_\mu\partial_\nu T(x_0)\eta^{\mu\nu}\log\Lambda + \cdots\right) \tag{7.19}$$

where by $\langle\xi^\mu\xi^\nu\rangle$ we indicate the one-loop quantum fluctuations (see figure 7.1) and where Λ is the ultraviolet cutoff for the one-loop integration.

Next we need to relate the Weyl factor ϕ with the cutoff Λ. A dilatation of the world-sheet metric induces a change $\Lambda \to \lambda\Lambda$ and $\mathrm{e}^\phi \to \lambda\mathrm{e}^\phi$, thus we can identify e^ϕ with Λ. Doing this we get from (7.18):

$$\Lambda^2\left[T(x_0) + \frac{\alpha'}{2}\partial_\mu\partial_\nu T(x_0)\eta^{\mu\nu}\log\Lambda\right]. \tag{7.20}$$

Expanding (7.20) in powers of $\log\Lambda$ we get, at first order in $\log\Lambda$, that the independence of the Weyl rescalings requires

$$\beta_T \equiv 2T(x_0) + \frac{\alpha'}{2}\partial_\mu\partial_\nu T(x_0)\eta^{\mu\nu} = 0 \tag{7.21}$$

[4] For solutions to the bosonic beta-function interpolating between AdS and linear dilaton see [39,40].

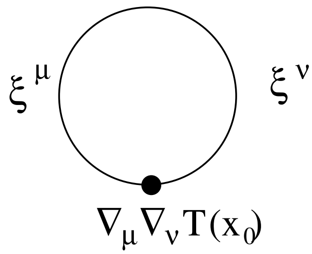

Figure 7.1. One-loop contribution to the tachyon beta-function.

which is the definition of the closed-string tachyon beta-function.

Repeating exactly the same steps for the open-string tachyon instead of (7.21) we get

$$\beta_T^o \equiv T(x_0) + \alpha' \partial_\mu \partial_\nu T(x_0) \eta^{\mu\nu} = 0. \tag{7.22}$$

If we interpret (7.21) and (7.22) as equations of motion they correspond to tachyonic spacetime fields of $(mass)^2$, respectively, $-\frac{4}{\alpha'}$ and $-\frac{1}{\alpha'}$.

What we learn from this simple exercise is that the tachyonic nature of background T introduced in (7.17) is tied to the simple fact that $\int_\Sigma \sqrt{h} T$ is not Weyl invariant. Note that although the usual dilaton term $\int_\Sigma \sqrt{h} \Phi R^{(2)}$ is not Weyl invariant it depends on ϕ only through the $(\partial\phi)^2$ terms.

7.5 Open strings and cosmological constant: the Fischler–Susskind mechanism

7.5.1 Fischler–Susskind mechanism: closed-string case

Let us start by considering one-loop divergences in the critical $D = 26$ closed bosonic string. For simplicity we will limit ourselves to amplitudes with M external tachyons. Divergences for this amplitude will arise in the limit where all the M external tachyon insertions coalesce (see figure 7.2).

The amplitude is given by

$$A(1, 2, \ldots, M) = \int \frac{\mathrm{d}^2\tau}{(\mathrm{Im}\,\tau)^2} C(\tau) F(\tau) \tag{7.23}$$

where

$$C(\tau) = \left(\frac{\mathrm{Im}\,\tau}{2}\right)^{-12} \mathrm{e}^{4\pi\,\mathrm{Im}\,\tau} |f(\mathrm{e}^{2\mathrm{i}\pi\tau})|^{-48} \tag{7.24}$$

and

$$F(\tau) = \kappa^M \,\mathrm{Im}\,\tau \int \prod^{M-1} \mathrm{d}^2\nu_r \prod_{r<s} (\chi_{rs})^{\frac{k_r k_s}{2}} \tag{7.25}$$

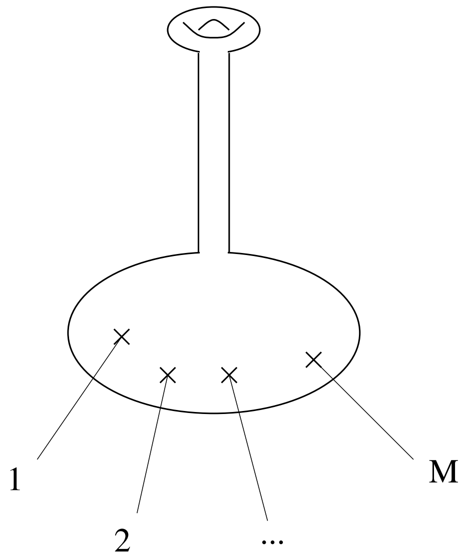

Figure 7.2. Relevant topology to describe the limit where the insertion points coalesce.

Expression (7.23) is invariant under $SL(2, Z)$ modular transformations:

$$\tau \to \frac{a\tau + b}{c\tau + d} \qquad ad - bc = 1. \tag{7.26}$$

Integration in (7.23) is reduced to the fundamental domain F. Using the conformal Killing vector on the torus we have fixed the position ν_M of one external tachyon. It is convenient to define the new variables:

$$\varepsilon\eta_r \equiv \nu_r - \nu_M \qquad r = 1, \ldots, M-2 \tag{7.27}$$

$$\eta_{M-1} \equiv \nu_{M-1} - \nu_M = \varepsilon \mathrm{e}^{\mathrm{i}\phi} \tag{7.28}$$

with ε and ϕ real variables. The Jacobian of the transformation is:

$$\prod^{M-1} \mathrm{d}^2\nu_r = \mathrm{i}\varepsilon^{2M-3}\,\mathrm{d}\varepsilon\,\mathrm{d}\phi \prod^{M-2} \mathrm{d}^2\eta_r. \tag{7.29}$$

In the limit where $\nu_{rs} = \nu_r - \nu_s \sim 0$ the Green function χ_{rs} in (7.25) behaves:

$$\chi_{rs} \sim 2\pi|\nu_{rs}|. \tag{7.30}$$

Expanding the integrand in (7.25) in this limit in powers of ε the leading divergence is:

$$\kappa^M \int_0^1 \frac{\mathrm{d}\varepsilon}{\varepsilon^3}\,\mathrm{d}\phi \prod^{M-2} \mathrm{d}^2\eta_r \prod_{1\le r\le s\le M-1} |\eta_r - \eta_s|^{\frac{k_r k_s}{2}} \int \frac{\mathrm{d}^2\tau}{(\mathrm{Im}\,\tau)} C(\tau) \tag{7.31}$$

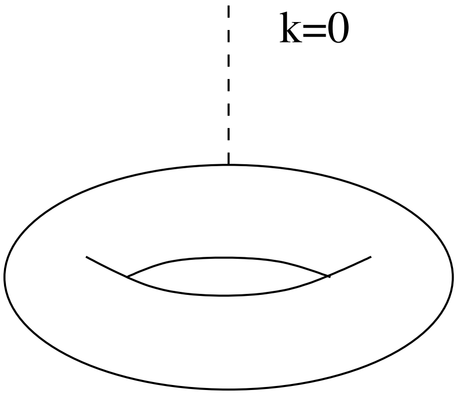

Figure 7.3. Dilaton tadpole graph.

where we have used the on-shell condition for the closed tachyon:

$$\sum_{1 \le r \le s \le M-1} k_r k_s = -4M. \tag{7.32}$$

The amplitude (7.31) corresponds to the propagation of a closed tachyon along the neck. The next subleading term in the expansion goes like $\frac{1}{\varepsilon}$ and corresponds to propagation along the neck of a massless dilaton. Thus the divergent contribution to the amplitude can be written:

$$\int_0^1 \frac{\mathrm{d}\varepsilon}{\varepsilon} A_0(k = 0, 1 \ldots, M)\kappa J \tag{7.33}$$

where A_0 is the genus-zero amplitude for M external tachyons and one dilaton at zero momentum and where κJ is proportional to the genus-one dilaton tadpole (see figure 7.3):

$$\kappa J = \kappa \int_F \frac{\mathrm{d}^2\tau}{(\mathrm{Im}\,\tau)^2} C(\tau) \tag{7.34}$$

The original idea of the Fischler–Susskind mechanism [43] consists in absorbing the genus-one divergence (7.33) into a renormalization of the world-sheet σ-model Lagrangian, namely

$$\eta_{\mu\nu}\partial x^\mu \partial x^\nu \to \eta_{\mu\nu}\left[1 + \kappa^2 J \int_0^1 \frac{\mathrm{d}\varepsilon}{\varepsilon}\right]\partial x^\mu \partial x^\nu. \tag{7.35}$$

The factor κ^2 in (7.35) appears because we want to use this counterterm on the sphere to cancel a genus-one divergence. Recall that the generic genus-one amplitudes follow κ^M while genus-zero amplitudes follow κ^{M-2}.

Obviously the renormalized Lagrangian defined in (7.35) explicitly breaks the conformal invariance. Introducing a cutoff Λ in the ε-integration the corresponding σ-model beta-function is:

$$\beta^{(1)}_{\mu\nu} = \kappa^2 J \eta_{\mu\nu} \sim \frac{\delta L_R}{\delta \log \Lambda} \tag{7.36}$$

for L_R the renormalized Lagrangian defined in (7.35). In principle, we can generalize (7.36) to curved spacetime just replacing $\eta_{\mu\nu}$ by $G_{\mu\nu}$. Once we do that we can compensate the σ-model beta-function arising from σ-model one-loop effects:

$$(\log \Lambda) R_{\mu\nu} \partial x^\mu \partial x^\nu \tag{7.37}$$

with the genus-one contribution, by imposing

$$R_{\mu\nu} = \kappa^2 J G_{\mu\nu}. \tag{7.38}$$

In summary the main message of the Fischler–Susskind mechanism is that σ-model divergences can be compensated by string loop divergences. We have shown that this is at least the case at genus one. Including the dilaton field and using the well-known relation

$$\kappa = \mathrm{e}^{\Phi} \tag{7.39}$$

we will get, instead of (7.38),

$$R_{\mu\nu} - 2\nabla_\mu \nabla_\nu \Phi = \mathrm{e}^{2\Phi} J G_{\mu\nu}. \tag{7.40}$$

7.5.2 Open-string contribution to the cosmological constant: the filling brane

This time we will consider the open-string one-loop amplitude for M external on-shell open tachyons (see figure 7.4)

In the planar case this amplitude is given by

$$\begin{aligned} A(1, 2, \ldots, M) = g^M \int_0^1 \prod^{M-1} \theta(\nu_{r+1} - \nu_r)\, \mathrm{d}\nu_r \\ \times \int_0^1 \frac{\mathrm{d}q}{q} q^{-2} [f(q^2)]^{-24} \prod_{r<s} [\Psi_{rs}]^{k_r k_s}. \end{aligned} \tag{7.41}$$

The divergences of this amplitude appear in the $q \to 0$ limit corresponding to the size of the hole of the annulus shrinking to zero. The structure of the divergences can be read from the annulus vacuum-to-vacuum amplitude:

$$Z_0^{(1)} = \int_0^1 \frac{\mathrm{d}q}{q} q^{-2} [f(q^2)]^{-24} = \int_0^1 \frac{\mathrm{d}q}{q^3} [1 + (26 - 2) q^2 + \cdots]. \tag{7.42}$$

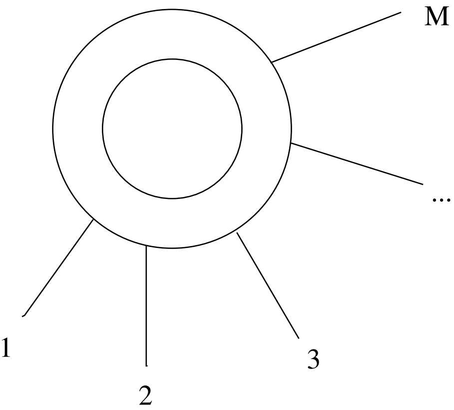

Figure 7.4. One-loop open-string amplitude.

Extending the Fischler–Susskind mechanism to (7.42) is equivalent to reproducing the coefficient of the divergences in terms of the expectation values of certain operators on the disc [44]. The divergence $26\int_0^1 \frac{\mathrm{d}q}{q}$ is easily reproduced by

$$\int_0^1 \frac{\mathrm{d}q}{q}\frac{\mathrm{e}^{\Phi}}{\alpha'}\langle \eta_{\mu\nu}\partial x^{\mu}\partial x^{\nu}\rangle_{\mathrm{disc}} \tag{7.43}$$

where we have included the dilaton factor required for matching the one-loop and disc amplitudes. The divergence $\int_0^1 \frac{\mathrm{d}q}{q^3}$ corresponds to

$$\int_0^1 \frac{\mathrm{d}q}{q^3}\mathrm{e}^{\Phi}\langle 1_d\rangle_{\mathrm{disc}}. \tag{7.44}$$

The logaritmic divergence $-2\int_0^1 \frac{\mathrm{d}q}{q}$ comes from the contribution of ghosts to the annulus partition function. The correct way to reproduce this divergence is in terms of the ghost dilaton vertex operator $D^{(\mathrm{ghost})}(k=0)$ as

$$\int_0^1 \frac{\mathrm{d}q}{q}\mathrm{e}^{\Phi}\langle D^{(\mathrm{ghost})}(k=0)\rangle_{\mathrm{disc}}. \tag{7.45}$$

In fact, the representation (7.45) of the divergence $-2\int_0^1 \frac{\mathrm{d}q}{q}$ is a direct consequence of the dilaton theorem [45]:

$$\left\langle \int D_{\mathrm{ghost}}(z,\bar{z})\Phi(p_1)\ldots\Phi(p_n)\right\rangle_{\Sigma} \sim 2g-2+n\langle\Phi(p_1)\ldots\Phi(p_n)\rangle_{\Sigma} \tag{7.46}$$

with

$$D_{\mathrm{ghost}}(z,\bar{z}) = \tfrac{1}{2}(c\partial^2 c - \bar{c}\bar{\partial}^2\bar{c}). \tag{7.47}$$

Let us concentrate on (7.43). The Fischler–Susskind counterterm needed to cancel this divergence induces a contribution to the $\beta_{\mu\nu}$ σ-model beta-function proportional to

$$\frac{e^{\Phi}}{\alpha'}\eta_{\mu\nu}. \tag{7.48}$$

In order to reproduce this term we need to add to the closed-string effective Lagrangian the open-string cosmological constant term:

$$\frac{1}{\kappa^2}\int d^{26}x\,\frac{e^{-\Phi}}{\alpha'}\sqrt{g}. \tag{7.49}$$

The reader can easily recognize in (7.49) the first term in the expansion of the $D-25$ filling brane Born–Infeld Lagrangian:

$$S_{\mathrm{BI}} = T_{25}e^{-\Phi}\int d^{26}x\,\sqrt{g+b+F} = T_{25}e^{-\Phi}\int d^{26}x\,\sqrt{g}+\cdots \tag{7.50}$$

with T_{25} the filling brane tension given by $\sim \frac{1}{\alpha'\kappa^2}$.

Thus we learn that the $D-25$ filling brane tension simply represents the open-string contribution to the cosmological constant.

Before finishing this section let us just summarize in the following table the different string contributions to the cosmological constant:

$D \neq D_{Cr}$	$\Lambda_{Cr} \sim e^{-2\Phi}\frac{D-D_{Cr}}{6\alpha'}$
Closed-string divergences	$\Lambda_{c} \sim J$
Open-string divergences	$\Lambda_{o} \sim e^{-\Phi}\frac{1}{\alpha'}$

The tachyon condensation is strongly connected with these string contributions to the cosmological constant. Generically closed tachyon condensation could change the value of Λ_{Cr} and open tachyon condensation, according to Sen's conjecture, can cancel Λ_{o}.

7.6 The effective action

7.6.1 A warming-up exercise

Let us start with the following open-string action

$$S(a) = \int_{\Sigma} d^2\sigma\,\sqrt{h}h^{\alpha\beta}\partial_\alpha x^\mu\partial_\beta x^\nu\eta_{\mu\nu} + \int_{\partial\Sigma} d\theta\, a \tag{7.51}$$

with a some constant and Σ the disc.

We will fix a world-sheet metric $h_{\alpha\beta} = e^{2\phi}\eta_{\alpha\beta}$, thus the open-string tachyon term in (7.51) is $\int_{\partial\Sigma} d\theta\, a e^{\phi}$.

The partition function $Z(a)$ is simply defined by

$$Z(a) = \int \mathrm{D}x\, \mathrm{e}^{-S(a)}. \tag{7.52}$$

If, as usual, we identify e^{ϕ} as the ultraviolet cutoff we get the beta-function for a:

$$\beta_a = -a. \tag{7.53}$$

The effective action will be defined, in this trivial case, by

$$\frac{\partial I(a)}{\partial a} = G_{TT}\beta_a \tag{7.54}$$

with β_a given in (7.53) and G_{TT} the Zamolodchikov metric[5] defined by the open-string amplitude on the disc of two open tachyon vertex operators at zero momentum:

$$G_{TT}(a) = \langle 1_d, 1_d\rangle_{\mathrm{disc}}(a) \tag{7.61}$$

with the expectation value in (7.61) computed for the action (7.51). In our case and assuming ghost decoupling it is obvious that G_{TT} is equal to $\mathrm{e}^{-a}Z(0)$. Thus using (7.54) we get the following relation for the effective action $I(a)$:

$$\frac{\partial I(a)}{\partial a} = G_{TT}\beta_a = -\mathrm{e}^{-a} a Z(0) \tag{7.62}$$

[5] For a formal derivation of (7.54) see [46]. Very briefly the proof is as follows. Let us define a family of two-dimensional field theories

$$L = L_0 + \lambda_i u^i(\xi) \tag{7.55}$$

parametrized by λ_i. The generating functional $Z(\lambda_1 \ldots \lambda_n)$ can be expanded in powers of λ. At order N we have

$$Z^N = \int \mathrm{d}^2\xi_1 \ldots \mathrm{d}^2\xi_N \,\langle u_{n_1}(\xi_1)\ldots u_{n_N}(\xi_N)\rangle \lambda_1 \ldots \lambda_n. \tag{7.56}$$

Using the OPE we get the logaritmic contribution

$$Z^n = \int \mathrm{d}^2\xi_1 \ldots \mathrm{d}^2\xi_N\, \mathrm{d}^2\xi\, f_{n_1n_2m}\frac{1}{|\xi|^2}\lambda_1 \ldots \lambda_n \langle u_m(\xi) u_{n_3}(\xi_3)\ldots u_{n_N}(\xi_N)\rangle \tag{7.57}$$

from (7.57) we can find the beta-function β_m:

$$\beta_m = \frac{\mathrm{d}\lambda_m^R}{\mathrm{d}\log\Lambda} = f_{mn_1n_2}\lambda_{n_1}\lambda_{n_2} \tag{7.58}$$

for $\lambda_m^R = \lambda_m^B + f_{mn_1n_2}\lambda_{n_1}\lambda_{n_2}\log\Lambda$ with Λ the ultraviolet cutoff in the integration (7.57). Defining now the effective action:

$$\Gamma(\lambda) = \sum \lambda^{i_1}\ldots\lambda^{i_N}\langle u_{i_1}\ldots u_{i_N}\rangle \tag{7.59}$$

we get

$$\frac{\partial\Gamma(\lambda)}{\partial\lambda_m} = \sum \lambda^{i_1}\ldots\lambda^{i_N} C_e^{i_1\ldots i_N}\langle u_m u_e\rangle = \sum \beta_e G_{me} \tag{7.60}$$

where we have used a generalized OPE and expression (7.58) for the beta-functions. In this section we will use (7.60) to define the effective action.

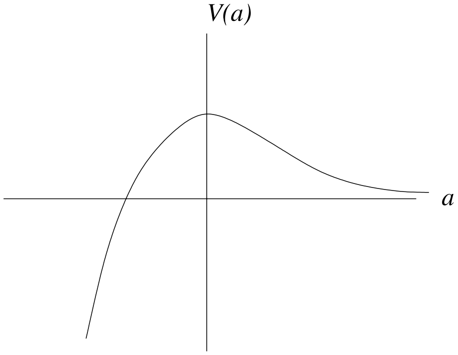

Figure 7.5. Open-string tachyon potential.

which can be trivially integrated to

$$I(a) = (1+a)Z(a) = (1+a)\mathrm{e}^{-a}Z(0). \tag{7.63}$$

This is an extremely interesting result since it defines a non-trivial potential for the tachyon constant a (figure 7.5), namely

$$V(a) = (1+a)\mathrm{e}^{-a}. \tag{7.64}$$

This potential has two extremal points at $a = \infty$ and $a = 0$.

The interpretation of the two extremal points in $V(a)$ is by no means obvious. The extremal point $a = 0$ is the standard open-string vacua with a vanishing expectation value for the open tachyon. It is a maximun reflecting the existence of open tachyons in the string spectrum. The extremal point $a = \infty$ is a bit more mysterious since apparently it describes a stable vacua (up to tunnelling processes to $a = -\infty$) of the open string in flat Minkowski spacetime and without open tachyons. What is the physical meaning of this vacua?

7.6.2 The effective action

Next we will consider, following [19–24] , a slightly more complicated action:

$$S(a, u_i) = \frac{1}{2\pi\alpha'}\int_\Sigma \mathrm{d}^2\sigma\,\sqrt{h}h^{\alpha\beta}\partial_\alpha x^\mu\partial_\beta x^\nu\eta_{\mu\nu} + \int_{\partial\Sigma}\mathrm{d}\theta\,\sqrt{h}T(x) \tag{7.65}$$

with

$$T(x) = a + \sum u_i x_i^2. \tag{7.66}$$

Identifying, as usual, the ultraviolet cutoff with the world-sheet Weyl factor we get, at one loop in the σ-model,

$$\Lambda[a + u_i\alpha' \log\Lambda] \tag{7.67}$$

from which we derive the beta-function β_a:

$$\beta_a = -a - \sum_i \alpha' u_i . \tag{7.68}$$

At this point we are interpreting x_i in (7.66) as representing quantum fluctuations, i.e. $\alpha' u_i = \frac{\partial^2 T}{\partial x_0 \partial x_0}$ and $T(x_0) = a$ for some background x_0. Thus we should replace u_i in (7.66) by $u_i \alpha'$.

In addition to β_a, we have, at tree level,

$$\Lambda[\alpha' u_i x_i^2] \tag{7.69}$$

which implies a beta-function

$$\beta_{u_i} = -u_i . \tag{7.70}$$

Using these tools we can define the effective action by

$$\mathrm{d}I = \frac{\partial I}{\partial a}\,\mathrm{d}a + \frac{\partial I}{\partial u_i}\,\mathrm{d}u_i \tag{7.71}$$

with

$$\frac{\partial I}{\partial a} = G_{aa}\beta_a + G_{au_i}\beta_{u_i} \tag{7.72}$$

$$\frac{\partial I}{\partial u_i} = G_{u_i u_j}\beta_{u_j} + G_{u_i a}\beta_a \tag{7.73}$$

where the ‘metric’ factors are defined by

$$G_{aa} = \int_0^{2\pi} \mathrm{d}\theta \,\langle 1_d, 1_d\rangle_{\mathrm{disc}} \tag{7.74}$$

$$G_{au_i} = \int_0^{2\pi} \mathrm{d}\theta \,\langle 1_d, x_i^2\rangle_{\mathrm{disc}} \tag{7.75}$$

$$G_{u_i u_j} = \int_0^{2\pi} \mathrm{d}\theta \,\langle x_i^2, x_j^2\rangle_{\mathrm{disc}} \tag{7.76}$$

In terms of the partition function

$$Z(a, u_i) = \int \mathrm{D}x \,\mathrm{e}^{-S(a,u_i)} \tag{7.77}$$

we get, from (7.71), (7.72),

$$\mathrm{d}I = \mathrm{d}\Big(\sum \alpha' u_i Z - \sum u_j \frac{\partial Z}{\partial u_j} + (1+a) Z \Big) \tag{7.78}$$

where we have used

$$G_{au_i} = \frac{\partial Z}{\partial u_i} \tag{7.79}$$

and

$$G_{u_i u_j} = \frac{\partial^2 Z}{\partial u_i \partial u_j}. \tag{7.80}$$

Integrating (7.78) we obtain the definition of the effective action:

$$I = \left(\sum \alpha' u_i - \sum u_j \frac{\partial}{\partial u_j} + (1+a) \right) Z(a, u_i). \tag{7.81}$$

In this formal derivation we have assumed the complete decoupling of ghosts. Note that the contribution $1 + a + \sum \alpha' u_i$ comes directly from the beta-function β_a defined in (7.68) while the contribution $\sum u_j \frac{\partial}{\partial u_j}$ comes from the β_{u_i} defined in (7.70). We can rewrite (7.81) in a more compact way as

$$I = \left(1 + \beta^a \frac{\partial}{\partial a} + \sum \beta^{u_i} \frac{\partial}{\partial u_i} \right) Z(a, u_i) \tag{7.82}$$

where we have used $Z(a, u_i) = \mathrm{e}^a \widetilde{Z}(u_i)$.

The next step is to compute $Z(a, u_i)$. In order to do this we need the Green function on the disc satisfying the boundary conditions

$$n_\alpha \partial^\alpha x^i + u_i x^i = 0 \tag{7.83}$$

on $\partial\Sigma$ with n_α a normal vector to the boundary. This Green function is given by

$$G^{(i)}(z, w) = -\log|z-w|^2 - \log|1 - z\bar{w}|^2 + \frac{2}{u} - 2u \sum_k \frac{1}{k(k+u)} ((z\bar{w})^k + (\bar{z}w)^k). \tag{7.84}$$

Integrating

$$\frac{\partial Z}{\partial u_i} = \int_0^{2\pi} \mathrm{d}\theta \, \langle x_i^2 \rangle_{\text{disc}} \tag{7.85}$$

and using

$$\langle x_i^2 \rangle = \lim_{\epsilon \to 0} G^i_{\mathrm{R}}(\theta, \theta + \epsilon) \tag{7.86}$$

for the renormalized Green function

$$G^i_{\mathrm{R}}(\theta, \theta) = \frac{2}{u} - 4u \sum_k \frac{1}{k(k+u)} \tag{7.87}$$

we get

$$Z(a, u_i) = \mathrm{e}^{-a} \prod_i \sqrt{\alpha' u_i} \, \Gamma(\alpha' u_i) \mathrm{e}^{\gamma \alpha' u_i} \tag{7.88}$$

for γ the Euler constant. For small u_i we can approximate (7.88) by

$$Z(a, u_i) \sim \mathrm{e}^{-a} \prod_i \frac{1}{\sqrt{\alpha' u_i}} \qquad u_i \to 0. \tag{7.89}$$

In this limit we get from (7.82):

$$I(a, u_i) \sim (1+a)\mathrm{e}^{-a} \prod_i \frac{1}{\sqrt{\alpha' u_i}} + \alpha' \left(\sum u_i \right) \mathrm{e}^{-a} \prod_i \frac{1}{\sqrt{\alpha' u_i}} + \cdots . \tag{7.90}$$

We can now compare the first term with

$$T_{25} \int \mathrm{d}^{26}x\, (1+T)\mathrm{e}^{-T} \tag{7.91}$$

for $T = a + \sum u_i x_i^2$, obtaining the well-known result for the filling brane tension:

$$T_{25} = \frac{1}{(2\pi\alpha')^{13}}. \tag{7.92}$$

Then the terms in (7.90) correspond to the kinetic term for the open tachyon:

$$T_{25} \int \mathrm{d}^{26}x\, \mathrm{e}^{-T} \partial T \partial T. \tag{7.93}$$

In order to define a potential we can change variables:

$$T \to \Phi = 2\mathrm{e}^{-\frac{T}{2}} \tag{7.94}$$

In these new variables the tachyon Lagrangian becomes

$$S = T_{25} \int \mathrm{d}^{26}x\, [\alpha' \partial\Phi\partial\Phi + V(\Phi)] \tag{7.95}$$

with (see figure 7.6)

$$V(\Phi) = \frac{\Phi^2}{4} \left(1 - \log \frac{\Phi^2}{4} \right). \tag{7.96}$$

The extremal corresponding to $T = \infty$ is $\Phi = 0$. The effective mass of the tachyon around this extremal is

$$m^2 = \left. \frac{\partial^2 V(\Phi)}{\partial\Phi\partial\Phi} \right|_{\Phi=0} = \infty. \tag{7.97}$$

The extremal $T = 0$, i.e. $\Phi = 2$, is a maximum reproducing the standard open tachyon mass:

$$m^2 = \left. \frac{\partial^2 V(\Phi)}{\partial\Phi\partial\Phi} \right|_{\Phi=2} = -\frac{1}{\alpha'}. \tag{7.98}$$

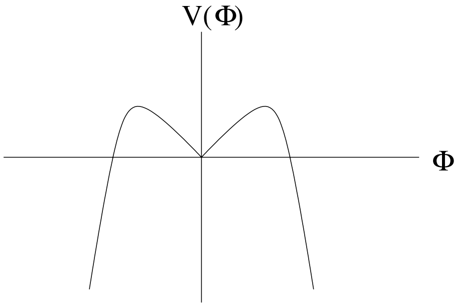

Figure 7.6. Open-string tachyon potential $V(\Phi)$.

As we can see for equation (7.97) open tachyon condensation at $T = \infty$ induces an infinite mass for the open tachyon. Using the string mass formula (7.1) we can interpret this as an effective normal ordering constant $a = -\infty$. If we do this the dominant contribution to the BRST charge (7.5) is just the cohomologically trivial BRST charge $Q = c_0$.

This heuristic argument indicates, in agreement with Sen's conjecture that no open-string degrees of freedom survive once the tachyon condenses to $T = \infty$. In summary we can interpret the vacuum defined by the $T = \infty$ condensate as the closed-string vacua. The closed-string tachyon can be interpreted as being associated with the quantum instability due to tunnelling processes from $\Phi = 0$ to $\Phi = \infty$.

7.6.3 Non-critical dimension and tachyon condensation

The spacetime Lagrangian for the open tachyon is given by

$$S = T_{25} \int \mathrm{d}^{26}x \, \mathrm{e}^{-T} [\alpha' \partial T \partial T + (1 + T)]. \tag{7.99}$$

The corresponding equation of motion is

$$2\alpha' \partial^\mu \partial_\mu T - \alpha' \partial^\mu T \partial_\mu T + T = 0. \tag{7.100}$$

A soliton solution for equation (7.100) is given by:

$$T(x) = a + \sum u_i x_i^2 \tag{7.101}$$

with $u_i = \frac{1}{4\alpha'}$ or $u_i = 0$ and $a = -n$ for n the number of non-vanishing u_i's. In terms of the field Φ defined in (7.94) the profile of the solution looks like the one depicted in figure 7.7.

This can be interpreted in a first approximation as the $D - (25 - n)$ soliton brane.

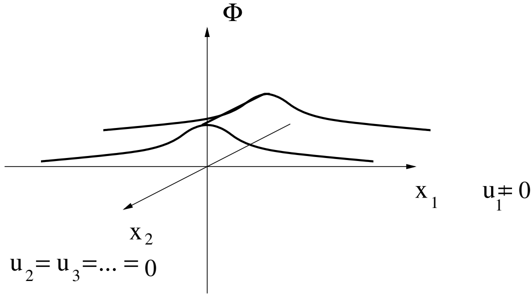

Figure 7.7. Soliton shape.

In principle, we can try to play the same game but include the effect of a non-trivial dilaton. The simplest example will be, of course, to work with non-critical dimension n and a linear dilaton background

$$\Phi = qy \tag{7.102}$$

with $q = \sqrt{\frac{n-26}{6\alpha'}}$. Inspired by the Liouville picture of non-critical strings we take the linear dilaton to depend only on one coordinate y. The Lagrangian including the effect of the dilaton would, most probably, be:

$$S = T_{25} \int \mathrm{d}^{26}x \, \mathrm{e}^{-\Phi} \mathrm{e}^{-T} [\alpha' \partial T \partial T + (1 + T)]. \tag{7.103}$$

The equation of motion becomes

$$2\alpha' \partial^{\mu} \partial_{\mu} T - \alpha' \partial^{\mu} T \partial_{\mu} T - \alpha' \partial^{\mu} \Phi \partial_{\mu} T + T = 0. \tag{7.104}$$

As a solution we can try

$$T(x) = a + \sum u_i x_i^2 \qquad a = -m \qquad u_i = \frac{1}{4\alpha'} \qquad i = 1, \ldots, m \tag{7.105}$$

with $u_y = 0$. This soliton defines a $D - (n - m - 1)$-brane that extends along the 'Liouville' direction. Note that we have no soliton solutions for $u_y \neq 0$ which seems to imply that tachyon condensation does not take place in the Liouville direction. This leads us to suggest the following conjecture: In non-critical open strings, open tachyon condensation cannot take place in the Liouville direction.

A trivial corollary of the previous conjecture is that in a spacetime dimension equal to two tachyon condensation does not take place, which is consistent with the fact that tachyons in $D = 2$ with the linear dilaton turned on are not real tachyons.

7.7 D-branes, tachyon condensation and K-theory

7.7.1 Extended objects and topological stability

Let us start by considering a gauge theory with a Higgs field Φ:

$$L = L_0(A^\mu, \Phi) + V(\Phi) \tag{7.106}$$

for some Higgs potential $V(\Phi)$. A necessary condition for the existence of topologically stable extended objects of space codimension p is the non-triviality of the homotopy group

$$\Pi_{p-1}(V) \tag{7.107}$$

for V the manifold of classical vacua of Lagrangian (7.106).

In fact, for an extended object of codimension p the condition of finite density of energy implies that at the infinity region in the transversal directions—whose topology is S^{p-1}—the field configuration must belong to the vacuum manifold V. Hence we associate with each configuration of finite density of energy a map

$$\Psi : S^{p-1} \to V \tag{7.108}$$

whose topological clasification is defined by the homotopy group (7.107).

The simplest example of a vacuum manifold corresponding to the spontaneous breaking of symmetry $G \to H$ is the homogeneous space

$$V = G/H. \tag{7.109}$$

So the 't Hooft–Polyakov monopole, for instance, is defined for $G = SU(2)$ and $H = U(1)$ by the topological condition $\Pi_2(G/H) = Z$ which coincides with its magnetic charge.

7.7.2 A gauge theory analogue for D-branes in type II strings

We know that in type II strings we have extended objects which are R–R charged and stable, namely the D-branes. For type IIA we have D_p-branes with p even and for type IIB D_p-branes with p odd. Since we are working in critical ten-dimensional spacetime the space codimension of those D_p-branes is odd $2k + 1$ for type IIA and even $2k$ for type IIB.

We will now consider the following formal problem. Obtain two gauge Higgs Lagrangians $L^{\mathrm{IIA(IIB)}}(A_\mu, \Phi)$ such that a one-to-one map can be established between type II D-branes and topological stable extended objects for those Lagrangians in the sense defined in previous section. We will denote this formal gauge theory the *gauge theory analogue* of type II strings.

Of course the hint for answering this question is Sen's tachyon condensation conjecture for type II strings. We will first present this construction in the case of type IIB strings.

7.7.2.1 Sen's conjecture for type IIB strings

In type IIB strings we have well-defined D_9 filling branes. Since they are charged under the R–R sector we can define the corresponding $D_{\bar{9}}$-antibranes. As is well known the low-energy physics on the world-volume of a set of N D_9-branes is a $U(N)$ gauge theory without open tachyons. In fact the open tachyon is projected out by the standard GSO projection

$$(-1)^F = +1 \tag{7.110}$$

with F the world-sheet fermion number operator. The situation changes when we consider N D_9-branes and N $D_{\bar{9}}$-antibranes. In this case the theory on the world-volume is $U(N) \times U(N)$ and not $U(2N)$ due to the fact that the GSO projection on open-string states with end points at a D_9-brane and a $D_{\bar{9}}$-antibrane is the opposite, namely

$$(-1)^F = -1 \tag{7.111}$$

This projection eliminates the massless gauge vector bosons that will enhance the $U(N) \times U(N)$ gauge symmetry to $U(2N)$ from the spectrum. In addition, projection (7.111) does not kill the tachyon in the $(9, \bar{9})$ and $(\bar{9}, 9)$ open-string sectors. Thus, the gauge theory associated with the configuration of N D_9-branes and N $D_{\bar{9}}$-antibranes is a $U(N) \times U(N)$ gauge theory with a Higgs field, the open tachyon, in the bifundamental $(N, \bar{N})$ representation.

This gauge theory will be a natural starting point for defining the gauge analogue model in the case of type IIB strings.

Of course in order to obtain a rigorous criterion for the topological stability of extended objects in this gauge theory we need to know the potential for the open tachyon. This potential is something that at this point we do not know how to calculate it in a rigorous way. However, we can assume that a tachyon condensation is generated with a vacuum expectation value

$$\langle T \rangle = T_0 \tag{7.112}$$

with T_0 diagonal and with equal eigenvalues. If this condensate takes place then the vacuum manifold is simply

$$V = \frac{U(N) \times U(N)}{U_D(N)} \sim U(N). \tag{7.113}$$

Thus the condition for topological stability for extended objects of space codimension $2k$ will be

$$\Pi_{2k-1}(U(N)) \neq 0 \tag{7.114}$$

which, by the Bott periodicity theorem in which

$$\Pi_j(U(k)) = \begin{cases} Z & j \text{ odd } j < 2k \\ 0 & j \text{ even } j < 2k \end{cases} \tag{7.115}$$

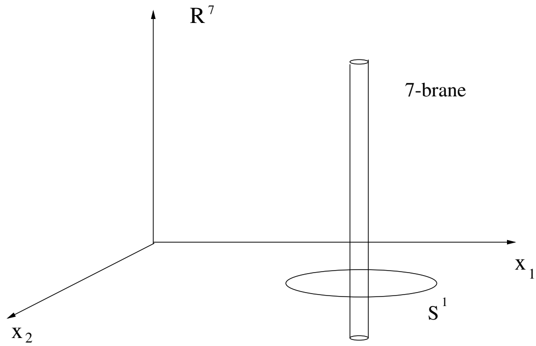

Figure 7.8. Topology of the D_7-brane.

we know is the case for big enough N.

The simplest example will be to take $k = 1$ corresponding to the extended object of the type of a D_7-brane. The condition of finite energy density defines a map from S^1 into $U(N)$. For just one pair of the D_9–$D_{\bar{9}}$ configuration we get $\Pi_1(U(1)) = Z$, with this winding number representing the 'magnetic charge' of the D_7-brane that looks topologically like a vortex line (see figure 7.8).

If we go to the following brane, namely the D_5-brane, we have $k = 2$ and we need a non-vanishing homotopy group $\Pi_3(U(N))$. The minimun N for which this is possible according to (7.115) is $N = 2$, i.e. two pairs of D_9–$D_{\bar{9}}$-branes. We can understand what is happening in two steps. First of all, we obtain a configuration of two D_7-branes and from this the D_5-brane.

For $k = 3$ we need the non-vanishing homotopy group $\Pi_5(U(N))$. The natural N we should choose is dictated by the step construction, namely $N = 4$. In general, for codimension $2k$ we will consider a gauge group $U(2^{k-1})$.

7.7.3 K-theory version of Sen's conjecture

The configuration of D_q–$D_{\bar{q}}$ branes naturally defines a couple of $U(N)$ vector bundles (E, F) on the spacetime. Sen's main idea of tachyon condensation is that a configuration characterized by a couple of vector bundles (E, E) with a topologically trivial tachyon field configuration decays into the closed-string vacua for type IIB string theory. This is exactly the same type of phenomena we have discussed in section 7.1 for the bosonic string. This phenomenon naturally leads us to consider, as far as we are concerned with D-brane charges, instead of the couple of bundles (E, F), the equivalence class defined by [32]

$$(E, F) \sim (E \oplus G, F \oplus G) \tag{7.116}$$

for $\oplus$, the direct sum of bundles. This is precisely the definition of the K-group of vector bundles on the spacetime X, $K(X)$. Let us here recall that the space

$A = V_{\mathrm{ec}}(X)$ of vector bundles on X is a semigroup with respect to the operation of direct sum. The way to associate with A a group $K(A)$ is as the quotient space in $A \times A$ defined by the equivalence relation

$$(m, n) \sim (m', n') \tag{7.117}$$

if $\exists p$ such that $m + n' + p = n + m' + p$ which is precisely what we are doing in the definition (7.116).

A different but equivalent way to define $K(A)$ for $A = V_{\mathrm{ec}}(X)$ is as the set of equivalence classes in $V_{\mathrm{ec}}(X)$ defined by the equivalence relation:

$$E \sim F \qquad \text{if } \exists G : E \oplus G = F \oplus G \tag{7.118}$$

where '=' means isomorphism.

It is convenient for our purposes to work with the reduced K-group $\tilde{K}(X)$ which is defined by

$$\mathrm{Ker}[K(X) \to K(p)] \tag{7.119}$$

for p a point in X. Note that $K(p)$ is just the group of integer numbers Z. This is the group naturally associated with the semigroup $V_{\mathrm{ec}}(p) = N$ where N here parametrizes the different dimensions of the vector bundles in $V_{\mathrm{ec}}(p)$.

In order to characterize type IIB D_p-branes in terms of K-theory we will need to consider the group $K(X, Y)$. We will consider X a compact space with Y also compact and contained in X.

In order to define $K(X, Y)$ we will use triplets (E, F, α) where E and F are vector bundles on X and where α is an isomorphism:

$$\alpha : E|_Y \to F|_Y \tag{7.120}$$

of the vector bundles E and F reduced to the subspace Y [37].

The definition of $K(X, Y)$ requires us to define *elementary triplets*. An elementary triplet is given by (E, F, α) with $E = F$ and α homotopic to the identity in the space of automorphisms of $E|_Y$. Once we have defined the elementary triplets, the equivalence relation defining $K(X, Y)$ is

$$\sigma = (E, F, \alpha) \qquad \sigma' = (E', F', \alpha') \tag{7.121}$$

$\sigma \sim \sigma'$ iff $\exists$ elementary triplets τ and τ' such that

$$\sigma + \tau = \sigma' + \tau' \tag{7.122}$$

where

$$\sigma + \tau = (E \oplus G, F \oplus G, \alpha \oplus 1_d). \tag{7.123}$$

Once we have defined $K(X, Y)$ we can try to put the topological characterization of a D_p-brane of space codimension $2k$ in this language. Namely we will take Y as the 'boundary' region in transversal directions, i.e. S^{2k-1}.

As space X we will take the ball B^{2k}. The tachyon field transforming in the bifundamental representation will define on S^{2k-1} an isomorphism between the two vector bundles E, F defined by the starting configurations of D_9–$D_{\bar{9}}$-branes. Finally the homotopy class of this map will define the charge of the D_q-brane of space codimension $2k$. The K-group we define in this way is

$$K(B^{2k}, S^{2k-1}). \tag{7.124}$$

Now we can use the well-known relation:

$$K(B^{2k}, S^{2k-1}) = \tilde{K}(B^{2k}/S^{2k-1}) \tag{7.125}$$

where X/Y is defined by contracting Y to a point.

It is easy to see that

$$B^{2k}/S^{2k-1} \sim S^{2k}. \tag{7.126}$$

Thus we can associate with type IIB D_p-branes of space codimension $2k$ elements in $\tilde{K}(S^{2k})$.

In order to define the tachyon field in this case we will specify the isomorphism α. For codimension $2k$ let us consider the $2^{k-1} \times 2^{k-1}$ gamma matrices Γ_i $(i = 1 \ldots 2k)$. Let v be a vector in C^{2k}. The isomorphism α is defined by

$$\alpha(x, v) = (x, x_i \Gamma^i(v)) \tag{7.127}$$

for $x \in S^{2k-1}$. The tachyon field is defined by

$$T(x)|_{x \in S^{2k-1}} = x_i \Gamma^i. \tag{7.128}$$

7.7.4 Type IIA strings

Next we will define a gauge analogue for type IIA strings. The gauge-Higgs Lagrangian will be defined in terms of a configuration of D_9-branes for type IIA D_9-branes are not BPS and therefore they are unstable. The manifestation of this instability is the existence of an open tachyon field T transforming in the adjoint representation. The gauge group for a configuration of N D_9-branes is $U(N)$. Note that in type IIA we cannot use $D_{\bar{9}}$-antibranes since type IIA D_9-branes are not R–R-charged.

We could now follow the same steps as for the type IIB case, namely look for a tachyon potential and compute the even homotopy groups of the corresponding vacuum manifold. Instead of doing this we will approach the problem from a different point of view, interpreting the type IIA D-branes as topological defects associated with the gauge-fixing topology. In order to describe this approach we need first to review some known facts about gauge-fixing topology for non-Abelian gauge theories.

7.7.4.1 't Hooft's Abelian projection

An important issue in the quantization of non-Abelian gauge theories is to fix the gauge. By an unitary gauge we mean a procedure to parametrize the space of gauge 'orbits', i.e. the space of physical configurations

$$R/G \tag{7.129}$$

for R the total space of field configurations, in terms of physical degrees of freedom whereby we mean those that contribute to the unitary S-matrix. This, in particular, means a ghost-free gauge fixing.

In [41] 't Hooft suggested a way to fix the non-Abelian gauge invariance in a unitary way. This type of gauge fixing, known as 'Abelian projection', reduces the physical degrees of freedom to a set of $U(1)$ photons and electrically charged vector bosons.

In addition to these particles there is an extra set of dynamical degrees of freedom we need to include in order to have a complete description of the non-Abelian gauge theory. These extra degrees of freedom are magnetic monopoles that appear as a consequence of the topology of the gauge fixing.

More precisely, let X be a field transforming in the adjoint representation

$$X \to gXg^{-1}. \tag{7.130}$$

The field X can be a functional $X(A)$ of the gauge field A or some extra field in the theory. The way to fix the gauge is to impose X to be diagonal:

$$X = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_N \end{pmatrix}. \tag{7.131}$$

The residual gauge invariance for a $U(N)$ gauge theory is $U(1)^N$, i.e. gauge transformations of the type

$$g = \begin{pmatrix} e^{i\alpha_1} & & \\ & \ddots & \\ & & e^{i\alpha_N} \end{pmatrix}. \tag{7.132}$$

The degrees of freedom of this gauge are:

- N $U(1)$ photons,
- $\frac{1}{2}N(N-1)$ charged vector bosons and
- N scalars fields λ_i.

Gauge-fixing singularities will appear whenever two eigenvalues coincide:

$$\lambda_i = \lambda_{i+1}. \tag{7.133}$$

Note that we can fix the gauge imposing X to be diagonal and that $\lambda_i > \lambda_{i+1} > \lambda_{i+2} > \cdots$. What is the physical meaning of these gauge-fixing singularities?

First of all, it is easy to see that generically these gauge-fixing singularities have codimension 3 in space. In particular, this means that if we are working in four-dimensional spacetime they behave as pointlike particles.

Second, if we consider the field X in a close neighbourhood of the singular point before gauge fixing:

$$X = \left(\begin{array}{c|cc|c} D_1 & \multicolumn{2}{c|}{0} & 0 \\ \hline 0 & \lambda + \epsilon_3 & \epsilon_1 - i\epsilon_2 & 0 \\ & \epsilon_1 + i\epsilon_2 & \lambda - \epsilon_3 & \\ \hline 0 & \multicolumn{2}{c|}{0} & D_2 \end{array}\right) \tag{7.134}$$

we can write the small two-by-two matrix in (7.134) as:

$$X = \lambda 1_d + \epsilon_i \sigma_i \tag{7.135}$$

for σ_i the Pauli matrices. The field $\epsilon(x)$ is equal to zero at the singular point and in a close neighbourhood this can be written as

$$\epsilon(x) = \sum_{i=1}^{3} x_i \sigma_i. \tag{7.136}$$

We can easily relate this field to a magnetic monopole. In fact let us consider S^2 in R^3 and let us define the field on S^2:

$$X(x)|_{x \in S^2} = \sum_{i=1}^{3} x_i \sigma_i. \tag{7.137}$$

Clearly $X^2(x)|_{x \in S^2} = 1_d$, thus we can define the projector:

$$\Pi_{\pm} = \tfrac{1}{2}(1 \pm X(x)). \tag{7.138}$$

The trivial bundle $S^2 \times C^2$ decomposes into

$$S^2 \times C^2 = E_+ \oplus E_- \tag{7.139}$$

where the line bundles $E_{\pm}$ are defined by the action of the projection $\Pi_{\pm}$ on C^2. The principal bundle associated with E_+, E_- defines the magnetic monopoles.

In summary the gauge-fixing singularities of gauge (7.131) corresponding to two equal eigenvalues should be interpreted as pointlike magnetically charged particles. It is important to stress that the existence of these magnetic monopoles is completly independent of being in a Higgs or confinement phase.

7.7.4.2 The D_6-brane

Here we will repeat the discussion from section 7.7.4.1 but for the $U(N)$ gauge theory defined by a configuration of D_9 unstable filling branes. We will use the open tachyon field transforming in the adjoint representation to fix the gauge. By imposing T to be diagonal we reduce the theory to pure Abelian degrees of freedom in addition to magnetically charged objects of space codimension 3 that very likely can be identified with D_6-branes.

Using expression (7.137) and replacing the X field by the open tachyon we find that in the close neighbourhood of a codimension 3 singular region the tachyon field is represented by

$$T(x)|_{x\in S^2} = \sum_{i=1}^{3} x_i\sigma_i \tag{7.140}$$

which is precisely the representation of the tachyon field around a D_6-brane suggested in [33].

7.7.4.3 K-theory description

The data we can naturally associate with a configuration of type IIA D_9-branes is a couple (E, T) with E a vector bundle and T the open tachyon field. We will translate these data into more mathematical language using the higher K-group $K^{-1}(X)$ [32,33].

In order to define $K^{-1}(X)$ we will start with couple (E, α) with E a vector bundle on X and α an automorphism of E. As we did in the definition of $K(X, Y)$ we define elementary pairs (E, α) if α is homotopic to the identity within automorphisms of E. Using elementary pairs $(E, \alpha) = \tau$ we define the equivalence relation

$$\sigma \sim \sigma' \tag{7.141}$$

iff $\exists\tau$, τ' elementary such that

$$\sigma \oplus \tau = \sigma' \oplus \tau'. \tag{7.142}$$

We can now define $K^{-1}(X, Y)$ as pairs $(E, \alpha) \in K^{-1}(X)$ such that $\alpha|_Y = 1_d$.

As before we will use the tachyon field T to define the automorphism α. In codimension 3 in the previous section we obtained

$$T(x)|_{x\in S^2} = \sum_{i=1}^{3} x_i\sigma_i. \tag{7.143}$$

Clearly $T^2(x)|_{x\in S^2} = 1_d$ and, therefore, if we define

$$\alpha = \mathrm{e}^{\mathrm{i}T} \tag{7.144}$$

and we identify $Y = S^2$ we get the condition

$$\alpha|_Y = 1_d \tag{7.145}$$

used in the definition of $K^{-1}(X, Y)$. Thus we associate the D_p-branes of codimension $2k + 1$ with elements in

$$K^{-1}(B^{2k+1}, S^{2k}) \tag{7.146}$$

Using again the relation

$$K^{-1}(B^{2k+1}, S^{2k}) = \tilde{K}^{-1}(B^{2k+1}/S^{2k}). \tag{7.147}$$

and

$$\tilde{K}^{-1}(X) = \tilde{K}(SX) \tag{7.148}$$

for SX the reduced suspension of X (in particular $SS^n = S^{n+1}$) we conclude that type IIA D_p-branes are associated with

$$\tilde{K}^{-1}(S^{2k+1}) = \tilde{K}(S^{2k+2}). \tag{7.149}$$

The reader can wonder at this point in what sense working with K-theory is relevant for this analysis. The simplest answer comes from remembering the group structure of $K^{-1}(X)$.

The group structure of in $K^{-1}(X)$ is associated with the definition of the inverse. Namely the inverse of (E, α) is (E, α^{-1}). The reason is that

$$(E, \alpha) \oplus (E, \alpha^{-1}) = (E \oplus E, \alpha \oplus \alpha^{-1}) \tag{7.150}$$

where

$$\alpha \oplus \alpha^{-1} = \begin{pmatrix} \alpha & 0 \\ 0 & \alpha^{-1} \end{pmatrix}. \tag{7.151}$$

Now there is a homotopy transforming matrix (7.151) into the identity

$$\begin{pmatrix} \alpha & 0 \\ 0 & \alpha^{-1} \end{pmatrix}(t) = \begin{pmatrix} \alpha & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \alpha^{-1} \end{pmatrix} \begin{pmatrix} \cos t & \sin t \\ \sin t & \cos t \end{pmatrix} \tag{7.152}$$

such that

$$\begin{pmatrix} \alpha & 0 \\ 0 & \alpha^{-1} \end{pmatrix}(t = 1) = \begin{pmatrix} \alpha\alpha^{-1} & 0 \\ 0 & 1 \end{pmatrix} = 1_d. \tag{7.153}$$

What this homotopy means is again Sen's tachyon condensation conjecture. In fact if we associate a D_6–$D_{\bar{6}}$ brane configuration with a matrix T with two pairs of eigenvalues $(\lambda_i = \lambda_{i+1})$ and $(\lambda_j = \lambda_{j+1})$ equal. This configuration is—because of homotopy (7.152)—topologically equivalent to the vacuum. We see once more how Sen's condensation is at the core of the K-theory promotion of $V_{\mathrm{ec}}(X)$ from a semigroup into a group.

7.8 Some final comments on gauge theories

The data associated with a gauge theory and a unitary gauge fixing of the type used in the Abelian projection can be summarized in the same type of couples used to define the higher K-group $K^{-1}(X)$, namely a $U(N)$ vector bundle E and an automorphism α. In this sense, the gauge-fixing topology is translated into the homotopy class of α within the automorphisms of E. In standard four-dimensional gauge theories the gauge-fixing topology is described in terms of magnetic monopoles and antimonopoles. In principle, we have different types of magnetic monopoles charged with respect to the different $U(1)$'s in the Cartan subalgebra. The group theory meaning of $K^{-1}(X)$ is reproduced, at the gauge theory level, by the homotopy (7.152) that is telling us that monopole–antimonopole pairs, although charged with respect to different $U(1)$'s in the Cartan subalgebra, annihilate into the vacuum, very much in the same way as, by Sen's tachyon condensation, a brane–antibrane pair decay into the vacuum.

In what we have denoted the gauge theory analogue of type II strings, namely a gauge-Higgs Lagrangian with topologically stable extended objects in one-to-one correspondence with type II stable D_p-branes, apparently one important dynamical aspect of D-filling brane configurations is absent. In fact, in the case of unstable filling branes, the decay into the vacuum comes together with the process of cancellation of the filling brane tension and thus with the 'confinement' of 'electric' open-string degrees of freedom. The resulting state is a closed-string vacua with stable D_p-branes that are sources of R–R fields which are part of the closed-string spectrum. The dynamics we lack in the gauge theory analogue is, on the one hand, the equivalent of the confinement of open-string degrees of freedom[6] and, on the other, the R–R closed string interpretation of the dual field created by the D_p-brane topological defects. Very likely the gauge theory interpretation of these two phenomena can shed some light on the quark confinement problem.

Acknowledgments

CG thanks the organizers of the Fourth SIGRAV School at Como for giving him the opportunity to present these lectures. We also thank E Alvarez for useful discussions. This work is supported by the Spanish grant AEN2000-1584. The work of PR is supported by a FPU UAM grant.

[6] i.e. the concept of a filling brane tension which is intimately related to the open–closed relation in string theory, namely to soft dilaton absorption to the vacuum.

References

[1] Sen A 1999 Descent relations among bosonic D-branes *Int. J. Mod. Phys.* A **14** 4061–78

[2] Sen A 1998 Stable non BPS bound states of BPS D-branes *J. High Energy Phys.* **9808** 010

[3] Sen A 1998 Tachyon condensation on the brane–antibrane system *J. High Energy Phys.* **9808** 012

[4] Sen A 1998 SO(32) spinors of type I and other solitons in brane–antibrane pair *J. High Energy Phys.* **9809** 023

[5] Sen A 1998 BPS D-branes on non-supersymmetric cycles *J. High Energy Phys.* **9812** 021

[6] Sen A 1999 Universality of the tachyon potential *J. High Energy Phys.* **9912** 027

[7] Sen A and Zwiebach B 2000 Tachyon condensation in string field theory *J. High Energy Phys.* **0003** 002

[8] Taylor W 2000 D-brane effective field theory from string field theory *Nucl. Phys.* B **585** 171–92

[9] Kostelecky V A and Potting R 1996 Expectation values, Lorentz invariance and CPT in the open bosonic string *Phys. Lett.* B **381** 89–96

Kostelecky V A and Samuel S 1988 The static tachyon potential in the open bosonic string theory *Phys. Lett.* B **207** 169

[10] Moeller N and Taylor W 2000 Level truncation and the tachyon in open bosonic string field theory *Nucl. Phys.* B **583** 105–44

[11] Harvey J A and Kraus P 2000 D-branes as unstable lumps in bosonic open string field theory *J. High Energy Phys.* **0004** 012

[12] de Mello Koch R, Jevicki A, Mihailescu M and Tatar R 2000 Lumps and p-branes in open string field theory *Phys. Lett.* B **482** 249–54

[13] Berkovits N, Sen A and Zwiebach B 2000 Tachyon condensation in superstring field theory *Nucl. Phys.* B **587** 147–78

[14] Berkovits N 2000 The tachyon potential in open Neveu-Schwarz string field theory *J. High Energy Phys.* **0004** 022

[15] De Smet P-J and Raeymaekers J 2000 Level four approximation to the tachyon potential in superstring field theory *J. High Energy Phys.* **0005** 051

[16] Iqbal A and Naqvi A 2000 Tachyon condensation on a non-BPS D-brane *Preprint* hep-th/0004015

[17] Harvey J A and Horava P 2000 D-sphalerons and the topology of string configuration space *J. High Energy Phys.* **0003** 021

[18] Bardakci K 1974 Dual models and spontaneous symmetry breaking *Nucl. Phys.* B **68** 331

Bardakci K and Halpern M B 1974 Explicit spontaneous breakdown in a dual model *Phys. Rev.* D **10** 4230

Bardakci K and Halpern M B 1975 Explicit spontaneous breakdown in a dual model II: N point functions *Nucl. Phys.* B **96** 285

Bardakci K 1978 Spontaneous symmetry breakdown in the standard dual string model *Nucl. Phys.* B **133** 297

[19] Witten E 1992 On background independent open string field theory *Phys. Rev.* D **46** 5467–73

[20] Witten E 1993 Some computations on background independent off-shell string theory *Phys. Rev.* D **47** 3405–10

[21] Shatashvili S L 1993 Comment on the background independent open string field theory *Phys. Lett.* B **311** 83–6

[22] Shatashvili S L 1993 On the problems with background independence in string theory *Preprint* hep-th/9311177

[23] Gerasimov A A and Shatashvili S L 2000 On exact tachyon potential in open string field theory *J. High Energy Phys.* **0010** 034

[24] Kutasov D, Marino M and Moore G 2000 Some exact results on tachyon condensation in string field theory *J. High Energy Phys.* **0010** 045

[25] Tseytlin A A 1991 On the tachyonic terms in the string effective action *Phys. Lett.* B **264** 311–18

[26] Tseytlin A A 2001 Sigma model approach to string theory effective actions with tachyons *Math. Phys.* **42** 2854

[27] Polyakov A M 1999 The wall of the cave *Int. J. Mod. Phys.* A **14** 645–58

[28] Rastelli L, Sen A and Zwiebach B 2000 String field theory around the tachyonic vacuum *Preprint* hep-th/0012251

[29] Sen A 2001 Fundamental strings in open string theory at the tachyonic vacuum *Math. Phys.* **42** 2844

[30] Kleban M, Lawrence A E and Shenker S 2001 Closed strings from nothing *Phys. Rev.* D **64** 066002

[31] Yi P 1999 Membranes from five-branes and fundamental strings from Dp-branes *Nucl. Phys.* B **550** 214–24

[32] Witten E, 1998 D-branes and K-theory *J. High Energy Phys.* **9812** 019

[33] Horava P 1999 Type IIA D-branes, K-theory, and matrix theory *Adv. Theor. Math. Phys.* **2** 1373–404

[34] Minasian R and Moore G 1997 K-theory and Ramond–Ramond charge *J. High Energy Phys.* **9711** 002

[35] Moore G and Witten E 2000 Selfduality, Ramond–Ramond fields, and K-theory *J. High Energy Phys.* **0005** 032

[36] Witten E 2001 Overview of K-theory applied to strings *Int. J. Mod. Phys.* A **16** 693–706

[37] Karoubi M 1978 *K-Theory: An Introduction* (Berlin: Springer)

[38] Gomez C 2002 A comment on K-theory meaning of the topology of gauge fixing *J. High Energy Phys.* **0206** 020

[39] Alvarez E, Gomez C and Hernandez L 2001 Non-critical Poincaré invariant bosonic string backgrounds *Nucl. Phys.* B **600** 185–96

[40] Alvarez E, Gomez C, Hernandez L and Resco P 2001 Non-critical non-singular bosonic strings, linear dilatons and holography *Nucl. Phys.* B **603** 286–96

[41] 't Hooft G 1981 Topology of the gauge condition and new confinement phases in non-Abelian gauge theories *Nucl. Phys.* B **190** 455

[42] Breitenlohner P and Freedman D Z 1982 Positive energy in anti-de Sitter backgrounds and gauged extended supergravity *Phys. Lett.* B **115** 197

Breitenlohner P and Freedman D Z 1982 Stability in gauged extended supergravity *Ann. Phys.* **144** 249

[43] Fischler W and Susskind L 1986 Dilaton tadpoles, string condensates and scale invariance *Phys. Lett.* B **171** 383

Fischler W and Susskind L 1986 Dilaton tadpoles, string condensates and scale invariance. 2 *Phys. Lett.* B **173** 262

[44] Callan C G, Lovelace C, Nappi C K and Yost S A 1987 String loop corrections to beta functions *Nucl. Phys.* B **288** 525

[45] Bergman O and Zwiebach B 1995 The dilaton theorem and closed string backgrounds *Nucl. Phys.* B **441** 76–118

[46] Polyakov A M 1987 *Gauge Fields and Strings* (Chur: Harwood)

This page intentionally left blank

PART 3

MATHEMATICAL DEVELOPMENTS

Kenji Fukaya, Antonella Grassi and Michele Rossi

This page intentionally left blank

Chapter 8

Deformation theory, homological algebra and mirror symmetry

Kenji Fukaya
Department of Mathematics, Kyoto University,
Kitashirakawa, Sakyo-ku, Kyoto, Japan

8.1 Introduction

In this article I wish to explain one of the relations between deformation theory and mirror symmetry. Deformation theory or theory of moduli is related to mirror symmetry in many ways. We discuss only one part of it here. The part we want to explain is related to a rather abstract and formal point of the theory of moduli, which was much studied in the 1950s and 1960s. Moduli are related to the definitions of schemes, stacks and their complex analytic analogues, and also to various parts of homological and homotopical algebra. Recently these topics have again caught the attention of several people working in areas closely related to mirror symmetry.

I first met moduli when was working [34] with K Ono on the construction of the Gromov–Witten invariant of general symplectic manifolds and studying the periodic orbit of periodic Hamiltonian systems. There we found that a C^∞ analogue of the notion of scheme and stack is appropriate to attack the transversality problem. The transversality problem we met was on the moduli space of holomorphic maps from the Riemann surface. Later I learned that in algebraic geometry, the same problem is studied by using stacks [8, 72]. This point, however, is not our main concern in this article. Our main focus is the next point.

I came across the relation of homological algebra to the theory of moduli while we (myself together with Y G Oh, H Ohta and K Ono [33]) were trying to find a good formulation for the Floer homology of the Lagrangian submanifold. There we first met the problem that the Floer homology of a

Lagrangian submanifold is not defined in general, so studying the condition when it is defined becomes an interesting problem. For this purpose, we developed an obstruction theory to enable the Floer homology to be defined. We next found that the Floer homology, even when it is defined, is not independent of the various choices involved.[1]

An example of this phenomenon is as follows. Let us consider a Lagrangian submanifold L in a symplectic manifold M. A problem, which is related to the definition of a Floer homology, is to count the number of holomorphic maps $\varphi : D^2 \to M$ such that $\varphi(\partial D^2) \subset L$.[2] Then the problem is that the number thus defined depends on the various choices involved. For example, it is *not* independent of the deformation of the (almost) complex structure of M. So unless we clarify the sense in which the number of holomorphic discs is independent of the various choices, it does not make mathematical sense to count it. It is at this essential point that we need deformation theory and homological algebra, i.e. we construct an algebraic structure using the number of discs, and the homotopy types of this algebraic structure are independent of the perturbation.

We thus developed a moduli theory for the deformations of Floer homology; i.e. we defined a moduli space $\mathcal{M}(L)$ for each Lagrangian submanifold L and the Floer homology is defined as a family of graded vector spaces parametrized by $\mathcal{M}(L_1) \times \mathcal{M}(L_2)$. The moduli space $\mathcal{M}(L)$ is related to the actual deformation of Lagrangian submanifolds but the relation is rather delicate. The algebraic machinery for constructing such a moduli space is one of A_∞ algebra and the Maurer–Cartan equation.

The A_∞ algebra we used there is a version of one I found in [23].[3] Using this A_∞ structure, Kontsevich [64,68] discovered a very interesting version of the mirror symmetry conjecture which he called the homological mirror symmetry conjecture. There it is conjectured that the Lagrangian submanifold corresponds to a coherent sheaf on the mirror bundle.

After developing a theory for the deformation of and obstruction to the Floer homology of the Lagrangian submanifold, we could make the homological mirror symmetry conjecture more precise. For example, we now conjecture that the moduli space $\mathcal{M}(L)$ will become a moduli space of the holomorphic vector bundles on the mirror. It is the purpose of this article to explain the formulation of homological mirror symmetry based on homological algebra and deformation theory.

[1] This problem is quite similar to Donaldson's gauge theory invariant of four manifolds with $b_2^+ = 1$ [17]. (Here $b_2^+ = 1$ is the number of positive eigenvalues of the intersection matrix on the second homology.)

[2] Definitions of several notions we need in symplectic geometry will be given at the beginning of section 8.3.2.

[3] My original motivation for introducing the A_∞ structure on the Floer homology was to use it to study the gauge theory Floer homology of 3-manifolds with a boundary. (The author was inspired by Segal and Donaldson to use category theory for this purpose.) The research toward this original direction is still in progress [25, 26] and the author believes that the relation of the A_∞ structure of Floer homology of Lagrangian submanifolds to gauge theory of 3-manifolds with a boundary will be related to some kind of duality in the future.

During the recent school on the 'Geometry and Physics of Branes', the author learned that there are several recent works by physicists which seem to be closely related to the theory we have developed. For example, the obstruction phenomenon seems to have been rediscovered. The fact that our Maurer–Cartan equation which controls the deformation of the Floer homology (see section 8.3.4) is inhomogeneous and zero is not its solution seems to be related to what is called 'tachyon condensation'.[4] I will not quote references to the papers by physicists on these points, since I expect these will be included in other parts of this book and I also find it hard to make the appropriate choice of papers to be quoted. It would be interesting to find a good dictionary between the physics and mathematics of the works. I hope that this book will be helpful for this purpose.

As we have already mentioned, the main purpose of this article is to describe a version of homological mirror symmetry precisely. For example, we want to state precisely the conjectured coincidence of the moduli spaces mentioned earlier. For this purpose, we need to review various basic aspects of moduli theory (especially its local version—deformation theory). Hence classical deformation theory of the holomorphic structures of vector bundles on complex manifolds (together with proofs of various parts of it) is included in this article.

Section 8.2 of this article is thus devoted to the theory of deformations. Deformation theory or the local theory of moduli is a classical topic initiated by Riemann for the moduli space of complex structures of Riemann surfaces. Kodaira and Spencer [61, 63], generalized it to higher dimensions and studied the deformation theory of complex structures of complex manifolds of higher dimension. This has been further amplified by many people, see, for example, [19, 71, 84, 104]. There are many versions of deformation theory, that is deformation theory of holomorphic vector bundles, complex submanifolds, holomorphic maps etc. But, as far as the points covered in section 8.2 are concerned, the differences among them are rather minor; so we mostly restrict ourselves to the case of vector bundles. In this article, we are taking the analytic point of view and use (nonlinear) partial differential equations, i.e. the algebraic theory of deformations (and of moduli). Some basic references are [4,20,45,78,92]. I have tried to make section 8.2 self-contained. In particular, I have tried to explain several abstract notions which are popular among algebraic geometers but not very popular among researchers in other fields. For example, we explain the notion of an analytic space (the complex analytic analogue of a scheme), the relation of category theory to the problem of moduli, especially the notion of a 'functor from an Artin ring', which is basic to a study of formal moduli. (Here 'formal moduli' means that we consider formal power series solutions of the defining equation of the moduli space.) Proofs of several of the main results of the local theory of moduli (the existence of the Kuranishi family, its completeness, versality, etc) are postponed to section 8.3, where we give a proof of them based on the homotopy theory of A_∞ algebras developed there. The content of section 8.1 is classical and no new

[4] I do not yet understand the precise relation of our theory to the one developed by physicists.

points of view are introduced. We include them here since most of the references I found require much background on algebraic geometry etc.

In section 8.3 we systematically explain how the homological algebra of A_∞ or L_∞ algebras can be applied to the problem of moduli. (A_∞ and L_∞ algebras are generalizations of differential graded algebras and of differential graded Lie algebras, respectively.) In section 8.2.1, we give a definition of them and define A_∞ and L_∞ homomorphisms. We also develop a homotopy theory of them, i.e. we define a homotopy between two A_∞ or L_∞ homomorphisms and the homotopy equivalence between two A_∞ or L_∞ algebras.

We next study the Maurer–Cartan equation from the point of view of a 'functor from an Artin ring' [92] which we explain in section 8.2.7.

We then sketch an important theorem which states that the gauge equivalence class of solutions of the Maurer–Cartan equation is invariant with respect to the homotopy types of A_∞ or L_∞ algebras. In differential graded algebras and differential graded Lie algebras, this result is due to [37, 38].

We then construct a Kuranishi family of the solutions of the Maurer–Cartan equation, as a quotient ring of an appropriate formal power series ring. We use a technique to sum over trees (calculation of the tree amplitude using the Feynman diagram) for this purpose. Several basic results postponed from section 8.2 (together with its generalization to A_∞ or L_∞ algebras) follows. In section 8.2.4, we briefly explain a translation of the theory in section 8.3 into the language of formal super geometry.

The theory developed in section 8.3 seems to have been studied by various people [7, 14, 46, 49, 50, 54, 56, 69, 77, 91, 96, 98] (I apologise to the authors of other papers on the subject which are not quoted here. I do not have enough knowledge to quote all important papers.) (Y Soibelman informed me that he and M Kontsevitch are preparing a book which overlaps this article.)

In section 8.4, we apply the discussion in sections 8.2 and 8.3 to homological mirror symmetry. We need to introduce a type of formal power series ring which we call the universal Novikov ring to study the instanton (or quantum) effect (in the symplectic geometry side of the theory). Our A_∞ algebra is a module over the universal Novikov ring and we need a slight modification of the definition of an A_∞ algebra which is explained in section 8.4.1.

In sections 8.4.2 and 8.4.3, Floer homology is explained. In section 8.4 we do not assume the reader to be familiar with global symplectic geometry. So we include section 8.4.2, which is an introduction to the part of global symplectic geometry related to section 8.4.3. In particular, we explain Floer's original construction [22]. In section 8.4.4, we explain the main construction of [33] which associates an A_∞ algebra with Lagrangian submanifolds. Our discussion in sections 8.4.2 and 8.4.3 is rather brief especially in the geometric and analytic points we need for the construction, since the main purpose of this article is to explain algebraic formalism rather than the basic geometric-analytic construction which is essential to give examples of the algebraic formalism. Details of the construction can be found in [33]; [28, 30, 83] are other surveys.

Section 8.4.5 is devoted to the definition of the moduli space $\mathcal{M}(L)$. To define it, we explain the modifications of the argument of section 8.2 which are necessary to apply it to the case when the coefficient ring is not $\mathbb{C}$ but a universal Novikov ring. Section 8.4.6 is devoted to explaining the complex geometry side of the story. An important point to be explained is what the Novikov ring in the complex geometry side corresponds to. Roughly speaking, the Novikov ring will become the ring of functions on the disc which parametrize the maximal degenerate family of mirror manifolds. We then discuss the fact that the mirror of a Lagrangian submanifold is a family of vector bundles (or more generally of objects of the derived category of coherent sheaves) over a maximally degenerate family of Calabi–Yau manifolds. A version of the homological mirror symmetry conjecture is then stated in which two A_∞ algebras over Novikov rings, one for Floer homology and the other for sheaf cohomology, coincide up to homotopy equivalence. There are some points which are not yet clear to me which are related to various deep problems in algebraic geometry. We conclude section 8.4 by giving an example.

My original plan was to include several other deformation theories related to mirror symmetry; for example, the extended deformation of the Calabi–Yau manifold due to [7], deformation quantization due to [69] and contact homology announced in [21].[5] They all can be treated by using the formalism in section 8.2. However, this article has already become too large and I will postpone this to another occasion.[6]

Parts of this article were announced in my joint paper [33] with Oh, Ohta and Ono. (A preliminary version of [33] was completed in December 2000 and is available from my home page at the time of writing this article. We are adding some new material to it, some of which is included in this article. The final version of [33] is now being completed.)

I would like to thank the organizers of the school 'Geometry and Physics of Branes' that gave me an opportunity to communicate with various researchers in a comfortable atmosphere and to write this article.

8.2 Classical deformation theory

8.2.1 Holomorphic structure on vector bundles

We start by describing deformation theory (that is a local theory of moduli) of holomorphic structures of complex vector bundles on complex manifolds. It is a classical theory and is a direct analogue of Kodaira and Spencer's study [61, 63] of the deformation of the complex structures of the complex manifold itself. We

5 Probably there is another example related to the period of the primitive form due to Saito [90]. I am unable to explain it at the time of writing this article. Some explanation from the point of view of mirror symmetry is found in [74].

6 The reader who speaks Japanese can find them in [29].

present it here since it is a prototype of the discussion which will appear later in less classical situations.

Let M be a complex manifold and $\Lambda^k_{\mathbb{C}} M = \oplus_{p+q=k} \Lambda^{p,q} M$ be the decomposition of the set of the complex valued k forms according to their types. We denote by $\Omega^{p,q}(M)$ the set of all smooth sections of $\Lambda^{p,q} M$. The complex structure of M is characterized by the Dolbault differential $\overline{\partial} : \Omega^{p,q}(M) \to \Omega^{p,q+1}(M)$. (See [40] for a standard textbook on complex manifold written from the transcendental point of view.) We remark that $(\Omega(M), \wedge, \overline{\partial})$ is a differential graded algebra, which we now define. Hereafter we denote by R a commutative ring with unit.

Definition 8.2.1. A *differential graded algebra* or *DGA* over R is a triple $(A^*, \cdot, d)$ with the following properties.

(1) For each $k \in \mathbb{Z}_{\geq 0}$, A^k is an R module. We write $\deg a = k$ if $a \in A^k$.
(2) $\cdot : A^k \otimes A^\ell \to A^{k+\ell}$ is an R module homomorphism, which is associative. Namely $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.
(3) $d : A^k \to A^{k+1}$ is an R module homomorphism such that $d \circ d = 0$.
(4) $d(a \cdot b) = d(a) \cdot b + (-1)^{\deg a} a \cdot d(b)$.

We may take either $(\oplus \Omega^{0,k}(M), \wedge, \overline{\partial})$ where k is the degree or $(\sum_{p,q} \Omega^{p,q}(M), \wedge, \overline{\partial})$ where the degree is the (total) degree of differential form. Let $\pi_E : E \to M$ be a complex vector bundle (see [58] for a standard textbook of differential geometry of holomorphic vector bundles.) We put $\Omega^{p,q}(M; E) = \Gamma(M, \Lambda^{p,q}(M; \Lambda^{p,q} M \otimes E))$. (We omit M to prevent confusion.) We define a wedge product: $\wedge : \Omega^{p,q}(M) \otimes \Omega^{p',q'}(M; E) \to \Omega^{p+p',q+q'}(M; E)$ in an obvious way. We define a holomorphic structure on our complex vector bundle as follows.

Definition 8.2.2. A *holomorphic structure* on E is a sequence of operators $\overline{\partial}_{\mathcal{E}} : \Omega^{0,q}(M; E) \to \Omega^{p,q+1}(M; E)$ $(p, q \in \{1, \dots, n\})$, such that (1) $\overline{\partial}_{\mathcal{E}} \circ \overline{\partial}_{\mathcal{E}} = 0$; and (2) $\overline{\partial}_{\mathcal{E}}(u \wedge \alpha) = \overline{\partial}(u) \wedge \alpha + (-1)^{p+q} u \wedge \overline{\partial}_{\mathcal{E}}(\alpha)$, for $u \in \Omega^{p,q}(M)$, $\alpha \in \Omega^*(M; E)$.

In other words, the holomorphic structure on E is a structure on $\Omega^*(M; E)$ of a left graded differential graded module over $(\Omega(M), \wedge, \overline{\partial})$, which we now define.

Definition 8.2.3. A *differential graded module* on a differential graded algebra $(A^*, \cdot, d)$ is, by definition, a triple $(C^*, \cdot, d)$ such that

(1) for each $k \in \mathbb{Z}$, M^k is an R module (we write $\deg a = k$ if $a \in M^k$);
(2) $\cdot : A^k \otimes M^\ell \to M^{k+\ell}$ is an R module homomorphism, which is associative (i.e. $(a \cdot b) \cdot x = a \cdot (b \cdot x)$, for $a, b \in A$, $x \in M$);
(3) $d : M^k \to M^{k+1}$ is an R module homomorphism such that $d \circ d = 0$;
(4) $d(a \cdot x) = d(a) \cdot x + (-1)^{\deg a} a \cdot d(x)$, for $a \in A$, $x \in M$.

We note that definition 8.2.2 coincides with another definition of a holomorphic vector bundle, the one which uses the local chart (see [58]).

From now on we *choose* one holomorphic structure $\overline{\partial}_{\mathcal{E}}$ on E and write $\mathcal{E} = (E, \overline{\partial}_{\mathcal{E}})$. Once we have fixed $\overline{\partial}_{\mathcal{E}}$, other holomorphic structures can be identified with elements of an affine space satisfying a differential equation, as we describe here. We consider the vector bundle $\mathrm{End}(E)$ whose fibre at p is $\mathrm{Hom}(E_p, E_p)$. (We omit M to prevent confusion.) Let $\Omega^{p,q}(M; \mathrm{End}(E))$ be the set of all smooth sections of $\Lambda^{p,q} \otimes \mathrm{End}(E)$. We define operators

$$\circ : \Omega^{p,q}(M; \mathrm{End}(E)) \otimes \Omega^{p',q'}(M; E) \to \Omega^{p+p',q+q'}(M; E)$$
$$\circ : \Omega^{p,q}(M; \mathrm{End}(E)) \otimes \Omega^{p',q'}(M; \mathrm{End}(E)) \to \Omega^{p+p',q+q'}(M; \mathrm{End}(E))$$

by using $\mathrm{End}(E) \otimes E \to E$, $\mathrm{End}(E) \otimes \mathrm{End}(E) \to \mathrm{End}(E)$ and the wedge product in an obvious way. A holomorphic structure $\overline{\partial}_{\mathcal{E}}$ on E induces a holomorphic structure (still denoted by $\overline{\partial}_{\mathcal{E}}$) on $\mathrm{End}(E)$ by

$$\overline{\partial}_{\mathcal{E}}(B) = \overline{\partial}_{\mathcal{E}} \circ B - (-1)^{\deg B} B \circ \overline{\partial}_{\mathcal{E}}. \tag{8.1}$$

Theorem 8.2.1. *Let $\overline{\partial}_{\mathcal{E}'}$ be another holomorphic structure on $\pi_E : E \to M$. Then, there exists a section $B \in \Omega^{0,1}(M; \mathrm{End}(E))$, such that*

$$\overline{\partial}_{\mathcal{E}'}(\alpha) = \overline{\partial}_{\mathcal{E}}(\alpha) + B \circ \alpha \tag{8.2}$$

B satisfies the differential equation

$$\overline{\partial}_{\mathcal{E}} B + B \circ B = 0. \tag{8.3}$$

However, let $B \in \Omega^{0,1}(M; \mathrm{End}(E))$ be a section satisfying (8.3). *We define $\overline{\partial}_{\mathcal{E}'}$ by* (8.2). *Then $\overline{\partial}_{\mathcal{E}'}$ defines a holomorphic structure on $\pi_E : E \to M$.*

Proof. By definition, we find $(\overline{\partial}_{\mathcal{E}'} - \overline{\partial}_{\mathcal{E}})(u \wedge \alpha) = (-1)^{\deg u} u \wedge (\overline{\partial}_{\mathcal{E}'} - \overline{\partial}_{\mathcal{E}})(\alpha)$. This implies that $\overline{\partial}_{\mathcal{E}'} - \overline{\partial}_{\mathcal{E}}$ is induced by a section of $\Omega^{0,1}(M; \mathrm{End}(E))$, which we denote by B. To show (8.3) we calculate

$$\begin{aligned}
\overline{\partial}_{\mathcal{E}'}(\overline{\partial}_{\mathcal{E}'}(\alpha)) &= \overline{\partial}_{\mathcal{E}'}(\overline{\partial}_{\mathcal{E}}(\alpha) + B \circ \alpha) \\
&= \overline{\partial}_{\mathcal{E}}(\overline{\partial}_{\mathcal{E}}(\alpha) + B \circ \alpha) + B \circ (\overline{\partial}_{\mathcal{E}}(\alpha) + B \circ \alpha) \\
&= \overline{\partial}_{\mathcal{E}} B \circ \alpha - B \circ \overline{\partial}_{\mathcal{E}}(\alpha) + B \circ \overline{\partial}_{\mathcal{E}}(\alpha) + B \circ B \circ \alpha \\
&= (\overline{\partial}_{\mathcal{E}} B + B \circ B) \circ \alpha.
\end{aligned} \tag{8.4}$$

Since (8.4) holds for any α, we have (8.3). The converse can be proved in the same way. □

Equation (8.3) is an example of the *Maurer–Cartan equation*, whose study is one of the main themes of this article.

8.2.2 Families of holomorphic structures on vector bundles

We study holomorphic vector bundles which are sufficiently close to $\mathcal{E}$. In other words, we are going to discuss a local moduli theory.

We first define a family of complex structures. Let $\mathcal{U} \subset \mathbb{C}^n$ be an open set. Let $\pi_{\hat{M}} : \hat{M} \to \mathcal{U}$ be a fibre bundle whose fibres are diffeomorphic to M.

Definition 8.2.4. A smooth (complex analytic) family of complex structures on M parametrized by $\mathcal{U}$ is a complex structure $J_{\hat{M}}$ on $\hat{M}$ such that $\pi_{\hat{M}} : (\hat{M}, J_{\hat{M}}) \to \mathcal{U}$ is holomorphic.

We next define a family of holomorphic vector bundles. Let $\hat{E} \to \hat{M}$ be a complex vector bundle. We assume that the restriction of $\hat{E}$ to $\pi_{\hat{M}}^{-1}(x) \cong M$ is isomorphic to E (as complex vector bundles).

Definition 8.2.5. A smooth (complex analytic) *family of the holomorphic structures* on E over $(\hat{M}, J_{\hat{M}})$ is a holomorphic structure $\overline{\partial}_{\hat{\mathcal{E}}}$ of the bundle $\hat{E}$.

One important case occurs when $(\hat{M}, J_{\hat{M}})$ is trivial, that is the case when $(\hat{M}, J_{\hat{M}})$ is isomorphic to the direct product $(M, J_M) \times \mathcal{U}$. (However, the case when the family $(\hat{M}, J_{\hat{M}})$ is non-trivial also appears later in our account in section 8.3.5.) In this case, we can use theorem 8.2.1 to identify a family of holomorphic structures on E to a map $\mathcal{U} \to \Omega^{0,1}(M; \mathrm{End}(E))$ as follows. Let $\overline{\partial}_{\hat{\mathcal{E}}}$ be a family of holomorphic structures on $\hat{E} = E \times \mathcal{U} \to \hat{M} = M \times \mathcal{U}$. Each $x \in \mathcal{U}$ determines a holomorphic structure $\overline{\partial}_{\mathcal{E}_x}$ on E, i.e. $\overline{\partial}_{\mathcal{E}_x}$ is the restriction of $\overline{\partial}_{\hat{\mathcal{E}}}$ to $M \times \{x\}$. We put $B_x = \overline{\partial}_{\mathcal{E}_x} - \overline{\partial}_{\mathcal{E}}$. Theorem 8.2.1 implies

$$\overline{\partial}_{\mathcal{E}} B_x + B_x \circ B_x = 0. \tag{8.5}$$

Using the fact that $\overline{\partial}_{\hat{\mathcal{E}}}$ is a holomorphic structure on $\hat{E}$ we can prove that the map

$$B : \mathcal{U} \to \Omega^{0,1}(M; \mathrm{End}(E)) \qquad x \mapsto B_x \tag{8.6}$$

is holomorphic. $(\Omega^{0,1}(M; \mathrm{End}(E))$ is a complex vector space (of infinite dimension). Hence, it makes sense to say that the map B is holomorphic. In contrast, given a holomorphic map (8.6) satisfying (8.5), we can define $\overline{\partial}_{\hat{\mathcal{E}}}$ by $\overline{\partial}_{\hat{\mathcal{E}}} = \overline{\partial}_{\mathcal{E}\times\mathcal{U}} + B$. Here $\overline{\partial}_{\mathcal{E}\times\mathcal{U}}$ is a holomorphic structure on $E \times \mathcal{U}$ (the direct product) and B is regarded as a smooth section of $\Omega^{0,1}(\hat{M}; \mathrm{End}(\hat{E}))$.

Our next purpose is to define and study notions of the completeness, versality and universality of families. We define them only in the case when the complex structure on M is fixed. The case of a family of complex structures on a manifold M and the case of holomorphic structures of vector bundles over $\hat{M}$ (with moving complex structures) are similar and are omitted.

We first need to define the morphism of two families, for this purpose. We first define the equivalence of holomorphic vector bundles. Let $\pi_{E_1} : E_1 \to M_1$

and $\pi_{E_2} : E_2 \to M_2$ be complex vector bundles. We consider a bundle homomorphism $\varphi : E_1 \to E_2$ over a holomorphic map $\overline{\varphi} : M_1 \to M_2$; i.e. $\pi_{E_2} \circ \varphi = \overline{\varphi} \circ \pi_{E_1}$ and φ is complex linear on each fibre. A bundle homomorphism φ induces $\varphi_* : \Omega^{p,q}(M; E_1) \to \Omega^{p,q}(M; E_2)$.

Definition 8.2.6. We say $\varphi : (E_1, \overline{\partial}_{\mathcal{E}_1}) \to (E_2, \overline{\partial}_{\mathcal{E}_2})$ is *holomorphic* if $\overline{\partial}_{\mathcal{E}_2} \circ \varphi_* = \varphi_* \circ \overline{\partial}_{\mathcal{E}_1}$. We say that $\varphi : (E_1, \overline{\partial}_{\mathcal{E}_1}) \to (E_2, \overline{\partial}_{\mathcal{E}_2})$ is an *isomorphism* if it is holomorphic and is a bundle isomorphism.

Let $\hat{E}_i = E \times \mathcal{U}_i$ and $\overline{\partial}_{\mathcal{E}_i}$ be a holomorphic structure on it, i.e. a deformation of holomorphic structures on E_i. We put $\hat{M}_i = M \times \mathcal{U}_i$.

Definition 8.2.7. A *morphism* from $(\hat{M}_1, \overline{\partial}_{\hat{\mathcal{E}}_1})$ to $(\hat{M}_2, \overline{\partial}_{\hat{\mathcal{E}}_2})$ is a pair (Φ, ϕ), where $\phi : \mathcal{U}_1 \to \mathcal{U}_2$ is a holomorphic map and Φ is a holomorphic bundle map $\Phi : (\hat{E}_1, \overline{\partial}_{\hat{\mathcal{E}}_1}) \to (\hat{E}_2, \overline{\partial}_{\hat{\mathcal{E}}_2})$ over $\mathrm{id} \times \phi : M \times \mathcal{U}_1 \to M \times \mathcal{U}_2$.

Let us define the notion of deformations of complex structure and of holomorphic vector bundle. It is the germ of a family and is defined as follows.

Definition 8.2.8. A *deformation* of a complex manifold (M, J) is a $\sim$ isomorphism class of a pair $(((\hat{M}, J), \mathcal{U}), i)$ where $(\hat{M}, J) = M \times \mathcal{U} \to \mathcal{U}$ is a family of complex structures and i is a (biholomorphic) isomorphism $\pi^{-1}(0) \cong (M, J)$.

We say $((M \times \mathcal{U}, J), i) \sim ((M \times \mathcal{U}', J'), i')$, if there exists an open neighbourhood $\mathcal{V}$ of 0 such that $\mathcal{V} \subset \mathcal{U} \cap \mathcal{U}'$ and if there exists a biholomorphic map $\varphi : (M \times \mathcal{V}, J) \to (M \times \mathcal{V}, J')$ which commutes with the projection: $M \times \mathcal{V} \to \mathcal{V}$ and which satisfies $i \circ \varphi = i$.

Let $\mathcal{E}$ be a holomorphic vector bundle on a complex manifold M. A deformation of $\mathcal{E}$ is an equivalence class of maps $B : \mathcal{U} \to \Omega^{0,1}(M, \mathrm{End}(E))$ such that $B(0) = 0$ and $\overline{\partial}_{\mathcal{E}} B(x) + B(x) \circ B(x) = 0$. We say that $B : \mathcal{U} \to \Omega^{0,1}(M, \mathrm{End}(E))$ is equivalent to $B' : \mathcal{U}' \to \Omega^{0,1}(M, \mathrm{End}(E))$ if there exists an open neighbourhood $\mathcal{V}$ of 0 with $\mathcal{V} \subset \mathcal{U} \cap \mathcal{U}'$ and a holomorphic map $\Phi : \mathcal{V} \to \Gamma(M, \mathrm{End}(E))$ such that $\Phi(0) = \mathrm{id}$ and $\Phi(x) \circ (\overline{\partial}_{\mathcal{E}} + B(x)) = (\overline{\partial}_{\mathcal{E}} + B'(x)) \circ \Phi(x)$.

We can define the deformation of a pair of complex structures and the vector bundle on it in a similar way. Hereafter we sometimes say B is a deformation of $\mathcal{E}$ or $(B, \mathcal{U})$ is a deformation of $\mathcal{E}$ by abuse of notation.

Now we define the completeness of a deformation. Roughly speaking, this means that all nearby holomorphic structures are contained in the family.

Definition 8.2.9. A deformation $(B, \mathcal{U})$ of $\mathcal{E}$ is said to be *complete* if the following condition holds.

Let $(B', \mathcal{U}')$ be another deformation of $\mathcal{E}$ and $\Phi_0 : E \to E$ is an automorphism of $\mathcal{E}$. Then, there exists a neighbourhood $\mathcal{V}$ of 0 in $\mathcal{U}'$ and a morphism

$$(\Phi, \phi) : (E \times \mathcal{V}, \overline{\partial}_{\mathcal{E}+B'}) \to (E \times \mathcal{U}, \overline{\partial}_{\mathcal{E}+B}) \tag{8.7}$$

of families in the sense of definition 8.2.7 such that

$$\Phi|_{M\times\{0\}} = \Phi_0. \tag{8.8}$$

The other important notion is the universality and versality of a deformation, which we now define. Let $(B, \mathcal{U})$ of $\mathcal{E}$ be a complete deformation.

Definition 8.2.10. We say that $(B, \mathcal{U})$ is *universal* if for each $(B', \mathcal{U}')$ as in definition 8.2.9 the morphism (Φ, ϕ) as in (8.7) satisfying (8.8) is unique.

We say that $(B, \mathcal{U})$ is *versal* if the differential of ϕ at 0 is unique; i.e. if (Φ', ϕ') as in (8.7) is another morphism satisfying (8.8) then $\mathrm{d}_0\phi' = \mathrm{d}_0\phi$. (Note they are both linear maps: $T_0\mathcal{V} \to T_0\mathcal{U}$.)

The difference between versality and universality is related to the stability of bundles.[7] We give an example of a versal family which is not universal in section 8.1.3.

8.2.3 Cohomology and deformations

The Maurer–Cartan equation (8.3) is a nonlinear partial differential equation. In this section, we study its linearization. The solution of a linearized equation is related to the cohomology group. Let $\mathcal{E}_i = (E_i, \overline{\partial}_{\mathcal{E}_i})$ be holomorphic vector bundles on M. We consider $\Omega^{0,q}(\mathrm{Hom}(E_1, E_2)) = \Gamma(M; \Lambda^{0,q} \otimes \mathrm{Hom}(E_1, E_2))$, where $\mathrm{Hom}(E_1, E_2)$ is a bundle whose fibre at p is $\mathrm{Hom}((E_1)_p, (E_2)_p)$ Operations $\overline{\partial}_{\mathcal{E}_1}$, $\overline{\partial}_{\mathcal{E}_2}$ define an operation $\overline{\partial}_{\mathcal{E}_1,\mathcal{E}_2} : \Omega^{0,q}(\mathrm{Hom}(E_1, E_2)) \to \Omega^{0,q+1}(\mathrm{Hom}(E_1, E_2))$ in the same way as (8.1). It is easy to see $\overline{\partial}_{\mathcal{E}_1,\mathcal{E}_2} \circ \overline{\partial}_{\mathcal{E}_1,\mathcal{E}_2} = 0$; i.e. $(\mathrm{Hom}(E_1, E_2), \overline{\partial}_{\mathcal{E}_1,\mathcal{E}_2})$ is a holomorphic vector bundle.

Definition 8.2.11. The extension $\mathrm{Ext}^q(\mathcal{E}, \mathcal{E}')$ is the qth cohomology of the chain complex $(\Omega^{0,*}(\mathrm{Hom}(E_1, E_2)), \overline{\partial}_{\mathcal{E}_1,\mathcal{E}_2})$.

Let $(B, \mathcal{U})$ be a deformation of $\mathcal{E}$. We are going to define a Kodaira–Spencer (KS) map $T_0\mathcal{U} \to \mathrm{Ext}^1(\mathcal{E}, \mathcal{E})$. By definition 8.2.8, we have

$$\overline{\partial}_{\mathcal{E}} B(x) + B(x) \circ B(x) = 0. \tag{8.9}$$

We differentiate (8.9) at 0. Then, in view of $B(0) = 0$, we have

$$\overline{\partial}_{\mathcal{E}} \left(\frac{\partial B}{\partial x^i}(0) \right) = 0.$$

Here $x = (x^1, \ldots, x^n)$ is a complex coordinate of $\mathcal{U} \subseteq \mathbb{C}^n$.

[7] See [78] for a definition of stability.

Definition 8.2.12. We put

$$\mathrm{KS}\left(\frac{\partial}{\partial x^i}\right) = \left[\frac{\partial B}{\partial x^i}(0)\right] \in \mathrm{Ext}^1(\mathcal{E}, \mathcal{E}).$$

KS is a linear map: $T_0\mathcal{U} \to \mathrm{Ext}^1(\mathcal{E}, \mathcal{E})$, which we call the *Kodaira–Spencer map* of our deformation.

A KS map is gauge equivariant; i.e. it is independent of the choice of the representative $(B, \mathcal{U})$ of the deformation. In other words, we have the following lemma. Let $(B', \mathcal{U}')$ be another representative of the deformation.

Lemma 8.2.1. *If* $(B, \mathcal{U})$ *is equivalent to* $(B', \mathcal{U}')$ *in the sense of definition* 8.2.8 *then*

$$\left[\frac{\partial B}{\partial x^i}(0)\right] = \left[\frac{\partial B'}{\partial x^i}(0)\right] \in \mathrm{Ext}^1(\mathcal{E}, \mathcal{E}).$$

Proof. $\mathcal{V} \subset \mathcal{U} \cap \mathcal{U}'$, $\Phi : \mathcal{V} \to \Gamma(M; \mathrm{End}(E))$ in definition 8.2.8 satisfies

$$\Phi(x) \circ (\overline{\partial}_{\mathcal{E}} + B(x)) = (\overline{\partial}_{\mathcal{E}} + B'(x)) \circ \Phi(x). \tag{8.10}$$

We differentiate (8.10) at 0 and obtain

$$\frac{\partial \Phi}{\partial x}(0) \circ \overline{\partial}_{\mathcal{E}} + \frac{\partial B'}{\partial x^i}(0) = \frac{\partial B}{\partial x^i}(0) + \overline{\partial}_{\mathcal{E}} \circ \frac{\partial \Phi}{\partial x^i}(0).$$

That is

$$\frac{\partial B}{\partial x^i}(0) - \frac{\partial B'}{\partial x^i}(0) = \overline{\partial}_{\mathcal{E}}\left(\frac{\partial \Phi}{\partial x^i}(0)\right).$$

□

In a similar way, we can prove the following lemma.

Lemma 8.2.2. *Let* $(B_1, \mathcal{U}_1)$, $(B_2, \mathcal{U}_2)$ *be deformations of* $\mathcal{E}_1, \mathcal{E}_2$ *and* $\Phi : (E_1 \times \mathcal{U}_1, \overline{\partial}_{\mathcal{E}_1 + B_1}) \to (E_2 \times \mathcal{U}_2, \overline{\partial}_{\mathcal{E}_2 + B_2})$, $\phi : \mathcal{U}_1 \to \mathcal{U}_2$ *be a morphism of a family of holomorphic structures in the sense of definition* 8.2.7. *We assume* $\phi(0) = 0$. *Then the following diagram commutes.*

$$\begin{array}{ccc} T_0\mathcal{U}_1 & \xrightarrow{\mathrm{KS}} & \mathrm{Ext}^1(\mathcal{E}, \mathcal{E}) \\ {\scriptstyle d_0\phi}\downarrow & & \Vert \\ T_0\mathcal{U}_2 & \xrightarrow{\mathrm{KS}} & \mathrm{Ext}^1(\mathcal{E}, \mathcal{E}) \end{array}$$

Diagram 1

The following result was proved in [62] in the case of deformation theory of complex structure.

Theorem 8.2.2. *If the KS map is surjective then the deformation is complete.*

We will prove this in section 8.2.3. Another main result of deformation theory is the following theorem which is due to Kodaira *et al* [60] in the case of the deformation theory of complex structures.

Theorem 8.2.3. *If* $\mathrm{Ext}^2(\mathcal{E},\mathcal{E}) = 0$ *then there exists a deformation of* $\mathcal{E}$ *such that the KS map is an isomorphism.*

We will prove this in section 8.1.5. We also prove that the family obtained in theorem 8.2.3 is unique up to an isomorphism in section 8.2.3.

Remark 8.2.1. The smooth family where the KS map is surjective does not exist, in general, in the case when $\mathrm{Ext}^2(\mathcal{E},\mathcal{E}) \neq 0$. (However there are cases where such families exist in the case $\mathrm{Ext}^2(\mathcal{E},\mathcal{E}) \neq 0$. See example 8.2.1.) Kuranishi [71] studied the case $\mathrm{Ext}^2(\mathcal{E},\mathcal{E}) \neq 0$. This leads us to the notion of a Kuranishi map: $\mathrm{Ext}^1(\mathcal{E},\mathcal{E}) \to \mathrm{Ext}^2(\mathcal{E},\mathcal{E})$. We will discuss this in section 8.1.6.

We note the following proposition.

Proposition 8.2.1. *The deformation obtained in theorem* 8.2.3 *is versal.*

Proof. Completeness follows from theorem 8.2.2. Let (Φ,ϕ) (Φ',ϕ') be morphisms as in (8.7) satisfying (8.8). By lemma 8.2.2 we have the following commutative diagram.

$$\begin{array}{ccccc}
T_0\mathcal{V} & \xrightarrow{\mathrm{KS}} & \mathrm{Ext}^1(\mathcal{E},\mathcal{E}) & \xleftarrow{\mathrm{KS}} & T_0\mathcal{V} \\
{\scriptstyle d_0\phi}\downarrow & & \| & & \downarrow{\scriptstyle d_0\phi'} \\
T_0\mathcal{U} & \xrightarrow{\mathrm{KS}} & \mathrm{Ext}^1(\mathcal{E},\mathcal{E}) & \xleftarrow{\mathrm{KS}} & T_0\mathcal{U}
\end{array}$$

Diagram 2

$\mathrm{d}_0\psi' = \mathrm{d}_0\psi$ follows immediately. □

We now give a few examples of deformations of holomorphic vector bundles.

Example 8.2.1. Let M be a complex manifold and let $L \to M$ be a complex *line* bundle. L has a holomorphic structure if its first Chern class is represented by a 1–1 form (see [40, 58]). We fix a holomorphic structure of L and denote it by $\overline{\partial}_{\mathcal{L}}$. The other holomorphic structure is equal to $\overline{\partial}_B = \overline{\partial}_{\mathcal{L}} + B$ where $B \in \Omega^{0,1}(M; \mathrm{End}(\mathcal{L}))$. Let us study equation (8.3) in this case.

Using the fact that L is a line bundle, it is easy to see that $\mathrm{End}(\mathcal{L})$ is isomorphic to the trivial line bundle (as a holomorphic line bundle); i.e. $B \in \Omega^{0,1}(M)$. Since B is a 1-form we have $B \circ B = 0$. Hence (8.3) reduces to a *linear* equation:

$$\overline{\partial} B_{\mathcal{L}} = 0.$$

We can find a vector subspace

$$\mathcal{U} \subset \operatorname{Ker}(\overline{\partial} : \Omega^{0,1}(M; \operatorname{End}(\mathcal{L})) \to \Omega^{0,2}(M; \operatorname{End}(\mathcal{L})))$$

such that the restriction $\mathcal{U} \to \operatorname{Ext}^1(\mathcal{L}, \mathcal{L})$ of the natural projection is an isomorphism. Hence, in the case of a line bundle, we always have a family whose KS map is an isomorphism. Note that if $\dim M \geq 2$, the condition $\operatorname{Ext}^2(\mathcal{L}, \mathcal{L}) \cong 0$ of theorem 1.3.2 may not be satisfied.

It is easy to see that our deformation is universal.

Example 8.2.2. Let $\mathcal{L}_1$, $\mathcal{L}_2$ be line bundles on M such that

$$\begin{aligned} &\operatorname{Ext}^1(\mathcal{L}_2, \mathcal{L}_1) \neq 0. \\ &\operatorname{Ext}^0(\mathcal{L}_2, \mathcal{L}_1) \cong \operatorname{Ext}^1(\mathcal{L}_1, \mathcal{L}_1) \cong \operatorname{Ext}^1(\mathcal{L}_2, \mathcal{L}_2) \cong \operatorname{Ext}^1(\mathcal{L}_1, \mathcal{L}_2) \cong 0. \end{aligned} \tag{8.11}$$

(Let $M = \mathbb{C}P^1$ and $c^1(L_1) = k_2[\mathbb{C}P^1]$, $c^1(L_1) = k_2[\mathbb{C}P^1]$, with $k_1 > k_2$. It is easy to check (8.11) in this case.) Let $B(x) \in \Omega^{0,1}(M; \operatorname{Hom}(L_2, L_1))$ be a form representing non-zero cohomology class x in $\operatorname{Ext}^1(\mathcal{L}_2, \mathcal{L}_1)$. We consider

$$\mathfrak{B}(x) = \begin{pmatrix} 0 & B(x) \\ 0 & 0 \end{pmatrix} \in \Omega^{0,1}(M; \operatorname{End}(L_1 \oplus L_2)).$$

Then

$$\overline{\partial}_{\mathfrak{B}(x)} = \overline{\partial}_{\mathcal{L}_1 \oplus \mathcal{L}_2} + \mathfrak{B}(x) = \begin{pmatrix} \overline{\partial}_{\mathcal{L}_1} & B(x) \\ 0 & \overline{\partial}_{\mathcal{L}_2} \end{pmatrix}$$

satisfies $\overline{\partial}_{\mathfrak{B}(x)} \circ \overline{\partial}_{\mathfrak{B}(x)} = 0$ and hence we have a deformation of $\mathcal{L}_1 \oplus \mathcal{L}_2$. By (8.11) we can easily find $\operatorname{Ext}^1(\mathcal{L}_1 \oplus \mathcal{L}_2, \mathcal{L}_1 \oplus \mathcal{L}_2) \cong \operatorname{Ext}^1(\mathcal{L}_2, \mathcal{L}_1)$. The KS map of our deformation is the identity: $\operatorname{Ext}^1(\mathcal{L}_2, \mathcal{L}_1) \to \operatorname{Ext}^1(\mathcal{L}_1 \oplus \mathcal{L}_2, \mathcal{L}_1 \oplus \mathcal{L}_2) \cong \operatorname{Ext}^1(\mathcal{L}_2, \mathcal{L}_1)$. We thus have a versal deformation.

However this deformation is *not* universal.

In fact, let $r : \operatorname{Ext}^1(\mathcal{L}_2, \mathcal{L}_1) \to \mathbb{C}\backslash\{0\}$ be a holomorphic function with $r(0) = 1$. We define $\phi : \operatorname{Ext}^1(\mathcal{L}_2, \mathcal{L}_1) \to \operatorname{Ext}^1(\mathcal{L}_2, \mathcal{L}_1)$ by $\phi(x) = r(x)x$. We also define $\Psi : \operatorname{Ext}^1(\mathcal{L}_2, \mathcal{L}_1) \to \operatorname{End}(L_1 \oplus L_2)$ by $\Psi(x)(v, w) = (r(x)v, w)$. Ψ defines a bundle homomorphism over ϕ. By definition it is easy to see that $\Psi(x) : ((L_1 \oplus L_2), \overline{\partial}_{\mathcal{L}+\mathfrak{B}(x)}) \to ((L_1 \oplus L_2), \overline{\partial}_{\mathcal{L}+\mathfrak{B}(x)})$ is holomorphic. We thus find a morphism of a family of holomorphic structures, which is the identity on the fibre of 0 but not at the fibre of other points. Hence our deformation is not universal.

We can explain this phenomenon as follows. We consider the group of (holomorphic) automorphisms of the bundles $(L_1 \oplus L_2; \overline{\partial}_{\mathfrak{B}(x)})$. In case $x = 0$ this bundle is a direct product. Hence we have $\operatorname{Aut}(L_1 \oplus L_2; \overline{\partial}_{\mathfrak{B}(0)}) \cong \operatorname{Aut}(\mathcal{L}_1) \times \operatorname{Aut}(\mathcal{L}_2) \times \operatorname{Ext}^0(\mathcal{L}_1, \mathcal{L}_2) \cong \mathbb{C}_*^2 \times \operatorname{Ext}^0(\mathcal{L}_1, \mathcal{L}_2)$. However, for $x \neq 0$ we have $\operatorname{Aut}(L_1 \oplus L_2; \overline{\partial}_{\mathfrak{B}(x)}) \cong \mathbb{C}_* \times \operatorname{Ext}^0(\mathcal{L}_1, \mathcal{L}_2)$ where $c \in \mathbb{C}_*$ acts on $L_1 \oplus L_2$ by $(v, w) \mapsto (cv, cw)$. The difference $\operatorname{Aut}(L_1 \oplus L_2; \overline{\partial}_{\mathfrak{B}(0)}) / \operatorname{Aut}(L_1 \oplus L_2; \overline{\partial}_{\mathfrak{B}(x)}) \cong$

$\mathbb{C}_*$ acts on $\mathrm{Ext}^1(\mathcal{L}_2, \mathcal{L}_1)$ as a scalar multiplication. We can easily check that $(L_1 \oplus L_2; \overline{\partial}_{\mathfrak{B}(x)})$ is isomorphic to $(L_1 \oplus L_2; \overline{\partial}_{\mathfrak{B}(x')})$ if and only if $x' = cx$, i.e. in the case when x and x' lie in the same orbit of $\mathbb{C}_*$-action. Hence the part of the automorphism of $(L_1 \oplus L_2, \overline{\partial}_{B(x)}) : x = 0$ which will be 'lost' for $x \neq 0$ will act on $\mathcal{U} = \mathrm{Ext}^1(\mathcal{L}_2, \mathcal{L}_1)$. The set of isomorphism classes of nearby holomorphic structures will be the orbit space of this $\mathbb{C}_*$-action. The quotient space of this action in our case is a union of $\mathbb{C}P^N$ and one point where $N + 1 = \operatorname{rank} \mathrm{Ext}^1(\mathcal{L}_2, \mathcal{L}_1)$. If we use the quotient topology then the quotient space is not Hausdorff at the orbit of origin. This is because $\mathbb{C}_*$ is non-compact.

The phenomenon explained here, i.e. that the jump of the dimension of the automorphism group is related to a versal but not universal family, is rather general. In fact, we can prove the following result. We recall that the Lie algebra of $\mathrm{Aut}(\mathcal{E})$ is identified with $\mathrm{Ext}^0(\mathcal{E}, \mathcal{E})$.

Theorem 8.2.4. *If the deformation $(B, \mathcal{U})$ is versal and if the rank of $\mathrm{Ext}^0((E, \overline{\partial}_{\mathcal{E}+B(x)}), (E, \overline{\partial}_{\mathcal{E}+B(x)}))$ is independent of x in a neighbourhood of 0, then $(B, \mathcal{U})$ is universal.*

We will prove this in section 2.3.

8.2.4 Bundle valued harmonic forms

We prove theorem 8.2.3 in the next section. To prove theorem 8.2.3, we use the harmonic theory of vector bundle valued forms, which we review in this section (see [58] for details). Let $\kappa : \Omega^{p,q}(M) \to \Omega^{q,p}(M)$ be the complex antilinear homomorphism defined by

$$\begin{aligned}\kappa(u_{i_1,\ldots,i_p;j_1,\ldots,j_q}\, \mathrm{d}z^{i_1} \wedge \ldots \mathrm{d}z^{i_p} \wedge \mathrm{d}\overline{z}^{j_1} \wedge \ldots \mathrm{d}\overline{z}^{j_q}) \\ = \overline{u}_{i_1,\ldots,i_p;J_1,\ldots,j_q}\, \mathrm{d}\overline{z}^{i_1} \wedge \ldots \mathrm{d}\overline{z}^{i_p} \wedge \mathrm{d}z^{j_1} \wedge \ldots \mathrm{d}z^{j_q}.\end{aligned}$$

We next fix a Hermitian metric g on M. Then g induces the Hodge $*$ operator $* : \Lambda^k(M) \to \Lambda^{n-k}M$ by $u \wedge \kappa(*v) = g(u, v)\,\mathrm{Vol}_g$ where $\mathrm{Vol}_g \in \Omega^{2n}(M)$ is the volume element and $g(u, v)$ is the inner product on $\Lambda^*(M)$ induced by g. We note that $*$ is complex linear, $\kappa \circ * = * \circ \kappa$ and $* : \Lambda^{p,q}(M) \to \Lambda^{n-q,n-p}(M)$. We can also check that

$$** = (-1)^{p+q}\,\mathrm{id} \qquad \text{on } \Lambda^{p,q}(M). \tag{8.12}$$

Definition 8.2.13. We define the operator $\overline{\partial}^*$ by $-\kappa \circ * \circ \overline{\partial} \circ \kappa \circ * : \Lambda^{p,q}(M) \to \Lambda^{p,q-1}(M)$.

$\overline{\partial}^*$ is complex linear. Let us now include the holomorphic vector bundle. Let $\mathcal{E} = (E, \overline{\partial}_{\mathcal{E}})$ be a holomorphic vector bundle over M. We take and fix a Hermitian inner product h on E. h induce an anticomplex linear homomorphism $I_h : E \to E^*$. I_h and κ induce an anticomplex linear homomorphism $\kappa_h : \Omega^{p,q}(M; E) \to$

$\Omega^{q,p}(M; E^*)$. We define $\wedge : \Omega^{p,q}(M; E) \otimes \Omega^{p',q'}(M; E^*) \to \Omega^{p+p',q+q'}(M)$ by $(u \otimes a) \wedge (v \otimes \alpha) = \alpha(a) u \wedge v$. Then we define $* : \Omega^{p,q}(M; E) \to \Omega^{n-q,n-p}(M; E^*)$ by $a \wedge \kappa_h(* B) = g_h(A, B)\Omega_g$, $A, B \in \Omega^{p,q}(M; E)$, where g_h is an inner product on $\Lambda^{p,q}(M; E)$ induced by g and h. Formula (8.12) holds. We define

$$\overline{\partial}^*_{\mathcal{E}} = -\kappa \circ * \circ \overline{\partial}_{\mathcal{E}} \circ \kappa \circ * : \Lambda^{p,q}(M; E) \to \Lambda^{p,q-1}(M; E).$$

We define the Hermitian inner product $\langle \cdot, \cdot \rangle$ on $\Omega^{p,q}(M)$ and $\Omega^{p,q}(E)$ by

$$\langle u, v \rangle = \int_M g(u, v) \,\mathrm{Vol}_g \qquad \langle A, B \rangle = \int_M g_h(A, B) \,\mathrm{Vol}_g .$$

Let $L^2(M; \Lambda^{p,q}(M))$, $L^2(M; \Lambda^{p,q}(M; E))$ be the completion. They are Hilbert spaces. We can prove $\langle \overline{\partial} u, v \rangle = \langle u, \overline{\partial}^* v \rangle$, $\langle \overline{\partial}_{\mathcal{E}} A, B \rangle = \langle A, \overline{\partial}^*_{\mathcal{E}} B \rangle$ by using Stokes theorem and (8.12). We now define the Laplace–Beltrami operator by $\Delta_{\overline{\partial}} = \overline{\partial}\,\overline{\partial}^* + \overline{\partial}^* \overline{\partial}$, $\Delta_{\overline{\partial}_{\mathcal{E}}} = \overline{\partial}_{\mathcal{E}} \overline{\partial}^*_{\mathcal{E}} + \overline{\partial}^*_{\mathcal{E}} \overline{\partial}_{\mathcal{E}}$, and the space of harmonic forms and sections by

$$\mathcal{H}^{p,q}(M) = \mathrm{Ker}(\Delta_{\overline{\partial}} : \Omega^{p,q}(M) \to \Omega^{p,q}(M))$$
$$\mathcal{H}^{p,q}(M; \mathcal{E}) = \mathrm{Ker}(\Delta_{\overline{\partial}_{\mathcal{E}}} : \Omega^{p,q}(M; E) \to \Omega^{p,q}(M; E)).$$

We can show that $\Delta_{\overline{\partial}}$, $\Delta_{\overline{\partial}_{\mathcal{E}}}$ are elliptic and hence $\mathcal{H}^{p,q}(M)$ and $\mathcal{H}^{p,q}(M; \mathcal{E})$ are finite dimensional if M is compact without boundary.

Let $\Pi_{\mathcal{H}} : L^2(M; \Lambda^{p,q}(M)) \to \mathcal{H}^{p,q}(M)$, $\Pi_{\mathcal{H},\mathcal{E}} : L^2(M; \Lambda^{p,q}(M; E)) \to \mathcal{H}^{p,q}(M; \mathcal{E})$ be orthonormal projections.

Now the basic result due to Hodge–Kodaira is:

Theorem 8.2.5. *There exists an orthonormal decomposition :*

$$L^2(M; \Lambda^{p,q}(M)) \cong \mathrm{Im}\,\overline{\partial} \oplus \mathrm{Im}\,\overline{\partial}^* \oplus \mathcal{H}^{p,q}(M)$$
$$L^2(M; \Lambda^{p,q}(M; E)) \cong \mathrm{Im}\,\overline{\partial}_{\mathcal{E}} \oplus \mathrm{Im}\,\overline{\partial}^*_{\mathcal{E}} \oplus \mathcal{H}^{p,q}(M; \mathcal{E}).$$

There exist operators $Q : L^2(M; \Lambda^{p,q}(M)) \to L^2(M; \Lambda^{p,q}(M))$, $Q_{\mathcal{E}} : L^2(M; \Lambda^{p,q}(M); \mathcal{E}) \to L^2(M; \Lambda^{p,q}(M); \mathcal{E})$ such that $\Delta_{\overline{\partial}} \circ Q = \mathrm{id} - \Pi_{\mathcal{H}}$, $\Delta_{\overline{\partial}_{\mathcal{E}}} \circ Q_{\mathcal{E}} = \mathrm{id} - \Pi_{\mathcal{H},\mathcal{E}}$.

(For the proof see, for example, [105].) We note that Q commutes with $\overline{\partial}$, $\overline{\partial}^*$, Δ^*. We put $G = \overline{\partial}^* \circ Q$, $G_{\mathcal{E}} = \overline{\partial}^*_{\mathcal{E}} \circ Q_{\mathcal{E}}$ and call them the *propagators*. G, $G_{\mathcal{E}}$ are the chain homotopy between the identity and the orthonormal projections $\Pi_{\mathcal{H}}$, $\Pi_{\mathcal{H},\mathcal{E}}$, i.e. we can easily prove that:

$$\mathrm{id} - \Pi_{\mathcal{H}} = G \circ \overline{\partial} + \overline{\partial} \circ G \qquad \mathrm{id} - \Pi_{\mathcal{H}_{\mathcal{E}}} = G_{\mathcal{E}} \circ \overline{\partial}_{\mathcal{E}} + \overline{\partial}_{\mathcal{E}} \circ G_{\mathcal{E}}. \qquad (8.13)$$

8.2.5 Construction of a versal family and Feynman diagrams

In this section we prove theorem 8.2.3. We are going to find a neighbourhood $\mathcal{U}$ with its origin in $\mathcal{H}^{p,q}(M; (\mathrm{End}(E), \overline{\partial}_{\mathcal{E}}))$ and construct a holomorphic map $B : \mathcal{U} \to \Omega^{0,1}(M; \mathrm{End}(E))$ such that

$$\overline{\partial}_{\mathcal{E}} B(b) + B(b) \circ B(b) = 0 \tag{8.14}$$

and that $\mathrm{d}_0 B : T_0\mathcal{U} = \mathcal{H}^{0,1}(M; (\mathrm{End}(E), \overline{\partial}_{\mathcal{E}})) \to \Omega^{0,1}(M; \mathrm{End}(E))$ is the identity. As we have already discussed, this is enough to show theorem 8.2.3. We take a formal parameter T and put

$$B(b) = Tb + \sum_{k=2}^{\infty} T^k B_k(b). \tag{8.15}$$

We solve equation (8.14) inductively on k, i.e. we solve

$$\overline{\partial}_{\mathcal{E}} B_k(b) = - \sum_{\ell+m=k} B_\ell(b) \circ B_m(b) \tag{8.16}$$

inductively on k. The solution of (8.16) is given by using the operator $G_{\mathcal{E}}$, the propagator, introduced in the last section. (Here we write $G_{\mathcal{E}}$ in place of $G_{\mathrm{End}(\mathcal{E})}$ for simplicity.)

Lemma 8.2.3. *We define $B_1(b) = b$ and*

$$B_k(b) = \sum_{\ell+m=k} G_{\mathcal{E}}(B_\ell(b) \circ B_m(b)) \tag{8.17}$$

inductively on k. Then it satisfies (8.16).

Proof. We remark that the harmonic projection

$$\Pi_{\mathcal{H},\mathrm{End}(\mathcal{E})} : L^2(M; \Lambda^{0,2}(M; \mathrm{End}(E))) \to \mathcal{H}^{0,2}(M; \mathrm{End}(E))$$

is zero because $\mathrm{Ext}^2(\mathcal{E}, \mathcal{E}) = 0$ by assumption. Hence, by (8.2.5), we have

$$\begin{aligned}\overline{\partial}_{\mathcal{E}} B_k(b) &= \sum_{\ell+m=k} \overline{\partial}_{\mathcal{E}} G_{\mathcal{E}}(B_\ell(b) \circ B_m(b)) \\ &= - \sum_{\ell+m=k} B_\ell(b) \circ B_m(b) - \sum_{\ell+m=k} G_{\mathcal{E}} \overline{\partial}_{\mathcal{E}}(B_\ell(b) \circ B_m(b)).\end{aligned}$$

By using the induction hypothesis, we have

$$\begin{aligned}
&\sum_{\ell+m=k} \overline{\partial}_{\mathcal{E}}(B_\ell(b) \circ B_m(b)) \\
&\quad = \sum_{\ell+m=k} \overline{\partial}_{\mathcal{E}} B_\ell(b) \circ B_m(b) - \sum_{\ell+m=k} B_\ell(b) \circ \overline{\partial}_{\mathcal{E}} B_m(b) \\
&\quad = - \sum_{\ell+m=k} \sum_{\ell_1+\ell_2=\ell} (B_{\ell_1}(b) \circ B_{\ell_2}(b)) \circ B_m(b) \\
&\qquad + \sum_{\ell+m=k} \sum_{m_1+m_2=m} B_\ell(b) \circ (B_{m_1}(b) \circ B_{m_2}(b)) = 0.
\end{aligned}$$

(We note that the associativity of $\circ$ plays an important role here.) □

To complete the proof of theorem 8.2.3, it suffices to show the following lemma.

Lemma 8.2.4. *There exists* $\epsilon > 0$ *such that if* $|T|\|b\| < \epsilon$ *then* (8.15) *converges.*

(Here $\|b\|$ is the Sobolev L^2_k norm of b with sufficiently large k, which will be introduced later.) The proof of lemma 8.2.4 is based on a standard result, in geometric analysis, which we briefly recall here. First for $A \in \Omega^{p,q}(\mathrm{End}(E))$ we define its Sobolev norm $\|A\|_{L^2_d}$ by

$$\|A\|^2_{L^2_d} = \sum_{i=0}^{k} \langle \nabla^i A, \nabla^i A \rangle.$$

Here $\nabla^i A$ is the ith covariant derivative of A and $\langle \nabla^i A, \nabla^i A \rangle$ is its appropriate L^2 inner product defined similarly to that in the last section. Let $L^2_d(M; \mathrm{End}(E))$ be the completion of $\Omega^{p,q}(\mathrm{End}(E))$ with respect to $\|A\|_{L^2_d}$. Then

(A) $Q_{\mathcal{E}}$ defines a bounded operator

$$L^2_d(M; \Lambda^{p,q}(M; \mathrm{End}(E))) \to L^2_{d+2}(M; \Lambda^{p,q}((M; \mathrm{End}(E))).$$

(B) $\circ : \Omega^{p,q}((M; \mathrm{End}(E)) \otimes \Omega^{p',q'}((M; \mathrm{End}(E)) \to \Omega^{p+p',q+q'}((M; \mathrm{End}(E))$ can be extended to a continuous operator

$$\begin{aligned}
L^2_{d+1}(M; \Lambda^{p,q}(M; \mathrm{End}(E))) &\otimes L^2_{d+1}(M; \Lambda^{p',q'}(M; \mathrm{End}(E))) \\
&\to L^2_d(M; \Lambda^{p+p',q+q'}(M; \mathrm{End}(E))).
\end{aligned}$$

if k is sufficiently large compared to $2n = \dim_{\mathbb{R}} M$.

The proof of (A) is given in many of the standard textbooks on harmonic theory, for example [105]. The proof of (B) is given in many of the standard textbooks on Sobolev spaces, see, for example, [36]. Now (A),(B) implies

$$\|G_{\mathcal{E}}(B_\ell(b) \circ B_m(b))\|_{L^2_k} < C \|B_\ell(b)\|_{L^2_k} \|B_m(b)\|_{L^2_k}$$

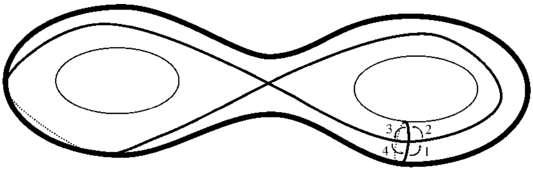

Figure 8.1.

if k is large. Here C is independent of k. Therefore, we can show

$$\|B_k(b)\|_{L^2_d} < (C'\|b\|_{L^2_d})^k$$

inductively on k. Lemma 8.2.4 follows immediately. □

We defined $B_k(b)$ inductively on k. We can rewrite it and define $B_k(b)$ as a sum over Feynman diagrams. In order to show the relation of theorem 8.2.3 to quantum field theory, let us do it here.

Definition 8.2.14. A *finite oriented graph* Γ consists of the following data:

(1) a finite set $\mathrm{Vertex}(\Gamma)$, the set of vertices;
(2) a finite set $\mathrm{Edge}(\Gamma)$, the set of edges; and
(3) maps $\partial_{\mathrm{source}} : \mathrm{Edge}(\Gamma) \to \mathrm{Vertex}(\Gamma)$, $\partial_{\mathrm{target}} : \mathrm{Edge}(\Gamma) \to \mathrm{Vertex}(\Gamma)$.

A *ribbon structure* of an oriented graph Γ is the cyclic ordering of the set $\partial^{-1}_{\mathrm{source}}(v) \cup \partial^{-1}_{\mathrm{target}}(v)$ for each $v \in \mathrm{Vertex}(\Gamma)$. A graph with a ribbon structure is called a *ribbon graph*.

We take copies of intervals $[0, 1]_e$ corresponding to each element of $e \in \mathrm{Edge}(\Gamma)$ and copies of points v corresponding to each element $v \in \mathrm{Vertex}(\Gamma)$. We identify $\{0\} \in [0, 1]_e$ with $\partial_{\mathrm{source}}(e)$ and $\{1\} \in [0, 1]_e$ with $\partial_{\mathrm{target}}(e)$. We thus obtain a one-dimensional complex, which we write as $|\Gamma|$.

An embedding of $|\Gamma|$ into an oriented surface (a real two-dimensional manifold) Σ induces a ribbon structure on Γ as in figure 8.1.

In this article we only consider finite oriented graphs which we call graphs. Now we consider ribbon graphs Γ satisfying the following conditions.

Condition 8.2.1.

(1) $\mathrm{Vertex}(\Gamma)$ is decomposed into a disjoint union of $\mathrm{Vertex}_{\mathrm{int}}(\Gamma)$ and $\mathrm{Vertex}_{\mathrm{ext}}(\Gamma)$.
(2) If $v \in \mathrm{Vertex}_{\mathrm{int}}(\Gamma)$ then $\sharp\partial^{-1}_{\mathrm{target}}(v) = 2$, $\sharp\partial^{-1}_{\mathrm{source}}(v) = 1$.
(3) There exists $v_{\mathrm{last}} \in \mathrm{Vertex}_{\mathrm{ext}}(\Gamma)$, the last vertex, such that $\sharp\partial^{-1}_{\mathrm{target}}(v) = 1$, $\sharp\partial^{-1}_{\mathrm{source}}(v) = 0$.
(4) If $v \in \mathrm{Vertex}_{\mathrm{ext}}(\Gamma)\setminus\{v_{\mathrm{last}}\}$ then $\sharp\partial^{-1}_{\mathrm{target}}(v) = 0$, $\sharp\partial^{-1}_{\mathrm{source}}(v) = 1$.
(5) $|\Gamma|$ is connected and simply connected.

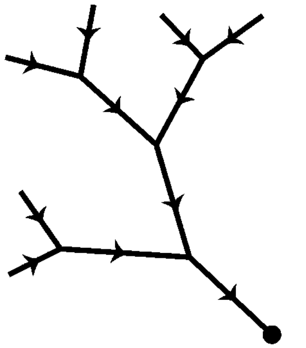

Figure 8.2.

We say an element of $\mathrm{Vertex}_{\mathrm{int}}(\Gamma)$ is an *interior vertex* and an element of $\mathrm{Vertex}_{\mathrm{ext}}(\Gamma)$ an *exterior vertex*. An edge e is called an *exterior edge* if $\{\partial_{\mathrm{target}}(e), \partial_{\mathrm{source}}(e)\} \cap \mathrm{Vertex}_{\mathrm{ext}}(\Gamma) \neq \emptyset$. Otherwise it is called an *interior edge*.

Let Γ be a ribbon graph such that $|\Gamma|$ is simply connected. It is then easy to see that there exists an embedding $|\Gamma| \to \mathbb{R}^2$ such that the ribbon structure is compatible with the orientation of $\mathbb{R}^2$.

We denote by $RG_{k,2}$ the set of all ribbon graphs Γ which satisfies condition 8.2.1 and has exactly $k+1$ exterior vertices.

Let $\Gamma \in RG_{k,2}$. Then, using the embedding $|\Gamma| \to \mathbb{R}^2$ compatible with the ribbon structure, we obtain a cyclic order on $\mathrm{Vertex}_{\mathrm{ext}}(\Gamma)$. By regarding the last vertex as the zeroth one, the cyclic order determines the order on $\mathrm{Vertex}_{\mathrm{ext}}(\Gamma)\backslash\{v_{\mathrm{last}}\}$. So we put $\mathrm{Vertex}_{\mathrm{ext}}(\Gamma)\backslash\{v_{\mathrm{last}}\} = \{v_1, \ldots, v_k\}$. We are now going to define $B_\Gamma : \mathcal{H}^{0,1}(M; (\mathrm{End}(E), \overline{\partial}_{\mathcal{E}}))^{k\otimes} \to \Omega^{0,1}(M; \mathrm{End}(E))$ for each $\Gamma \in Gr_{k,2}$ such that

$$B_k(b) = \sum_{\Gamma \in Gr_{k,2}} B_\Gamma(b, \ldots, b). \tag{8.18}$$

We define B_Γ by induction on k. Let $\Gamma \in RG_{k,2}$ and v_{last} be its last vertex. Let e_{last} be the unique edge such that $\partial_{\mathrm{target}}(e_{\mathrm{last}}) = v_{\mathrm{last}}$. We remove $[0,1]_{e_{\mathrm{last}}}$ together with its two vertices from $|\Gamma|$. Then $|\Gamma|\backslash[0,1]_{e_{\mathrm{last}}}$ is a union of two components which can be regarded as $|\Gamma_1|$ and $|\Gamma_2|$, where $\Gamma_1 \in RG_{2,\ell}$, $\Gamma_2 \in RG_{2,m}$ with $\ell + m = k$. We may number Γ_1, Γ_2 so that $v_1, \ldots, v_\ell \in \Gamma_1$.

Now we put

$$B_\Gamma(b_1, \ldots, b_k) = G_{\mathcal{E}}(B_{\Gamma_1}(b_1, \ldots, b_\ell) \circ B_{\Gamma_2}(b_{\ell+1}, \ldots, b_k)). \tag{8.19}$$

(8.18) is quite obvious by definition.

The way we rewrite the induction process into a sum over trees is a straightforward matter. However, rewriting the definition of B_k as in (8.18) leads us naturally to the following questions.

Question 8.2.1. (1) Can we generalize to the case when the interior vertex has more than two edges?
(2) When we consider a tree instead of a ribbon tree, is there any corresponding theory?
(3) What happens when we include the graph which is not simply connected?

Remarkably they all have good answers.

Answer 8.2.1. (1) We then study the deformation of A_∞ algebras in place of differential graded algebras.
(2) We then study the deformation of differential graded Lie algebras or, more generally, L_∞ algebras.
(3) This corresponds to Reidemeister or analytic torsion [87] (in the case $H_1(|\Gamma|) = \mathbb{Z}$), Chern–Simons perturbation theory [5, 6, 24, 66], quantum KS theory [10] or a pseudoholomorphic map from a higher-genus Riemann surface (with or without boundary).

We will explain the first two answers in later sections. The detailed study of the third one is left to the future, since mathematical theory, in the case $H_1(|\Gamma|) \neq \mathbb{Z}$, is not sufficiently well developed.

8.2.6 The Kuranishi family

In this section, we remove the assumption $\mathrm{Ext}^2(\mathcal{E}, \mathcal{E}) = 0$ from theorem 8.2.2. We need to study deformations parametrized by a singular variety for this purpose. Let us start by briefly reviewing the notion of analytic variety. Since we only discuss here the local moduli theory it is enough to consider the case of an analytic subspace of $\mathbb{C}^N$. (See [20] for more details on analytic variety.)

Definition 8.2.15. Let $X \subset \mathbb{C}^N$ be a locally closed subset. We say that X is a (complex) *analytic subset* if the following holds.

For each $p \in X$ there exists a neighbourhood U of p in $\mathbb{C}^N$ and holomorphic functions $f_1, \ldots, f_m$ such that $X \cap U = \{z | f_1(z) = \cdots = f_m(z) = 0\}$.

Definition 8.2.16. For $p \in X$, we put

$$\mathfrak{I}_{X,p} = \{f \in \mathcal{O}_p | f \equiv 0 \text{ on } X\}. \tag{8.20}$$

Here $\mathcal{O}_p$ is the germ of holomorphic functions at p, i.e. the set of convergent power series in a neighbourhood of p.

The germ of holomorphic functions $\mathcal{O}_{X,p}$ of X at p is defined by $\mathcal{O}_{X,p} = \mathcal{O}_{X,p}/\mathfrak{I}_{X,p}$. $\mathcal{O}_{X,p}$ is a local ring and its maximal ideal is $\mathcal{O}_{X,p,+} = \{[f] \in \mathcal{O}_{X,p} | f(p) = 0\}$. (Here and hereafter, we denote by $\mathcal{R}_+$ its maximal ideal of a local ring $\mathcal{R}$.)

Let $X \subset \mathbb{C}^N$, $X' \subset \mathbb{C}^{N'}$ be analytic sets. A map $F : X \to X'$ is said to be a *holomorphic map* if for each $p \in X$ there exists a neighbourhood U of p in $\mathbb{C}^N$

such that the restriction of F to $U \cap X$ is extended to a holomorphic map from U to $\mathbb{C}^{N'}$.

A *germ of analytic subset* at $0 \in \mathbb{C}^N$ is an equivalence class of analytic subsets X containing 0, where X is equivalent to X' if there exists a neighbourhood $\mathcal{V}$ of 0 such that $\mathcal{V} \cap X = \mathcal{V} \cap X'$.

To study the problem of moduli, we need to consider the case when $\mathfrak{I}_{X,p}$ does not satisfy (8.20), i.e. a complex analytic analogue of a scheme. The simplest example is $X = \{0\} \subset \mathbb{C}$ and $\mathfrak{I}_{X,p} = (x^2)$, that is the set of holomorphic functions $f(x)$ such that $f(0) = f'(0) = 0$. Let us define such objects. We need only its germ at 0 so we restrict ourselves to such a case.

Definition 8.2.17. A *germ at* 0 *of analytic subspace* $\mathfrak{X}$ of $\mathbb{C}^N$ is a germ of the analytic subset X together with ideals $\mathfrak{I}_{\mathfrak{X},0} \subset \mathcal{O}_{X,0}$ such that the following holds. Let $f_1, \dots, f_m$ be a generator of $\mathfrak{I}_{\mathfrak{X},0}$. (Since $\mathcal{O}_0$ is Noetherian it follows that we can choose such a generator.) Let f_i be defined on $\mathcal{U}$. Then

$$\mathcal{U} \cap X = \{x \mid f_i(x) = 0, i = 1, 2, \dots, m\}. \tag{8.21}$$

Remark 8.2.2. By Hilbert's nullstellensatz, (8.21) is equivalent to $\mathfrak{I}_{X,0} = \{f \mid f^n \in \mathfrak{I}_{\mathfrak{X},0}$ for some $n.\}$

One can define an analytic variety as a ringed space[8] which is locally isomorphic to $(\mathfrak{X}, \mathfrak{I}_{\mathfrak{X},0})$, in other words as a space obtained by gluing the germs of the analytic subspaces in $\mathbb{C}^N$ defined in definition 8.2.17. We do not try to do so since we do not use it (see [20]).

Example 8.2.3. Let $F = (f^1, \dots, f^k) : \mathcal{U} \to \mathbb{C}^k$ be a holomorphic map, where $\mathcal{U}$ is an open neighbourhood of 0 in $\mathbb{C}^N$. We assume $f^i(0) = 0$. We put $X = F^{-1}(0)$. We let $\mathfrak{I}_{\mathfrak{X},0}$ be the ideal generated by $f^1, \dots, f^k$. We then obtain a germ of the analytic subspace. We denote it by $F^{-1}(0)$ by abuse of notation.

We put $\mathcal{O}_{\mathfrak{X},0} = \mathcal{O}_0/\mathfrak{I}_{\mathfrak{X},0}$. $\mathcal{O}_{\mathfrak{X},0,+} = \{[f] \in \mathcal{O}_{\mathfrak{X},0} \mid f(0) = 0\}$.

Definition 8.2.18. Let $\mathfrak{X} = (X, \mathfrak{I}_{\mathfrak{X},0})$, $(\mathfrak{X}', \mathfrak{I}'_0(\mathfrak{X}'))$ be germs of analytic varieties. A morphism $\mathfrak{F}$ from $\mathfrak{X}$ to $\mathfrak{X}'$ is a ring homomorphism $F^* : \mathcal{O}_{\mathfrak{X}',0} \to \mathcal{O}_{\mathfrak{X},0}$.

(Here and hereafter all ring homomorphisms between commutative rings are assumed to preserve the unit.) We note that the morphism of a germ of analytic subspaces induces a map between analytic sets as follows.

Lemma 8.2.5. *Let* $\mathfrak{F} : \mathfrak{X} \to \mathfrak{X}'$ *be a morphism as in definition* 8.2.18. *Here* $\mathfrak{X} = (X, \mathfrak{I}_{\mathfrak{X},0})$, $X \subset \mathbb{C}^N$, $\mathfrak{X}' = (X', \mathfrak{I}'_{\mathfrak{X}',0})$, $X' \subset \mathbb{C}^{N'}$. *Then there exists a neighbourhood* $\mathcal{U}$ *of* 0 *in* $\mathbb{C}^N$ *and a holomorphic map* $\tilde{F} : \mathcal{U} \to \mathbb{C}^{N'}$ *with* $\tilde{F}(0) = 0$, *such that:*

[8] We do not define this notion here since we do not use it. See [48].

(1) *$\tilde{F}(X \cap \mathcal{U}) \subset X'$;*
(2) *if $f \in \mathfrak{I}'\mathfrak{I}'_{\mathfrak{X}',0}$ then $f \circ F \in \mathfrak{I}_{\mathfrak{X},0}$; and*
(3) *by (2) we have a ring homomorphism $\mathcal{O}_{\mathfrak{X}',0} \to \mathcal{O}_{\mathfrak{X},0}$ induced by $f \mapsto f \circ \tilde{F}$. This homomorphism coincides with F^*.*

Proof. Let x^i, $i = 1, \dots, N'$ be the coordinate function on $\mathbb{C}^{N'}$. We have $F^*x_i \in \mathcal{O}_{\mathfrak{X},0}$. Let $\tilde{f}^i \in \mathcal{O}_0$ be any elements which represent F^*x_i. It is easy to see that $\tilde{F} = (\tilde{f}^1, \dots, \tilde{f}^{N'})$ has the required properties. □

To define a deformation of complex structures parametrized by a germ of an analytic subspace of $\mathbb{C}^N$, we need to define a fibre bundle over a complex analytic variety etc. It is a straightforward analogue of the case for a complex manifold but since our main purpose is to study vector bundles, we restrict ourselves to the deformation of holomorphic structures of a complex vector bundle on a complex manifold M with a *fixed* complex structure.

Let $\mathcal{E}$ be a holomorphic vector bundle on M. Let $\mathcal{U} \subset \mathbb{C}^N$ be an open neighbourhood of the origin and let $X \subset \mathcal{U}$ be a germ of a complex analytic subset.

Definition 8.2.19. A *deformation* of $\mathcal{E}$ parametrized by X is a germ of a holomorphic map $B : \mathcal{U} \to \Omega^{0,1}(M; \mathrm{End}(E))$ such that $B(0) = 0$ and that

$$\overline{\partial}_{\mathcal{E}} B(x) + B(x) \circ B(x) = 0 \tag{8.22}$$

holds for each $x \in X$.

Using (8.20), it is easy to see that (8.22) is equivalent to the following.

There exists an open covering $\cup U_i = M$, and an open neighbourhood $\mathcal{U}$ of 0, and there exists a smooth section $e_i(x, q)$ of $\Lambda^{0,1}(\mathcal{U} \times U_i, \mathrm{End}(E))$ such that

$$\overline{\partial}_{\mathcal{E}} B(x) + B(x) \circ B(x) = \sum f_i(x) e_i(x, q). \tag{8.23}$$

where $f_i \in \mathfrak{I}_{X,0}$. Hereafter we say that (8.23) holds locally and do not mention $U_i, \mathcal{U}$.

In view of (8.23), we can generalize definition 8.2.19 to the case of a germ of an analytic subspace as follows.

Definition 8.2.20. Let $\mathfrak{X} = (X, \mathfrak{I}_{\mathfrak{X},0})$ be a germ of an analytic subspace and $B : \mathcal{U} \to \Omega^{0,1}(M; \mathrm{End}(E))$ be a deformation of the holomorphic structures on E parametrized by X. We say that B is a *deformation* of $\mathcal{E}$ parametrized by $\mathfrak{X}$ if

$$\overline{\partial}_{\mathcal{E}} B(x) + B(x) \circ B(x) = \sum f_i(x) e_i(q, x) \tag{8.24}$$

holds locally with $f_i \in \mathfrak{I}_{\mathfrak{X},0}$.

We say that B is the *same* deformation as B' if

$$B(x) - B'(x) = \sum f_i(x) e_i(q, x)$$

holds locally with $f_i \in \mathfrak{I}_{\mathfrak{X},0}$.

Example 8.2.4. Let $X = \{0\} \subset \mathbb{C}$ and $\mathfrak{I}_{\mathfrak{X},0} = (x^2)$. Let us write $B(x) = xB_1 + x^2B_2 + \cdots$. Then (8.24) implies $\overline{\partial}_{\mathcal{E}} B_1 = 0$. Note any choice of $B_2, \ldots$ defines the same family by definition. Hence the set of deformations parametrized by $(\{0\}, (x^2))$ is identified with the kernel of $\overline{\partial}_{\mathcal{E}} : \Omega^{0,1}(\mathrm{End}(E)) \to \Omega^{0,2}(\mathrm{End}(E))$.

To define completeness and the universality of a deformation parametrized by a germ of an analytic subspace, we define a morphism between deformations. Let $\mathfrak{X} = (X, \mathfrak{I}_{\mathfrak{X},0})$, $\mathfrak{X}' = (X', \mathfrak{I}_{\mathfrak{X}',0})$ be germs of analytic subspaces and $B : \mathcal{U} \to \Omega^{0,1}(M; \mathrm{End}(E))$, $B' : \mathcal{U}' \to \Omega^{0,1}(M; \mathrm{End}(E))$ be deformations of $\mathcal{E}$ parametrized by $\mathfrak{X}$, $\mathfrak{X}'$ respectively.

Definition 8.2.21. A morphism from $(\mathfrak{X}, B)$ to $(\mathfrak{X}', B')$ is a pair $(\mathfrak{F}, \Psi)$, where $\mathfrak{F} : \mathfrak{X} \to \mathfrak{X}'$ is a morphism of an analytic subspace and $\Psi : \mathcal{U} \to \Gamma(M; \mathrm{End}(E))$ is a holomorphic map such that

$$\overline{\partial}_{\mathcal{E}}(\Psi(x)) + B'(\tilde{F}(x)) \circ \Psi(x) - \Psi(x) \circ B(x) = \sum f_i(x) e_i(q, x) \qquad (8.25)$$

holds locally with $f_i \in \mathfrak{I}_{\mathfrak{X},0}$. Here $\tilde{F}$ is as in lemma 8.2.5.

We regard $(\mathfrak{F}, \Psi)$ as the same morphism as $(\mathfrak{F}', \Psi')$ if $\mathfrak{F} = \mathfrak{F}'$ in the sense of definition 8.2.18 and if

$$\Psi(x) - \Psi'(x) = \sum f_i(x) e_i(q, x)$$

locally with $f_i \in \mathfrak{I}_{\mathfrak{X},0}$. Here $e_i(q, x)$ is a local section of $\Lambda^{0,1} \otimes \mathrm{End}(E)$.

Example 8.2.5. Let us consider a family on $\mathfrak{X} = (\{0\}, (x^2))$ as in example 8.2.4. Let $B = xB_1$ and $B' = xB_1'$ be two such families. Let $(\mathfrak{F}, \Psi)$ be a morphism from $(\mathfrak{X}, B)$ to $(\mathfrak{X}, B')$. We may choose $\tilde{F}$ in lemma 8.2.5 so that $\tilde{F}(x) = xF_1$, when $F_1 \in \mathbb{C}$. We may also write $\Psi = \Psi_0 + x\Psi_1$. Equation (8.25) can be written as $\overline{\partial}_{\mathcal{E}}\Psi_0 = 0$, $F_1\overline{\partial}_{\mathcal{E}}\Psi_1 + B_1' \circ \Psi_0 - \Psi_0 \circ B_1 = 0$. Hence $B_1' - B_1 = F_1\overline{\partial}_{\mathcal{E}}\Psi_1$. Thus the set of the deformations of $\mathcal{E}$ parametrized by $\mathfrak{X} = (\{0\}, (x^2))$ is identified with $\mathrm{Ext}^1(\mathcal{E}, \mathcal{E})$.

Now that the morphism of a deformation has been defined, we can define completeness and universality in exactly the same way as for deformations parametrized by complex manifolds. We leave this to the reader.

To define versality and the KS map we need to define the Zariski tangent space of a germ of an analytic subspace.

Definition 8.2.22. Let $\mathfrak{X} \subset \mathbb{C}^N$ be a germ of an analytic subspace. The *Zariski tangent space* $T_0\mathfrak{X}$ is defined by

$$T_0\mathfrak{X} = \{V \in T_p\mathbb{C}^N \,|\, V(f) = 0 \text{ if } f \in \mathfrak{I}_{\mathfrak{X},0}\}.$$

Let $\mathfrak{F} : \mathfrak{X} \to \mathfrak{X}'$ be a morphism. Let $\tilde{F} : \mathcal{U} \to \mathbb{C}^{N'}$ be as in lemma 8.2.5. Then it is easy to see that $\mathrm{d}_0\tilde{F} : T_0\mathbb{C}^N \to T_0\mathbb{C}^{N'}$ induces $\mathrm{d}_0\mathfrak{F} : T_0\mathfrak{X} \to T_0\mathfrak{X}'$.

Now we can generalize the definition of versality in the same way as in section 8.1.2 by using the Zariski tangent space in the same way as in definition 8.2.10.

We next generalize the KS map. Let us consider the situation of definition 8.2.19. Let f_i, e_i be as in (8.23). Let $V = \sum V^i \frac{\partial}{\partial x^i} \in T_0\mathfrak{X}$. By differentiating (8.23) we have

$$\begin{aligned} &\overline{\partial}_{\mathcal{E}} V(B(x)) + V(B(x)) \circ B(x) + B(x) \circ V(B(x)) \\ &\quad = \sum V(f_i)(0)e_i(0) + f_i(0)V(e_i)(0) = 0. \end{aligned}$$

It follows that

$$V(B(x)) \in \mathrm{Ker}(\overline{\partial}_{\mathcal{E}} : \Omega^{0,1}(M; \mathrm{End}(\mathcal{E})) \to \Omega^{0,2}(M; \mathrm{End}(\mathcal{E}))).$$

Definition 8.2.23. We put

$$\mathrm{KS}(V) = [V(B(x))] \in \mathrm{Ext}^1(\mathcal{E}, \mathcal{E}).$$

KS is a linear map: $T_0\mathfrak{X} \to \mathrm{Ext}^1(\mathcal{E}, \mathcal{E})$, which we call the *KS map* of our family.

Lemmata 8.2.1 and 8.2.2 also hold in our case. However, theorem 8.2.2 does *not* hold. In fact, we have the following counter example. Let us consider a universal family parametrized by an open neighbourhood of 0 in $\mathbb{C}^2$. We restrict it to $X = \{(x, y)|xy = 0\}$. Since $T_0X = \mathbb{C}^2$ it follows that this family still satisfies the assumption of theorem 8.2.2. It is not complete however.

Now we have the following generalization of theorem 8.2.3.

Theorem 8.2.6. *Let $\mathcal{E}$ be a holomorphic vector bundle on M. There exists a germ of complex analytic variety $\mathfrak{X}$ and a deformation B of $\mathcal{E}$ parametrized by $\mathfrak{X}$ such that* (1) *B is complete; and* (2) *the KS map,* $\mathrm{KS} : T_0\mathfrak{X} \to \mathrm{Ext}^1(\mathcal{E}, \mathcal{E})$ *is an isomorphism.*

Moreover there exists an open neighbourhood $\mathcal{U}$ of the origin in $\mathrm{Ext}^1(\mathcal{E}, \mathcal{E})$ *and a holomorphic map* $\mathrm{Kura} : \mathcal{U} \to \mathrm{Ext}^2(\mathcal{E}, \mathcal{E})$ *such that* $\mathfrak{X} = \mathrm{Kura}^{-1}(0)$. *Here the* $\mathrm{Kura}^{-1}(0)$ *is as in example* 8.2.3.

Definition 8.2.24. We call the map Kura the *Kuranishi map*.

We will give a proof of theorem 8.2.6 in section 8.2.3. Theorem 8.2.6 is a vector bundle analogue of [71], the work in which has been generalized to various other situations by many people (see, for example, [4, 45, 84, 92, 104]). Theorem 8.2.6 also has an analogue in the case of the moduli space of the gauge equivalence classes of Yang–Mills equations [16, 101].

Proposition 8.2.1 can be generalized directly to our situation by the same proof, i.e. the deformation obtained in theorem 8.2.6 is versal. Theorem 8.2.2 still holds in the case of a deformation parametrized by a germ of an analytic subspace. (We prove it in section 8.2.3.)

We close this section by giving an example where the Kuranishi map is non-zero.

Example 8.2.6. Let $\mathbb{T}$ be an elliptic curve and $\mathcal{L} \to \mathbb{T}$ be a line bundle with $c^1(\mathcal{L}) \cap [\mathbb{T}] = -m < 0$. We have

$$\mathrm{Ext}^k(\mathbb{T}; \mathbb{C}, \mathcal{L}) \cong \begin{cases} \mathbb{C}^m & k = 1 \\ 0 & k = 0. \end{cases}$$

Let $\mathfrak{b}_i : i = 1, \ldots, m$ be a generator of $\mathrm{Ext}^1(\mathbb{T}; \mathbb{C}, \mathcal{L})$ which we may regard as a $\mathcal{L}$ valued $(0, 1)$-form.

By Serre duality, we have $\mathrm{Ext}^0(\mathbb{T}; \mathcal{L}, \mathbb{C}) \cong \mathrm{Ext}^1(\mathbb{T}; \mathbb{C}, \mathcal{L})^*$. Let $\mathfrak{b}^i : i = 1, \ldots, m$ be the dual basis.

We consider $M = \mathbb{T} \times \mathbb{T}$ and let $\mathrm{pr}_i : i = 1, 2$ be projections to first and second factors. We put

$$\mathcal{E} = \mathrm{pr}_1^* \mathcal{L} \oplus \mathrm{pr}_2^* \mathcal{L} = (\mathcal{L} \boxtimes 1) \oplus (1 \boxtimes \mathcal{L})$$

where $\boxtimes$ is the exterior product and 1 is the trivial line bundle. We consider

$$\begin{aligned} \mathrm{Ext}^1(M; \mathcal{E}, \mathcal{E}) \cong & \mathrm{Ext}^1(M; \mathcal{L} \boxtimes 1, \mathcal{L} \boxtimes 1) \oplus \mathrm{Ext}^1(M; 1 \boxtimes \mathcal{L}, 1 \boxtimes \mathcal{L}) \\ & \oplus \mathrm{Ext}^1(M; 1 \boxtimes \mathcal{L}, \mathcal{L} \boxtimes 1) \oplus \mathrm{Ext}^1(M; \mathcal{L} \boxtimes 1, 1 \boxtimes \mathcal{L}). \end{aligned}$$

The first two factors are isomorphic to $H^{0,1}(M) \cong \mathbb{C}^2$. The third and fourth factors are isomorphic to $\mathbb{C}^{m^2}$ and their generators are $\mathrm{pr}_1^* \mathfrak{b}_i \wedge \mathrm{pr}_2^* \mathfrak{b}^j$ and $\mathrm{pr}_1^* \mathfrak{b}^i \wedge \mathrm{pr}_2^* \mathfrak{b}_j$, respectively.

Using these bases we define coordinate a_i, b_i, x_{ij}, y_{ij} of $\mathrm{Ext}^1(M; \mathcal{E}, \mathcal{E})$ as follows. Let $B \in \mathrm{Ext}^1(M; \mathcal{E}, \mathcal{E})$, then we put

$$\overline{\partial}_{\mathcal{E}+B} = \overline{\partial}_{\mathcal{E}} + \begin{pmatrix} \sum_i a_i \, d\overline{z}^i & \sum_{i,j} x_{ij} (\mathrm{pr}_1^* \mathfrak{b}_i \wedge \mathrm{pr}_2^* \mathfrak{b}^j) \\ \sum_{i,j} y_{ij} (\mathrm{pr}_1^* \mathfrak{b}^i \wedge \mathrm{pr}_2^* \mathfrak{b}_j) & \sum b_i \, d\overline{z}^i. \end{pmatrix}$$

where z_1, z_2 are complex coordinates of the first and the second factor $M = \mathbb{T} \times \mathbb{T}$. We also have

$$\mathrm{Ext}^2(M; \mathcal{E}, \mathcal{E}) \cong \mathrm{Ext}^2(M; \mathcal{L} \boxtimes 1, \mathcal{L} \boxtimes 1) \oplus \mathrm{Ext}^2(M; 1 \boxtimes \mathcal{L}, 1 \boxtimes \mathcal{L}) \cong \mathbb{C}^2.$$

We find

$$(\mathrm{pr}_1^* \mathfrak{b}_i \wedge \mathrm{pr}_2^* \mathfrak{b}^j) \wedge (\mathrm{pr}_1^* \mathfrak{b}^k \wedge \mathrm{pr}_2^* \mathfrak{b}_\ell) = \begin{cases} (1, 0) & (i, j) = (k, \ell) \\ 0 & \text{otherwise} \end{cases}$$

$$(\mathrm{pr}_1^* \mathfrak{b}^i \wedge \mathrm{pr}_2^* \mathfrak{b}_j) \wedge (\mathrm{pr}_1^* \mathfrak{b}_k \wedge \mathrm{pr}_2^* \mathfrak{b}^\ell) = \begin{cases} (-1, 0) & (i, j) = (k, \ell) \\ 0 & \text{otherwise.} \end{cases}$$

Other components of the products $\mathrm{Ext}^1(M; \mathcal{E}, \mathcal{E}) \times \mathrm{Ext}^1(M; \mathcal{E}, \mathcal{E}) \to \mathrm{Ext}^2(M; \mathcal{E}, \mathcal{E})$ vanish. Therefore, the Kuranishi map is

$$(a_1, a_2; b_1, b_2; (x_{i,j}), (y_{i,j})) \mapsto \left(\sum_{i,j} x_{i,j} y_{i,j}, -\sum_{i,j} x_{i,j} y_{i,j} \right).$$

Hence the origin is a singular point of $\mathrm{Kura}^{-1}(0)$.

We note that the origin is not a stable bundle in the sense of [78]. We need to consider the three-dimensional case to obtain a stable bundle whose Kuranishi map is non-trivial. Thomas [102] found such an example for the Calabi–Yau threefold. We will discuss a mirror of example 8.2.6 in section 8.3.6.

8.2.7 Formal deformations

In section 8.1.4, we *proved* the convergence of the series (8.15). In a less classical situation, which we will study in later chapters, the convergence of a similar series has not yet been proved. So until the convergence has been proved (in the future hopefully) we need to regard a series like (8.18) as a formal power series. This leads us to the formal deformation theory for this developed in algebraic geometry; [4, 45, 92] seem to be standard references. We review formal deformation in this section.

We need to translate the notion of the deformation of a family of structures into more algebraic language. Let R be a commutative ring with unit. Here we consider the case in which R is a field. We consider a local ring $\mathcal{R}$ such that $R \cong \mathcal{R}/\mathcal{R}_+$. We assume that there exists an embedding $R \to \mathcal{R}$ preserving unit. Hence the composition $R \to \mathcal{R} \to \mathcal{R}/\mathcal{R}_+ \cong R$ is the identity. Let $(A, \cdot, \mathrm{d})$ be a differential graded algebra defined over R. Hereafter we assume A is free as the R module. We simply write A in place of $(A, \cdot, \mathrm{d})$ sometimes for simplicity.

Definition 8.2.25. A *deformation* of A over $\mathcal{R}$ is a pair $(A_{\mathcal{R}}, i)$ where $A_{\mathcal{R}}$ is a differential graded $\mathcal{R}$ algebra and $i : A_{\mathcal{R}}/\mathcal{R}_+ A_{\mathcal{R}} \cong A$, is an isomorphism of differential graded R algebras.

In a similar way, we can define the deformation of a differential graded module as follows. Let $(C, \cdot, \mathrm{d})$ be a differential graded module over $(A, \cdot, \mathrm{d})$. We sometimes write C in place of $(C, \cdot, \mathrm{d})$.

Definition 8.2.26. A *deformation* of C over $A_{\mathcal{R}}$ is a pair of $(C_{\mathcal{R}}, i)$ such that $C_{\mathcal{R}}$ is a differential graded $A_{\mathcal{R}}$ module and $i : C_{\mathcal{R}}/\mathcal{R}_+ C_{\mathcal{R}} \cong C$ is an isomorphism of differential graded A modules.

When $A_{\mathcal{R}} = A \otimes_R \mathcal{R}$, (that is a trivial deformation of A), we say $C_{\mathcal{R}}$ is a deformation of C over $\mathcal{R}$.

Let us explain the relation between definitions 8.2.25 and 8.2.26 and the definitions in section 8.1.2. Let $\hat{M} \to \mathcal{U}$ be a family of complex structures on

M. We consider a vector bundle $\Lambda^{p,q}(\hat{M}/\mathcal{U})$ whose fibre at $\hat{x} \in \hat{M}$ is $\Lambda^{p,q}_{\hat{x}}(M_x)$ where $x = \pi(\hat{x}) \in \mathcal{U}$.

Definition 8.2.27. A section ω of $\Lambda^{p,q}(\hat{M}/\mathcal{U})$ is said to be *holomorphic in the $\mathcal{U}$ direction* if the following holds. Let $\hat{x} \in \hat{M}$. We choose a complex coordinate $w^1, \ldots, w^N$ of a neighbourhood of $x = \pi(\hat{x}) \in \mathcal{U}$. We choose $z^1, \ldots, z^n$ so that $z^1, \ldots, z^n, w^1, \ldots, w^N$ is a complex coordinate of a neighbourhood of $\hat{x}$. (Here we identify w^i with $w^i \circ \pi$.) Now we may write

$$\omega = \sum \omega_{i_1,\ldots,i_p,j_1,\ldots,j_q}(z, w)\, \mathrm{d}z^{i_1} \wedge \cdots \wedge \mathrm{d}z^{i_p} \wedge \mathrm{d}\overline{z}^{j_1} \wedge \cdots \wedge \mathrm{d}\overline{z}^{j_q}. \qquad (8.26)$$

Now we say that ω is holomorphic in the base direction if $\omega_{i_1,\ldots,i_p,j_1,\ldots,j_q}(z, w)$ is holomorphic with respect to w. We assume ω to be smooth in both directions. We denote by $\Omega^{p,q}(\hat{M}/\mathcal{U})$ the set of fiberwise (p, q) forms which are holomorphic in the base direction.

We note that the holomorphicity in the base direction is independent of the choice of coordinate z, w. It is also easy to see that $\Omega^{p,q}(\hat{M}/\mathcal{U})$ is a module over $\mathcal{O}(\mathcal{U})$, the ring of holomorphic functions on $\mathcal{U}$.

We define an operator $\overline{\partial} : \Omega^{p,q}(\hat{M}/\mathcal{U}) \to \Omega^{p,q+1}(\hat{M}/\mathcal{U})$ by

$$\begin{aligned}&\overline{\partial}(\omega_{i_1,\ldots,i_p,j_1,\ldots,j_q}(z, w)\, \mathrm{d}z^{i_1} \wedge \cdots \wedge \mathrm{d}z^{i_p} \wedge \mathrm{d}\overline{z}^{j_1} \wedge \cdots \wedge \mathrm{d}\overline{z}^{j_q}) \\ &\quad = \sum_{\ell} (-1)^p \frac{\partial \omega_{i_1,\ldots,i_p,j_1,\ldots,j_q}}{\partial \overline{z}^{\ell}}\, \mathrm{d}z^{i_1} \wedge \cdots \wedge \mathrm{d}z^{i_p} \wedge \mathrm{d}\overline{z}^{\ell} \wedge \mathrm{d}\overline{z}^{j_1} \wedge \cdots \wedge \mathrm{d}\overline{z}^{j_q}.\end{aligned}$$

We can define a wedge product $\wedge$ between elements of $\Omega^{*,*}(\hat{M}/\mathcal{U})$ in an obvious way. Thus $(\Omega^{0,*}(\hat{M}/\mathcal{U}), \overline{\partial}, \wedge)$ is a differential graded algebra over $\mathcal{O}(\mathcal{U})$. We consider its germ as follows. Let $\mathcal{V} \subset \mathcal{U}$ be open neighbourhoods of 0. We put $\hat{M}(\mathcal{V}) = \pi^{-1}(\mathcal{V})$. We consider pairs $(\omega, \mathcal{V})$ where $\omega \in \Omega^{p,q}(\hat{M}(\mathcal{V})/\mathcal{V})$. We say $(\omega, \mathcal{V}) \sim (\omega', \mathcal{V}')$ if $\omega = \omega'$ on $\hat{M}(\mathcal{V} \cap \mathcal{V}')$. The set of $\sim$ equivalence classes of all such pairs is denoted by $\Omega^{p,q}(\hat{M}/\mathcal{U})_0$. It is obvious that $\Omega^{p,q}(\hat{M}/\mathcal{U})_0$ is an $\mathcal{O}_0$ module. It is easy to see that

$$\frac{\Omega^{p,q}(\hat{M}/\mathcal{U})_0}{\mathcal{O}_{0,+}\Omega^{p,q}(\hat{M}/\mathcal{U})_0} \cong \Omega^{p,q}(M).$$

We write the isomorphism by i. The following lemma is now obvious.

Lemma 8.2.6. *$((\Omega^{0,*}(\hat{M}/\mathcal{U})_0, \overline{\partial}, \wedge), i)$ is a deformation of $(\Omega^{0,*}(M), \overline{\partial}_0, \wedge)$ over $\mathcal{O}_0$.*

Remark 8.2.3. In fact, to study the deformation of complex structures, we need to study the deformation of the differential graded Lie algebra $\Gamma(M; \Lambda^{0,*} \otimes TM)$ (defined in example 8.3.1), as discussed in [61,63,71], instead of the deformation of differential graded Lie algebra $\Omega^{p,q}(M)$.

Let $(B, \mathcal{U})$ be a deformation of a holomorphic vector bundle $\mathcal{E}$ on $\hat{M}$. In a similar way, we can define $\Omega^{p,q}(\hat{M}/\mathcal{U};\, E)$ the set of all E valued p, q forms which is holomorphic with respect to the base direction ($\hat{M} = M \times \mathcal{U}$). We can also define its germ $\Omega^{p,q}(\hat{M}/\mathcal{U};\, E)_0$. We define $\overline{\partial}_{\mathcal{E}+B} : \Omega^{p,q}(\hat{M}/\mathcal{U};\, E)_0 \to \Omega^{p,q+1}(\hat{M}/\mathcal{U};\, E)_0$ and module structure $\wedge : \Omega^{p,q}(\hat{M}/\mathcal{U})_0 \otimes \Omega^{p',q'}(\hat{M}/\mathcal{U};\, E)_0 \to \Omega^{p+p',q+q'}(\hat{M}/\mathcal{U};\, E)_0$. Thus we have the following lemma.

Lemma 8.2.7. $((\Omega^{0,*}(\hat{M}/\mathcal{U};\, E)_0, \overline{\partial}_{\mathcal{E}+B}, \wedge), i)$, *defined earlier is a deformation of* $(\Omega^{0,*}(M;\, E), \overline{\partial}_{\mathcal{E}}, \wedge)$ *over* $((\Omega^{0,*}(\hat{M}/\mathcal{U})_0, \overline{\partial}, \wedge), i)$.

The generalization to the case of families parametrized by the germ of an analytic subspace as in section 8.1.6 is straightforward. When we do not move the complex structure of M, it is described as follows.

Let $\mathfrak{X} = (X, \mathfrak{I}_{\mathfrak{X},0})$ be a germ of an analytic subspace ($X \subseteq \mathcal{U}$). We put $\hat{M} = M \times X$. Let $B : \mathcal{U} \to \Omega^{0,1}(M; \mathrm{End}(E))$ be as in definition 8.2.20. We then have $\overline{\partial}_{\mathcal{E}+B} : \Omega^{p,q}(\hat{M}/\mathcal{U};\, E)_0 \to \Omega^{p,q+1}(\hat{M}/\mathcal{U};\, E)_0$. By definition 8.2.20 we have $(\overline{\partial}_{\mathcal{E}+B} \circ \overline{\partial}_{\mathcal{E}+B})(\omega) = \sum f_i e_i$ where $f_i \in \mathfrak{I}_{\mathfrak{X},0}$. We now put

$$\Omega^{p,q}(\hat{M}/\mathfrak{X};\, E)_0 = \frac{\Omega^{p,q}(\hat{M}/\mathcal{U};\, E)_0}{\mathfrak{I}_{\mathfrak{X},0}\Omega^{p,q}(\hat{M}/\mathcal{U};\, E)_0}.$$

Then $\overline{\partial}_{\mathcal{E}+B}$ induces a homomorphism $\overline{\partial}_{\mathcal{E}+B} : \Omega^{p,q}(\hat{M}/\mathfrak{X};\, E)_0 \to \Omega^{p,q+1}(\hat{M}/\mathfrak{X};\, E)_0$ such that $\overline{\partial}_{\mathcal{E}+B} \circ \overline{\partial}_{\mathcal{E}+B} = 0$. We thus have the following lemma.

Lemma 8.2.8. $((\Omega^{0,*}(\hat{M}/\mathfrak{X};\, E)_0, \overline{\partial}_{\mathcal{E}+B}, \wedge), i)$, *defined earlier, is a deformation of* $(\Omega^{0,*}(M;\, E), \overline{\partial}_{\mathcal{E}}, \wedge)$ *over* $((\Omega^{0,*}(M) \otimes \mathcal{O}_{\mathfrak{X},0}, \overline{\partial}, \wedge), i)$.

Now we will study formal deformation. This means that we are going to study a formal power series ring rather than a convergent power series ring. We first briefly review the formal power series ring and projective limit. We consider the convergent power series ring $\mathcal{O}_{\mathfrak{X},0}$. It is a local ring and its maximal ideal is $\mathcal{O}_{\mathfrak{X},0,+}$. We consider the quotient ring $\mathcal{O}_{\mathfrak{X},0}/\mathcal{O}^m_{\mathfrak{X},0,+}$. It is finite dimensional as a vector space over $\mathbb{C}$. We put

$$\hat{\mathcal{O}}_{\mathfrak{X},0} = \varprojlim \mathcal{O}_{\mathfrak{X},0}/\mathcal{O}^m_{\mathfrak{X},0,+}. \tag{8.27}$$

Here the right-hand side is the projective limit. Let us recall its definition for the convenience of the reader. Note that an obvious homomorphism $\pi_{m,m'} : \mathcal{O}_{\mathfrak{X},0}/\mathcal{O}^m_{\mathfrak{X},0,+} \to \mathcal{O}_{\mathfrak{X},0}/\mathcal{O}^{m'}_{\mathfrak{X},0,+}$ for $m < m'$ exists.

Definition 8.2.28. Let S_m be the set for each $m \in \mathbb{Z}_{>0}$ and $\pi_{m,m'} : S_m \to S_{m'}$ be the map for $m < m'$ such that $\pi_{m',m''} \circ \pi_{m,m'} = \pi_{m,m''}$. We consider the direct product $\prod S_m$. The *projective limit* $\varprojlim S_m$ is a subset of $\prod S_m$ consisting of $(x_1, x_2, \dots)$ such that $\pi_{m,m'}(x_m) = x_{m'}$.

When the S_m are groups, rings, modules etc and $\pi_{m,m'}$ are homomorphisms, the projective limit $\varprojlim S_m$ is also a group, ring, module etc.

Remark 8.2.4. We have defined the projective limit only for family S_m parametrized by $m \in \mathbb{Z}_{\geq 0}$. The projective limit can be defined for a more general family.

Remark 8.2.5. We define $d : \varprojlim S_m \times \varprojlim S_m \to \mathbb{R}_{\geq 0}$ by

$$d((x_1, x_2, \ldots), (y_1, y_2, \ldots)) = \exp(-\inf\{m | x_m \neq y_m\}).$$

Then $(\varprojlim S_m, d)$ is a *complete* metric space.

Example 8.2.7. Let us consider the case when $\mathfrak{X} = \mathcal{U}$ is an open neighbourhood of 0 in $\mathbb{C}^N$. In this case, $\mathcal{O}_{\mathcal{U},0}$ is a convergent power series ring $\mathbb{C}\{z_1, \ldots, z_N\}$ of N variables. Then $\mathcal{O}_{\mathcal{U},0}/\mathcal{O}^m_{\mathcal{U},0,+} = \mathbb{C}[z_1, \ldots, z_N]/(z_1, \ldots, z_N)^m$ is the set of all polynomials modulo the terms of order $> m$. Now let $(P_1, \ldots, P_m, \ldots) \in \prod \mathcal{O}_{\mathcal{U},P_0,0}/\mathcal{O}^m_{\mathcal{U},P_0,0}$ be an element of the projective limit $\varprojlim \mathcal{O}_{\mathfrak{X},0}/\mathcal{O}^m_{\mathfrak{X},0,+}$. This means that P_m coincides with $P_{m'}$ up to order m. Therefore, $(P_1, \ldots, P_m, \ldots)$ determines a formal power series of z_i; i.e. we have $\hat{\mathcal{O}}_{\mathcal{U},x_0,0} \cong \mathbb{C}[[z_1, \ldots, z_N]]$. Here $\mathbb{C}[[z_1, \ldots, z_N]]$ is a formal power series ring.

In general, if the ideal $\mathfrak{I}_{\mathfrak{X},0}$ is generated by $f_1, \ldots, f_m \in \mathcal{O}_0$ then

$$\hat{\mathcal{O}}_{\mathfrak{X},0} \cong \frac{\mathbb{C}[[z_1, \ldots, z_N]]}{(f_1, \ldots, f_m)}. \tag{8.28}$$

Here we regard $f_i \in \mathbb{C}[[z_1, \ldots, z_N]]$ by taking its Taylor series at 0 and $(f_1, \ldots, f_m)$ is an ideal generated by them.

We note that the ring $\mathcal{O}_{\mathfrak{X},0}/\mathcal{O}^m_{\mathfrak{X},+}$ in (8.27) is of finite dimension over $\mathbb{C}$ (as a vector space). (In the case of a general local ring $\mathcal{R}$ over R, the ring $\mathcal{R}/\mathcal{R}^m_+$ is an Artin algebra over R.) In other words, a formal power series ring can be regarded as a projective limit of finite dimensional $\mathbb{C}$ algebras (or of Artin R algebras). This is a reason why the theory of formal deformation is based on 'a functor from an Artin ring' [92], which we review here.

To make the exposition elementary, we only consider the case when R is an algebraically closed field. (We usually take $R = \mathbb{C}$.) In this case, we do not need the notion of an Artin R algebra and consider only an R algebra of finite dimension (as a vector space over R).

Definition 8.2.29. We define the category of finite dimensional local R algebra, abbreviated by $\{\text{f. d. Alg.}/R\}$, as follows.

(1) Its object is a local R algebra $\mathcal{R}$ which is commutative with unit and is finite dimensional over R as a vector space.
(2) The morphism $\mathcal{R} \to \mathcal{R}'$ is an R algebra homomorphism.

Definition 8.2.30. A *formal moduli functor* is a covariant functor from {f. d. Alg. $/R$} to {Sets}, the category of sets.

For the reader who is not familiar with category theory, let us review what definitions 8.2.29, 8.2.30, mean. (See [20, Exposé 11] for more detail on the relation of category theory to the theory of moduli.) Let $\mathfrak{F}$ be a formal moduli functor in the sense of definition 8.2.30. It consists of two kinds of data.

One is $\mathfrak{F}_0$ which associates a set $\mathfrak{F}_0(\mathcal{R})$ with any local ring $\mathcal{R}$ which is commutative with unit and is of finite dimension over R.

Let $\mathcal{R}$, $\mathcal{R}'$ be two such rings and let $\varphi : \mathcal{R} \to \mathcal{R}'$ be an R algebra homomorphism. Then the second data $\mathfrak{F}_1$ associates with φ a map $\mathfrak{F}_1(\varphi) : \mathfrak{F}_0(\mathcal{R}) \to \mathfrak{F}_0(\mathcal{R}')$.

The condition for $\mathfrak{F}_0$, $\mathfrak{F}_1$ to define a covariant functor is $\mathfrak{F}_1(\varphi' \circ \varphi) = \mathfrak{F}_1(\varphi') \circ \mathfrak{F}_1(\varphi)$, where $\varphi' : \mathcal{R}' \to \mathcal{R}''$.

Our main example of a formal moduli functor is one such that $\mathfrak{F}_0(\mathcal{R})$ is the set of isomorphism classes of the deformation of a given differential graded module. To define it we first give the following definition.

Definition 8.2.31. Let $(C_{\mathcal{R}}, i)$, $(C'_{\mathcal{R}}, i')$ be deformations of differential graded A module C over $\mathcal{R}$. We say that $(C_{\mathcal{R}}, i)$ is isomorphic to $(C'_{\mathcal{R}}, i')$ if there exists an isomorphism $\Phi : C_{\mathcal{R}} \to C'_{\mathcal{R}}$ of differential graded $A_{\mathcal{R}}$ modules such that the induced isomorphism $\overline{\Phi} : C_{\mathcal{R}}/\mathcal{R}_+C_{\mathcal{R}} \to C'_{\mathcal{R}}/\mathcal{R}_+C'_{\mathcal{R}}$ satisfies $i' \circ \overline{\Phi} = i$.

We now define a functor $\mathfrak{Def}_C$: {f. d. Alg. $/R$} $\to$ {Sets} for each differential graded A module C, where A is a differential graded ring over R. Let $\mathcal{R}$ be an object of {f. d. Alg. $/R$}.

Definition 8.2.32. $\mathfrak{Def}_{C,0}(\mathcal{R})$ is the set of all isomorphism classes of deformations of C over $\mathcal{R}$. Let $\varphi : \mathcal{R} \to \mathcal{R}'$ be a morphism of the category {f. d. Alg. $/R$}. Let $C_{\mathcal{R}}$ be a deformation of C over $\mathcal{R}$. We put $\mathfrak{Def}_{C,1}(\varphi)(C_{\mathcal{R}}) = C_{\mathcal{R}} \otimes_{\mathcal{R}} \mathcal{R}'$.

We note that if $C_{\mathcal{R}}$ is isomorphic to $C'_{\mathcal{R}}$ in the sense of definition 8.2.31 then $C_{\mathcal{R}} \otimes_{\mathcal{R}} \mathcal{R}'$ is isomorphic to $C'_{\mathcal{R}} \otimes_{\mathcal{R}} \mathcal{R}'$. Hence $\mathfrak{Def}_{C,1}(\varphi) : \mathfrak{Def}_{C,0}(\mathcal{R}) \to \mathfrak{Def}_{C,0}(\mathcal{R}')$ is well defined. It is easy to see that $\mathfrak{Def}_C$ is a covariant functor: {f. d. Alg. $/R$} $\to$ {Sets}, i.e. we can check $\mathfrak{Def}_{C,1}(\varphi') \circ \mathfrak{Def}_{C,1}(\varphi) = \mathfrak{Def}_{C,1}(\varphi' \circ \varphi)$.

To show the relation of moduli functor to moduli space, we need to discuss the representability of the functor. In our formal deformation theory, we need a notion of pro-representability, which we define here.

Definition 8.2.33. Let $\mathfrak{R}$ be a local R algebra (commutative with unit). We say that $\mathfrak{R}$ is a *pro* {f. d. Alg. $/R$} *object* if the following hold.

(1) Let us denote by $\mathfrak{R}_+$ the maximal ideal of $\mathfrak{R}$. Then, for each m, the quotient ring $\mathfrak{R}/\mathfrak{R}_+^m$ is an object of {f. d. Alg. $/R$}. (In other words, $\mathfrak{R}/\mathfrak{R}_+^m$ is of finite dimension over R.)

(2) $\mathfrak{R} \cong \varprojlim \mathfrak{R}/\mathfrak{R}_+^m$. In other words, $\mathfrak{R}$ is complete with respect to the $\mathfrak{R}_+$ adic metric (which was defined in remark 8.2.5).

Definition 8.2.34. Let $\mathfrak{R}$ be a pro {f. d. Alg. / R} object. We define a covariant functor $\mathfrak{F}_{\mathfrak{R}}$: {f. d. Alg. / R} $\to$ {Sets} as follows.

(1) For an object $\mathcal{R}$ of {f. d. Alg. / R}, $\mathfrak{F}_{\mathfrak{R},0}(\mathcal{R})$ is the set of all R algebra homomorphisms: $\mathfrak{R} \to \mathcal{R}$.
(2) Let $\varphi : \mathcal{R} \to \mathcal{R}'$ be a morphism in the category {f. d. Alg. / R}. Let $\psi \in \mathfrak{F}_{\mathfrak{R},0}(\mathcal{R})$. Then $(\mathfrak{F}_{\mathfrak{R},1}(\varphi))(\psi) = \varphi \circ \psi$.

If $\mathfrak{R}$ itself is an object of {f. d. Alg. / R} (i.e. $\mathfrak{R}$ is of finite dimension over R), then the functor $\mathfrak{F}_{\mathfrak{R}}$ defined in definition 8.2.34 is the functor represented by $\mathfrak{R}$ in the usual sense of category theory.

Definition 8.2.35. A covariant functor {f. d. Alg. / R} $\to$ {Sets} is said to be *pro-representable* if there exists a pro {f. d. Alg. / R} object $\mathfrak{R}$ such that $\mathfrak{F}$ is equivalent to $\mathfrak{F}_{\mathfrak{R}}$ defined in definition 8.2.34.

We recall that two functors $\mathfrak{F}, \mathfrak{F}'$: {f. d. Alg. / R} $\to$ {Sets} are said to be equivalent to each other if the following holds: For each object $\mathcal{R}$ of {f. d. Alg. / R} there exists a bijection $\mathcal{H}_{\mathcal{R}} : \mathfrak{F}_0(\mathcal{R}) \to \mathfrak{F}'_0(\mathcal{R})$ such that the following diagram commutes for any morphism $\varphi : \mathcal{R} \to \mathcal{R}'$.

$$\begin{array}{ccc}
\mathfrak{F}_0(\mathcal{R}) & \xrightarrow{\mathcal{H}_{\mathcal{R}}} & \mathfrak{F}'_0(\mathcal{R}) \\
{\scriptstyle \mathfrak{F}_1(\varphi)}\big\downarrow & & \big\downarrow{\scriptstyle \mathfrak{F}'_1(\varphi)} \\
\mathfrak{F}_0(\mathcal{R}') & \xrightarrow{\mathcal{H}_{\mathcal{R}'}} & \mathfrak{F}'_0(\mathcal{R}')
\end{array}$$

Diagram 3

We now define the universal formal moduli space of the deformation of a differential graded module. Let A be a differential graded algebra over R and C be a differential graded A module.

Definition 8.2.36. A pro {f. d. Alg. / R} object $\mathfrak{R}$ is said to be a *universal formal moduli space* of the deformation of C if the functor $\mathfrak{F}_{\mathfrak{R},0}(\mathcal{R})$ in definition 8.2.34 is equivalent to the functor $\mathfrak{Def}_C$ in definition 8.2.32.

Remark 8.2.6. It is more precise to say that Spec $\mathfrak{R}$ is a universal formal moduli space rather than to say $\mathfrak{R}$ is a universal moduli space. Here Spec $\mathfrak{R}$ is a formal scheme (see [45]). Since we do not introduce the notion of a formal scheme, we say $\mathfrak{R}$ is a moduli space by abuse of language.

Exercise 8.2.1. Prove that the universal formal moduli space in the sense of definition 8.2.36 is unique if it exists.

Now we consider our geometric situation of the deformation of holomorphic vector bundles. Let $\mathfrak{X} = (X, \mathfrak{I}_{\mathfrak{X},0})$ be an analytic subspace and $B : \mathcal{U} \to \Omega^{0,1}(M; \mathrm{End}(E))$ be a deformation of $\mathcal{E}$ parametrized by $\mathcal{X}$. We define $\hat{\mathcal{O}}_{\mathfrak{X},0}$ by (8.27). $\hat{\mathcal{O}}_{\mathfrak{X},0}$ is a pro {f. d. Alg. /$\mathbb{C}$} object.

Proposition 8.2.2. *If the deformation B of $\mathcal{E}$ is universal, then $\hat{\mathcal{O}}_{\mathfrak{X},0}$ is a universal moduli space of the deformation of* $(\Omega^*(M; \mathcal{E}|_{M\times\{x_0\}}), d, \wedge)$.

In contrast, if there exists a universal moduli space $\mathfrak{R}$ of the deformation of $(\Omega^*(M; \mathcal{E}|_{M\times\{x_0\}}), d, \wedge)$*, then there exists a universal family of holomorphic structures of $\mathfrak{R}$ on $\mathfrak{X}$ such that $\hat{\mathcal{O}}_{\mathfrak{X},0}$ is isomorphic to $\mathfrak{R}$.*

Proof (sketch). First we assume that our family is universal. Let $\mathfrak{R}$ be an object of {f. d. Alg. /$\mathbb{C}$}. We may write $\mathcal{R} = \mathbb{C}[[z_1, \ldots, z_N]]/(f_1, \ldots, f_k)$. f_i is, *a priori*, a formal power series. However, using finite dimensionality of $\mathcal{R}$ we may take polynomials for f_i.

We consider a germ of analytic subspace $\mathfrak{Y} = (\{0\}, (f_1, \ldots, f_k))$. It is easy to see that a morphism $\mathfrak{Y} \to \mathfrak{X}$ is in one-to-one correspondence with the $\mathbb{C}$ algebra homomorphism $\hat{\mathcal{O}}_{\mathfrak{X},0} \to \mathcal{R}$. It is an immediate consequence of lemma 8.2.7 (which still holds in the case of a deformation parametrized by a germ of an analytic subspace) that the deformation of $\mathcal{E}$ parametrized by $\mathfrak{Y}$ is in one-to-one correspondence with the deformation of $(\Omega^*(M; \mathcal{E}|_{M\times\{x_0\}}), d, \wedge)$ over $\mathcal{R}$. We can then prove easily that $\hat{\mathcal{O}}_{\mathfrak{X},0}$ is a universal moduli space of the deformation of $(\Omega^*(M; \mathcal{E}|_{M\times\{x_0\}}), d, \wedge)$.

The proof of the converse is more involved since we need to see the relation between the deformation in the category of formal power series and of convergent power series. We do not attempt it here. □

8.3 Homological algebra and deformation theory

8.3.1 Homotopy theory of A_∞ and L_∞ algebras

Now we are going to discuss the less classical part of the story. We have so far studied the equation

$$\overline{\partial} B + B \circ B = 0 \tag{8.29}$$

in which the second term is second order. We mentioned in section 8.1.5 the possibility of considering an equation with terms of third or higher order. To do so while keeping the gauge invariance of the equation, we need to consider the A_∞ algebra, due to Stasheff [97], which we define in this section. As we consider the deformation of complex manifolds rather than holomorphic vector bundles on it, we need to consider the equation

$$\overline{\partial} B + \tfrac{1}{2}[B, B] = 0 \tag{8.30}$$

in place of (8.29). Here $B \in \Omega^{0,1}(M; TM)$ and $[B, B]$ is a combination of the wedge product in the $\Lambda^{0,1}$ factor and the bracket of the vector field in the TM factor (see example 8.3.1). To generalize (8.30) so that it includes terms of third or higher order we introduce the notion of L_∞ algebras.[9]

To define A_∞ and L_∞ algebras we need to review coalgebras, coderivations etc. Let C be a graded R module. (Here the grading starts from 0.) We define its suspension $C[1]$ by $C[1]^k = C^{k+1}$. Hereafter we denote $\deg x$ as the degree of elements of $x \in C$ and $\deg' x$ is the degree of the same element regarded as an element of $C[1]$; i.e. $\deg' x = \deg x - 1$. We define its *Bar complex* $BC[1]$ by

$$B_k C[1] = C[1]^{k\otimes} \qquad BC[1] = \bigoplus_{k=0}^{\infty} B_k C[1].$$

We note that $B_0 C[1] = R$. We define the action of the group $\mathfrak{S}_k$ of all permutations of k elements on $B_k C[1]$ by

$$\sigma(x_1 \otimes \cdots \otimes x_k) = \pm x_{\sigma(1)} \otimes \cdots \otimes x_{\sigma(k)}$$

where

$$\pm = (-1)^{\sum_{i,j \text{ with } i<j, \sigma(i)>\sigma(j)} \deg' x_i \deg' x_j}. \tag{8.31}$$

We define $E_k C[1]$ to be the submodule consisting of fixed points of the $\mathfrak{S}_k$ action on $B_k C[1]$ and $EC[1] = \oplus E_k C[1]$.

We define $\Delta : BC[1] \to BC[1] \otimes BC[1]$ by

$$\Delta(x_1 \otimes \cdots \otimes x_k) = \sum_{i=0}^{k} (x_1 \otimes \cdots \otimes x_i) \otimes (x_{i+1} \otimes \cdots \otimes x_k). \tag{8.32}$$

Note the term of (8.32) when $i = 0$ becomes $1 \otimes (x_1 \otimes \cdots \otimes x_k) \in B_0 C[1] \otimes B_k C[1]$. The restriction of Δ induces $\Delta : EC[1] \to EC[1] \otimes EC[1]$.

Definition 8.3.1. A graded coalgebra $(D = \oplus D^k, \Delta, \epsilon)$ is a graded R module together with $\Delta : D \to D \otimes D$, $\epsilon : D^0 \to R$ such that the following diagrams commute.

$$\begin{tikzcd}
D \otimes_R D \otimes D & D \otimes D \arrow[l, "\Delta\otimes 1"'] \\
D \otimes D \arrow[u, "1\otimes\Delta"] & D \arrow[l, "\Delta"] \arrow[u, "\Delta"']
\end{tikzcd}
\qquad
\begin{tikzcd}
D & D \otimes D \arrow[l, "\epsilon\otimes 1"'] \arrow[r, "1\otimes\epsilon"] & D \\
D \arrow[u, equal] \arrow[r, equal] & D \arrow[u, "\Delta"] \arrow[r, equal] & D \arrow[u, equal]
\end{tikzcd}$$

Diagram 4 Diagram 5

Coalgebra (D, Δ, ϵ) is said to be *graded cocommutative* if $R \circ \Delta = \Delta : D \to D \otimes D$, where $R(x \otimes y) = (-1)^{\deg x \deg y}(y \otimes x)$.

[9] We note that the A in A_∞ algebras stands for associative and the L in L_∞ algebras for Lie.

The following lemma is easy to check.

Lemma 8.3.1. *$(BC[1], \Delta, \epsilon)$ is a colagebra where ϵ is an obvious isomorphism $B_0C[1] \cong R$. We define a degree of elements of $BC[1]$ by $\deg'(x_1 \otimes \cdots \otimes x_k) = \sum \deg' x_i = \sum \deg x_i - k$. $(EC[1], \Delta, \epsilon)$ is a graded cocommutative coalgebra. (We note that we need to take $\deg'$ and not $\deg$ in the definition of R.)*

Definition 8.3.2. A graded homomorphism $\delta : D \to D$ of degree 1 from a coalgebra D to itself is said to be a *coderivation* if the following diagram commutes.

$$\begin{tikzcd} D \otimes D & D \arrow[l, "\Delta"'] \\ D \otimes D \arrow[u, "1\hat{\otimes}\delta+\delta\hat{\otimes}1"] & D \arrow[l, "\Delta"'] \arrow[u, "\delta"'] \end{tikzcd}$$

Diagram 6.

Here we define the graded tensor product $A\hat{\otimes}B$ between two graded homomorphisms A,B by $(A\hat{\otimes}B)(x \otimes y) = (-1)^{\deg B \deg' x}(A(x)) \otimes (B(y))$.

Lemma 8.3.2. *For any sequence of homomorphisms $f_k : B_kC[1] \to C[1]$ of degree 1 for $k = 1, 2, \ldots,$ there exists a unique coderivation $\delta : BC[1] \to BC[1]$ whose $\mathrm{Hom}(B_kC[1], B_1C[1])$ component is f_k and whose $\mathrm{Hom}(B_0C[1], B_1C[1])$ component is zero. The same holds for $EC[1]$ in place of $BC[1]$.*

Proof. Put

$$\begin{aligned} \hat{f}_k(x_1 \otimes \cdots \otimes x_n) = & \sum_{i=1}^{n-k+1} (-1)^{\deg f_k(\deg' x_1+\cdots+\deg' x_{i-1})} \\ & \times x_1 \otimes \cdots \otimes x_{i-1} \otimes f_k(x_i \otimes \cdots \otimes x_{i+k-1}) \\ & \otimes x_{i+k} \otimes \cdots \otimes x_n \end{aligned} \tag{8.33}$$

and $\delta = \sum \hat{f}_k$. □

To simplify formulae like (8.33) we introduce the following notation. Let D be a coalgebra. We define $\Delta^k : D \to D^{k\otimes}$ by

$$\Delta^k = \underbrace{\cdots (\Delta \otimes 1 \otimes 1) \circ (\Delta \otimes 1) \circ \Delta}_{k-1 \text{ times}}.$$

Then, for an element $\mathbf{x}$ of D, we put

$$\Delta^k(\mathbf{x}) = \sum_a \mathbf{x}_a^{(k;1)} \otimes \cdots \otimes \mathbf{x}_a^{(k;k)}. \tag{8.34}$$

In general, we use bold face letters such as $\mathbf{x}$ for elements of the bar complex $BC[1]$ and roman letters such as x_k for elements of $C[1]$.

Now formula (8.33) can be written:

$$\hat{f}_k(\mathbf{x}) = \sum_a (-1)^{\deg f_k \deg' \mathbf{x}_a^{(3;1)}} \mathbf{x}_a^{(3;1)} \otimes f_k(\mathbf{x}_a^{(3;2)}) \otimes \mathbf{x}_a^{(3;1)}.$$

Now we are ready to define A_∞ and L_∞ algebras.

Definition 8.3.3. The structure of an A_∞ *algebra* on $C[1]$ is a series of R module homomorphisms $\mathfrak{m}_k : B_k C[1] \to C[1]$ $(k = 1, 2, \dots)$ of degree $+1$ such that the coderivation δ obtained by lemma 8.3.2 satisfies $\delta\delta = 0$. If we replace B by E, then it will be the definition of the structure of an L_∞ *algebra* on $C[1]$.

We can write the condition $\delta\delta = 0$ more explicitly as follows.

$$\begin{aligned} \sum_{i=1}^{n-k+1} & (-1)^{\deg' x_1 + \cdots + \deg' x_{i-1}} \\ & \times \mathfrak{m}_{n-k+1}(x_1 \otimes \cdots \otimes x_{i-1} \otimes \mathfrak{m}_k(x_i \otimes \cdots \otimes x_{i+k-1}) \\ & \otimes x_{i+k} \otimes \cdots \otimes x_n) = 0. \end{aligned} \tag{8.35}$$

In particular, we have $\mathfrak{m}_1\mathfrak{m}_1 = 0$. Hence we can define the $\mathfrak{m}_1$ cohomology $H(C, \mathfrak{m}_1)$.

Let $(C, \mathrm{d}, \cdot)$ be a graded differential algebra. We put

$$\mathfrak{m}_1(x) = (-1)^{\deg x} \, \mathrm{d}x \qquad \mathfrak{m}_2(x, y) = (-1)^{\deg x (\deg y + 1)} x \cdot y$$

and $\mathfrak{m}_k = 0$ for $k \geq 3$.

Lemma 8.3.3. $\mathfrak{m}_k$ *determines a structure of* A_∞ *algebra on* C.

Proof. It suffices to show that $\mathrm{Hom}(B_k C[1], C[1])$ component of $\delta\delta$ is zero for $k = 1, 2, 3$. The case $k = 1$ is obvious. Let us check the case $k = 3$, and leave the case $k = 2$ to the reader. Let $\pi_1 : BC[1] \to B_1 C[1] = C[1]$ be the projection. Then we have:

$$\begin{aligned} \pi_1 \delta\delta(x \otimes y \otimes z) &= (-1)^{\deg' x} \mathfrak{m}_2(x, \mathfrak{m}_2(y, z)) + \mathfrak{m}_2(\mathfrak{m}_2(x, y), z) \\ &= (-1)^{\deg x + 1 + \deg y (\deg z + 1) + \deg x (\deg y + \deg z + 1)} x \cdot (y \cdot z) \\ &\quad + (-1)^{\deg x (\deg y + 1) + (\deg x + \deg y)(\deg z + 1)} (x \cdot y) \cdot z = 0. \end{aligned}$$

The last equality follows from the associativity of $\cdot$. (Change of degree and sign is important so that the relation $\delta\delta = 0$ becomes an associativity relation.) □

We next discuss the L_∞ case.

Definition 8.3.4. A *differential graded Lie algebra* is a graded R module C together with operations $[\ ,\ ] : C \otimes C \to C$ of degree 0 and $d : C \to C$ of degree 1 such that $dd = 0$ and

(1) $\mathrm{d}[x, y] = [\mathrm{d}x, y] + (-1)^{\deg x}[x, \mathrm{d}y]$;
(2) $[x, y] = (-1)^{\deg x \deg y+1}[y, x]$; and
(3) $[[x, y], z] + (-1)^{(\deg x+\deg y)\deg z}[[z, x], y] + (-1)^{(\deg y+\deg z)\deg x}[[y, z], x] = 0$.

Example 8.3.1. Let M be a complex manifold and $T_{\mathbb{C}}M$ be a complex tangent bundle (i.e. the holomorphic vector bundle whose local frame is $\frac{\partial}{\partial z^i}$ where z^i, $i = 1, \dots, n$ is a local complex coordinate). Let $C^k = \Gamma^k(M; T_{\mathbb{C}}M \otimes \Lambda^{0,k})$. We put $\mathrm{d} = \overline{\partial}$ and

$$\left[f\frac{\partial}{\partial z^i} \otimes \mathrm{d}\overline{z}^{i_1} \wedge \cdots \wedge \mathrm{d}\overline{z}^{i_k}, g\frac{\partial}{\partial z^j} \otimes \mathrm{d}\overline{z}^{j_1} \wedge \cdots \mathrm{d}\overline{z}^{j_\ell}\right]$$
$$= \left(f\frac{\partial g}{\partial z^i}\frac{\partial}{\partial z^j} - g\frac{\partial f}{\partial z^j}\frac{\partial}{\partial z^i}\right) \otimes \mathrm{d}\overline{z}^{i_1} \wedge \cdots \mathrm{d}\overline{z}^{i_k} \otimes \mathrm{d}\overline{z}^{j_1} \wedge \cdots \mathrm{d}\overline{z}^{j_\ell}.$$

We obtain a differential graded Lie algebra.

Let $(C, [\ ,\ ], \mathrm{d})$ be a differential graded Lie algebra. We put

$$\mathfrak{m}_1(x) = (-1)^{\deg x}\,\mathrm{d}x \qquad \mathfrak{m}_2(x, y) = (-1)^{\deg x(\deg y+1)}[x, y]$$

and $\mathfrak{m}_k = 0$ for $k \geq 3$.

Lemma 8.3.4. *$\mathfrak{m}_k$ determines the structure of an L_∞ algebra on C.*

Proof. Let us first note that $\mathfrak{m}_2$ is a well-defined operator on $E_2C[1]$, i.e. we have

$$\begin{aligned}\mathfrak{m}_2(x, y) &= (-1)^{\deg x(\deg y+1)}[x, y] = (-1)^{\deg x(\deg y+1)+\deg x \deg y+1}[y, x] \\ &= (-1)^{\deg x(\deg y+1)+\deg x \deg y+1+\deg y(\deg x+1)}\mathfrak{m}_2(y, x) \\ &= (-1)^{(\deg x+1)(\deg y+1)}\mathfrak{m}_2(y, x).\end{aligned}$$

It then suffices to show that the $\mathrm{Hom}(E_kC[1], C[1])$ component of $\delta\delta$ is zero for $k = 1, 2, 3$. The case $k = 1$ is obvious. Let us check the case $k = 3$, and leave the case $k = 2$ to the reader. Let $\pi_1 : EC[1] \to E_1C[1] = C[1]$ be the projection.

We put $x_1 \times \cdots \times x_k = \sum_{\sigma\in\mathfrak{S}_k} \pm x_{\sigma(1)} \otimes \cdots \otimes x_{\sigma(k)}$, where $\pm$ is as in formula (8.31). Then we have:

$$\begin{aligned}&\tfrac{1}{2}\pi_1\delta\delta(x \times y \times z) \\ &\quad= \mathfrak{m}_2(\mathfrak{m}_2(x, y), z) + (-1)^{\deg' z(\deg' x+\deg' y)}\mathfrak{m}_2(\mathfrak{m}_2(z, x), y) \\ &\qquad+ (-1)^{\deg' x(\deg' y+\deg' z)}\mathfrak{m}_2(\mathfrak{m}_2(y, z), x) \\ &\quad= (-1)^{\deg x(\deg y+1)+(\deg x+\deg y)(\deg z+1)}[[x, y], z] \\ &\qquad+ (-1)^{\deg z(\deg x+1)+(\deg y+\deg z)(\deg x+1)+(\deg z+1)(\deg x+\deg y)}[[z, x], y] \\ &\qquad+ (-1)^{\deg y(\deg z+1)+(\deg y+\deg z)(\deg x+1)+(\deg x+1)(\deg y+\deg z)}[[y, z], x] \\ &\quad= 0.\end{aligned}$$

(Note this calculation is a bit problematic in case 2 as it is not invertible. We do not try to correct it since the case when R is a field of characteristic zero is our main concern.) $\square$

Our next purpose is to define the homotopy equivalence of A_∞ and L_∞ algebras. For this purpose we first define A_∞ and L_∞ homomorphisms.

Definition 8.3.5. Let (D, Δ, ϵ), (D', Δ', ϵ') be coalgebras. An R module homomorphism $\varphi : D \to D'$ of degree 0 is said to be a *coalgebra homomorphism* if the following diagram commutes.

$$\begin{tikzcd} D' \otimes D' & D' \arrow[l, "\Delta'"'] \\ D \otimes D \arrow[u, "\varphi\otimes\varphi"] & D \arrow[l, "\Delta"] \arrow[u, "\varphi"'] \end{tikzcd} \qquad \begin{tikzcd} R & D' \arrow[l, "\epsilon'"'] \\ R \arrow[u, equal] & D \arrow[l, "\epsilon"] \arrow[u, "\varphi"'] \end{tikzcd}$$

Diagram 7 Diagram 8

Lemma 8.3.5. *Let* $\varphi_k : B_k C[1] \to C'[1]$, $k = 1, 2, \ldots$ *be a sequence of degree*-0 *R module homomorphisms. Then there exists a unique colagebra homomorphism* $\hat{\varphi} : BC[1] \to BC'[1]$ *such that its* $\mathrm{Hom}(B_k C[1], B_1 C'[1])$ *component is* φ_k. *The same statement holds if we replace B by E.*

Proof. Let $\varphi : BC[1] \to C'[1]$ be a homomorphism whose $\mathrm{Hom}(B_k C[1], B_1 C'[1])$ component is φ_k. We then set (using notation (8.34))

$$\hat{\varphi}(\mathbf{x}) = \sum_k \sum_a \varphi(\mathbf{x}_a^{k;1}) \otimes \cdots \otimes \varphi(\mathbf{x}_a^{k;k}).$$

It is easy to check that this $\hat{\varphi}$ has the required property. The L_∞ case is similar. □

Definition 8.3.6. Let $(C[1], \mathfrak{m}_k)$, $(C'[1], \mathfrak{m}'_k)$ be A_∞ algebras. A sequence of homomorphisms $\varphi_k : B_k C[1] \to C'[1]$ is said to be an A_∞ *homomorphism* if the coalgebra homomorphism $\hat{\varphi} : BC[1] \to BC'[1]$ obtained by lemma 8.3.5 satisfies $\hat{\varphi} \circ \delta = \delta \circ \hat{\varphi}$. The definition of the L_∞ *homomorphism* is similar.

We can define a composition of A_∞ and L_∞ homomorphisms by $\widehat{\varphi \circ \psi} = \hat{\varphi} \circ \hat{\psi}$.

We next define a homotopy between A_∞ and L_∞ homomorphisms. The definition we give here is an analogy of a similar definition in the case of the differential graded algebra given in [41, chapter X]. There is another way to define homotopy which is an analogy of one in [100] (which we do not discuss any further here).

Let $(C[1], \mathfrak{m}_k)$ be an A_∞ algebra. We define an A_∞ algebra $C[1] \otimes R[t, \mathrm{d}t]$ as follows.

Definition 8.3.7. An element of $C[1] \otimes R[t, \mathrm{d}t]$ is written as $P(t) + Q(t)\,\mathrm{d}t$, where $P, Q \in C[t]$ are polynomials with coefficients on C. We put $\deg \mathrm{d}t = 1$.

The operator $\mathfrak{m}_k$ is defined as follows. Let $x_i = P_i(t) + Q_i(t)\,\mathrm{d}t$

$$\mathfrak{m}_1(P(t) + Q(t)\,\mathrm{d}t) = \mathfrak{m}_1(P(t)) - \mathfrak{m}_1(Q(t))\,\mathrm{d}t - \frac{\mathrm{d}P}{\mathrm{d}t}\,\mathrm{d}t \tag{8.36}$$

$$\begin{aligned}\mathfrak{m}_k(x_1, \ldots, x_k) = \mathfrak{m}_k(P_1, \ldots, P_k) + \sum_{i=1}^{k}(-1)^{\deg' P_1 + \cdots \deg' P_{i-1}+1} \\ \times \mathfrak{m}_k(P_1, \ldots, Q_i, \ldots, P_k)\,\mathrm{d}t\end{aligned} \tag{8.37}$$

Here we extend $\mathfrak{m}_k$ to $B_k C[t]$ in an obvious way.

When $(C[1], \mathfrak{m}_k)$ is an L_∞ algebra, we define $C[1] \otimes R[t, \mathrm{d}t]$ in the same way as an R module. The definition of operations $\mathfrak{m}_k$ is also the same by using (8.36), (8.37).

We omit the proof of the A_∞ and L_∞ formulae (see the final version of [33]). For $t_0 \in \mathbb{R}$, we define an A_∞ homomorphism $\mathrm{Eval}_{t=t_0} : C[1] \otimes R[t, \mathrm{d}t] \to C[1]$ by

$$\mathrm{Eval}_{t=t_0}(P(t) + Q(t)\,\mathrm{d}t) = P(t_0). \tag{8.38}$$

More precisely, we define the $B_1(C \otimes R[t, \mathrm{d}t])[1] \to C[1]$ component by (8.38) and set all the other components to 0.

Definition 8.3.8. Two A_∞ homomorphisms $\varphi, \varphi' : C \to C'$ are said to be *homotopic* to each other, if there exists an A_∞ homomorphism $H : C \to C' \otimes R[t, \mathrm{d}t]$ such that $\mathrm{Eval}_{t=0} \circ H = \varphi$, $\mathrm{Eval}_{t=1} \circ H = \varphi'$. We define the homotopy between L_∞ homomorphisms in the same way.

In theorems 8.3.1 and 8.3.2, we assume R contains $\mathbb{Q}$.

Theorem 8.3.1. *If φ is homotopic to φ' and φ' is homotopic to φ'' then φ is homotopic to φ''.*

When C, C' are differential graded algebras, theorem 8.3.1 is proved in [41]. The general A_∞ algebra case is similar and is proved in detail in [33]. We omit the proof in this article.

Definition 8.3.9. An A_∞ homomorphism $\varphi : C \to C'$ is said to be a *homotopy equivalence* if there exists an A_∞ homomorphism $\psi : C' \to C$ such that the compositions $\psi \circ \varphi$, $\varphi \circ \psi$ are homotopic to the identity. Two A_∞ algebras are said to be *homotopy equivalent* if there exists a homotopy equivalence between them. The homotopy equivalence of L_∞ algebras is defined in the same way.

The following theorem is useful to show that an A_∞ homomorphism is a homotopy equivalence.

Theorem 8.3.2. *If $\varphi : C \to C'$ is an A_∞ homomorphism which induces an isomorphism on an $\mathfrak{m}_1$ cohomology, then φ is a homotopy equivalence. The same holds for the L_∞ case.*

Remark 8.3.1. Theorem 8.3.2 does *not* hold in the category of differential graded algebras; i.e. if $\varphi : C \to C'$ is a differential graded algebra homomorphism (of degree 0) which induces an isomorphism on cohomology. Then theorem 8.3.2 implies that we can find a homotopy inverse of it which is an A_∞ homomorphism. However, it is not, in general, possible to find a homotopy inverse which is a differential graded algebra homomorphism.

Theorem 8.3.2 is proved in a somewhat weaker version in the 2000 December version of [33]. The proof of the general case will be included in the final version of [33].

8.3.2 Maurer–Cartan equation and moduli functors

We now generalize the Maurer–Cartan equation (8.5) to the case of A_∞ and L_∞ algebras, i.e. we consider the equation

$$\sum_k \mathfrak{m}_k(b, \ldots, b) = 0 \qquad A_\infty \text{ case} \tag{8.39a}$$

$$\sum_k \frac{1}{k!}\mathfrak{m}_k(b, \ldots, b) = 0 \qquad L_\infty \text{ case.} \tag{8.39b}$$

where $b \in C[1]^0 = C^1$. Note that (8.39a) coincides with (8.5) in a differential graded algebra and (8.39b) coincides with (8.30) in a differential graded Lie algebra.

There is, however, one problem in making sense of equations (8.39a), (8.39b), i.e. the left-hand side is an infinite sum when infinitely many of $\mathfrak{m}_k$ are non-zero. There are two ways to make sense of (8.39a), (8.39b). One is to define a topology on C and consider the case when the left-hand side converges. (We may either consider a non-Archimedean valuation on our coefficient ring ($R = \mathbb{C}[[T]]$ for example) or convergence in the classical sense ($R = \mathbb{R}$ or $\mathbb{C}$). Both play a role in the story of mirror symmetry.) The other possibility is to consider b which is nilpotent (i.e. the product of several of them vanishes). This second point is related to the 'functor from the Artin ring' discussed in section 8.1.7. Let us take this second point of view in this section. (The first point of view also appears later.) Let us again consider the case when R is an algebraically closed field of characteristic zero.

Let $\mathcal{R}$ be a finite dimensional local R algebra (commutative with unit). Let $\mathcal{R}_+$ be the maximal ideal of $\mathcal{R}$. There exists N such that $\mathcal{R}_+^N = 0$. Let C be an A_∞ algebra. $C_\mathcal{R} = C \otimes_R \mathcal{R}$ has a structure of A_∞ algebra. Let $b \in C^1 \otimes_R \mathcal{R}_+$. Obviously $\mathfrak{m}_k(b, \ldots, b) = 0$ if $k > N$. Hence equation (8.39) makes sense.

Definition 8.3.10. b is said to be a *Maurer–Cartan element* of $C_\mathcal{R}$ if it satisfies equation (8.39).

To simplify the notation we introduce the following notation.

$$e^b = \sum_{k=0}^{\infty} \underbrace{b \otimes \cdots \otimes b}_{k \text{ times}} \qquad A_\infty \text{ case} \tag{8.40a}$$

$$e^b = \sum_{k=0}^{\infty} \frac{1}{k!} \underbrace{b \otimes \cdots \otimes b}_{k \text{ times}} \qquad L_\infty \text{ case.} \tag{8.40b}$$

We note that $\mathfrak{m} : BC[1] \to C[1]$ is a homomorphism which is $\mathfrak{m}_k$ on $B_k C[1]$. ($\mathfrak{m} : EC[1] \to C[1]$ is similar.)

Then equation (8.39) can be written as $\mathfrak{m}(e^b) = 0$. Before going further let us explain the meaning of equation (8.39). Let us define a deformed boundary operator $\mathfrak{m}_1^b$ by

$$\mathfrak{m}_1^b(x) = \mathfrak{m}(e^b, x, e^b) \qquad A_\infty \text{ case} \tag{8.41a}$$

$$\mathfrak{m}_1^b(x) = \mathfrak{m}(e^b, x) \qquad L_\infty \text{ case.} \tag{8.41b}$$

Lemma 8.3.6. $\mathfrak{m}_1^b \mathfrak{m}_1^b = 0$ *if and only if* (8.39) *is satisfied.*

The proof is easy and is omitted. We can also deform $\mathfrak{m}_k$ by

$$\mathfrak{m}_k^b(x_1, \ldots, x_k) = \mathfrak{m}(e^b, x_1, e^b, x_1, \ldots, x_{k-1}, e^b, x_k, e^b) \qquad A_\infty \text{ case,}$$
$$\mathfrak{m}_k^b(x_1, \ldots, x_k) = \mathfrak{m}(e^b, x_1, \ldots, x_k) \qquad L_\infty \text{ case.}$$

Then we obtain either an A_∞ or L_∞ algebra $(C, \mathfrak{m}_*^b)$ (see [33]).

We now define a functor $\widetilde{\mathcal{MC}}(C) : \{\text{f. d. Alg.}/R\} \to \{\text{Sets}\}$ as follows. $\widetilde{\mathcal{MC}}(C)(\mathcal{R})$ is the set of all Maurer–Cartan elements of $C_{\mathcal{R}}$. If $\psi : \mathcal{R} \to \mathcal{R}'$ is a morphism in $\{\text{f. d. Alg.}/R\}$ and if b is a Maurer–Cartan element of $C_{\mathcal{R}}$, then $(1 \otimes \psi)(b) \in C_{\mathcal{R}'}$ is a Maurer–Cartan elements of $C_{\mathcal{R}'}$. Thus we obtain $\psi_* : \widetilde{\mathcal{MC}}(C)(\mathcal{R}) \to \widetilde{\mathcal{MC}}(C)(\mathcal{R}')$. We thus obtain a covariant functor $\widetilde{\mathcal{MC}}(C) : \{\text{f. d. Alg.}/R\} \to \{\text{Sets}\}$.

However, the set $\widetilde{\mathcal{MC}}(C)(\mathcal{R})$ is usually too big. So we divide it by an appropriate gauge equivalence, which we now define.

Definition 8.3.11. Let $b, b' \in \widetilde{\mathcal{MC}}(C)(\mathcal{R})$. We say that b is *gauge equivalent* to b' if there exists an element $\tilde{b} \in \widetilde{\mathcal{MC}}(C \otimes R[t, \mathrm{d}t])(\mathcal{R})$ such that $\mathrm{Eval}_{t=0}\, \tilde{b} = b$, $\mathrm{Eval}_{t=1}\, \tilde{b} = b'$. Here $\mathrm{Eval}_{t=t_0} : C_{\mathcal{R}} \otimes R[t, \mathrm{d}t] \to C_{\mathcal{R}}$ is an A_∞ homomorphism as in the last section.

In the following theorem, we assume R contains $\mathbb{Q}$.

Theorem 8.3.3. *Let $b, b', b'' \in \widetilde{\mathcal{MC}}(C)(\mathcal{R})$. If b is gauge equivalent to b' and b' is gauge equivalent to b'', then b is gauge equivalent to b''.*

The proof is similar to that of theorem 8.3.1 and will be given in the final version of [33].

Let us rewrite the definition of gauge equivalence in a more concrete way. Here we only discuss the A_∞ case (the L_∞ case is similar). Let $\tilde{b}$ be as in definition 8.3.11. We put $\tilde{b} = b(t) + c(t)\,\mathrm{d}t$ where $b(t) \in C^1 \otimes \mathcal{R}_+[t]$, $c(t) \in C^0 \otimes \mathcal{R}_+[t]$. Equation (8.39a) for $\tilde{b}$ becomes

$$\frac{\mathrm{d}b(t)}{\mathrm{d}t} + \sum_{k=1}^{N}\sum_{i=1}^{k} \mathfrak{m}_k(b(t)^{\otimes i-1}, c(t), b(t)^{\otimes k-i}) = 0 \tag{8.42}$$

$$\sum_{k=1}^{N} \mathfrak{m}_k(b(t)^{\otimes k}) = 0. \tag{8.43}$$

The condition $\mathrm{Eval}_{t=0}\,\tilde{b} = b$, $\mathrm{Eval}_{t=1}\,\tilde{b} = b'$ is a boundary condition $b(0) = b, \quad b(1) = b'$. We remark that (8.42) and equation (8.39a) for b imply (8.43). In fact $\frac{\mathrm{d}}{\mathrm{d}t}\sum_{k=1}^{N} \mathfrak{m}_k(b(t)^{\otimes k})$ is the left-hand side of (8.42).

We note that, using (8.41a), equation (8.42) can be written as

$$\frac{\mathrm{d}b(t)}{\mathrm{d}t} + \mathfrak{m}_1^{b(t)}(c(t)) = 0. \tag{8.44}$$

Let us consider the case when C is a differential graded algebra and $R = \mathbb{C}$ or $\mathbb{R}$. Then (8.44) is

$$\frac{\mathrm{d}b(t)}{\mathrm{d}t} + \mathrm{d}(c(t)) - b(t)\cdot c(t) + c(t)\cdot b(t) = 0. \tag{8.45}$$

Let $g(t)$ be the solution of the differential equation :

$$\frac{\mathrm{d}g(t)}{\mathrm{d}t} = g(t)\cdot c(t) \qquad g(0) = I \tag{8.46}$$

We note that we use the fact that R is a ring *with characteristic 0* to solve equation (8.46); and $g(t) - 1 \in \mathcal{R}_+$ for each t hence $g(t)$ is invertible in the ring $C^0_{\mathcal{R}}$.

Lemma 8.3.7. $g(t)^{-1}\mathfrak{m}_1^b(g(t)) = b - b(t)$.

Proof. We may assume $b = 0$ by replacing $\mathfrak{m}_1$ by $\mathfrak{m}_1^b$. The lemma is obvious for $t = 0$. By (8.46), we have:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathfrak{m}_1(g(t)) &= \mathfrak{m}_1(g(t)\cdot c(t)) = d(g(t)\cdot c(t))\\ &= \mathfrak{m}_1(g(t))\cdot c(t) + g(t)\cdot \mathrm{d}(c(t))\\ &= -g(t)\cdot b(t)\cdot c(t) + g(t)\cdot \mathrm{d}(c(t))\end{aligned}$$

However, we have, by (8.46),

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(g(t)\cdot b(t)) &= \frac{\mathrm{d}}{\mathrm{d}t}g(t)\cdot b(t) + g(t)\cdot\frac{\mathrm{d}}{\mathrm{d}t}b(t)\\ &= g(t)\cdot c(t)\cdot b(t) - g(t)\cdot \mathrm{d}(c(t))\\ &\quad + g(t)\cdot b(t)\cdot c(t) - g(t)\cdot c(t)\cdot b(t).\end{aligned}$$

The lemma follows. $\square$

When $C = (\Omega^{0,*}(\mathcal{E}), \overline{\partial}_{\mathcal{E}}, \wedge)$ which we studied in chapter 1, we have $g \in \Gamma(\mathrm{End}(\mathcal{E}))$, $\mathfrak{m}_1^B = \overline{\partial}_{\mathcal{E}-B}$. Lemma 8.3.7 then implies $\overline{\partial}_{\mathcal{E}+B} = g^{-1} \circ \overline{\partial}_{\mathcal{E}} \circ g$. In other words g is an isomorphism from $(E, \overline{\partial}_{\mathcal{E}+B})$ to $\overline{\partial}_{\mathcal{E}}$. This justifies our terminology—gauge equivalence.

Definition 8.3.12. $\mathcal{MC}(C)(\mathcal{R})$ is the set of all gauge equivalence classes of elements of $\widetilde{\mathcal{MC}}(C)(\mathcal{R})$.

If $\psi : \mathcal{R} \to \mathcal{R}'$ is a morphism in $\{\text{f. d. Alg.}/R\}$, we can construct $\psi_* : \mathcal{MC}(C)(\mathcal{R}) \to \mathcal{MC}(C)(\mathcal{R}')$ in an obvious way. Hence $\mathcal{MC}(C)$ defines a functor: $\{\text{f. d. Alg.}/R\} \to \{\text{Sets}\}$. We call it the *Maurer–Cartan functor* associated with either an A_∞ or L_∞ algebra C.

Our next goal is to show that the Maurer–Cartan functor is homotopy invariant. Let $\varphi_k : B_k C[1] \to C'[1]$ be an A_∞ or L_∞ homomorphism. It induces $\hat{\varphi} : BC[1] \to BC'[1]$, or $EC[1] \to EC'[1]$.

Lemma 8.3.8. *There exists $\varphi_* : C[1]^0 \to C'[1]^0$ such that $\hat{\varphi}(e^b) = e^{\varphi_*(b)}$.*

Proof. Put
$$\varphi_*(b) = \varphi(e^b) = \sum_k \varphi_k(b, \dots, b).$$
Here $\varphi : BC[1] \to C'[1]$ is a homomorphism which is φ_k on $B_k C[1]$. The L_∞ case is similar. $\square$

Using the fact that $\hat{\varphi}$ is a chain map we have the following lemma.

Lemma 8.3.9. *If $b \in \widetilde{\mathcal{MC}}(C)(\mathcal{R})$ then $\varphi_*(b) \in \widetilde{\mathcal{MC}}(C')(\mathcal{R})$.*

We also have the following one.

Lemma 8.3.10. *If $b \sim b'$ and if φ_k is homotopic to φ_k' then $\varphi_*(b) \sim \varphi_*'(b')$.*

Proof. Let $H : C \to C' \otimes R[t, \mathrm{d}t]$ be as in definition 8.3.8. It induces $\tilde{H} : C \otimes R[t, \mathrm{d}t] \to C' \otimes R[t, \mathrm{d}t]$ as follows. We put $H_k(v_1, \dots, v_k) = H_k^1(v_1, \dots, v_k) + H_k^2(v_1, \dots, v_k)\,\mathrm{d}t$, where $H_k^i : B_k C[1] \to C'[1] \otimes R[t]$. We extend H_k^i to $B_k(C[1] \otimes R[t]) \to C'[1] \otimes R[t]$ in an obvious way and denote it by the same symbol. Let $x_i = P_i(t) + Q_i(t)\,\mathrm{d}t$. Then $\tilde{H}_k(x_1, \dots, x_k) = P(t) + Q(t)\,\mathrm{d}t$ where
$$\begin{aligned} P(t) &= H_k^1(P_1(t), \dots, P_k(t)) \\ Q(t) &= H_k^2(P_1(t), \dots, P_k(t)) \\ &\qquad + \sum_i (-1)^{\deg' P_1 + \cdots \deg' P_{i-1} + 1} H_k^1(P_1(t), \dots, Q_i(t), \dots, P_k(t)). \end{aligned}$$
It is easy to check that $\tilde{H}$ is an A_∞ homomorphism. Let $\tilde{b} \in \widetilde{\mathcal{MC}}(C \otimes R[t, \mathrm{d}t])$ such that $\mathrm{Eval}_{t=0}\,\tilde{b} = b$, $\mathrm{Eval}_{t=1}\,\tilde{b} = b$ (definition 8.3.11). Now $\tilde{H}_*(\tilde{b}) \in \widetilde{\mathcal{MC}}(C' \otimes R[t, \mathrm{d}t])$ and $\mathrm{Eval}_{t=0}\,\tilde{H}_*(\tilde{b}) = \varphi_*(b)$, $\mathrm{Eval}_{t=1}\,\tilde{H}_*(\tilde{b}) = \varphi_*'(b)$. $\square$

By lemmata 8.3.9 and 8.3.10 we obtain a map $\varphi_*(\mathcal{R}) : \mathcal{MC}(C)(\mathcal{R}) \to \mathcal{MC}(C')(\mathcal{R})$. It is easy to see that the following diagram commutes for each morphism $\psi : \mathcal{R} \to \mathcal{R}'$ in {f. d. Alg. $/R$}.

$$\begin{array}{ccc}
\mathcal{MC}(C)(\mathcal{R}) & \xrightarrow{\psi_*} & \mathcal{MC}(C)(\mathcal{R}') \\
{\scriptstyle\varphi_*}\downarrow & & \downarrow{\scriptstyle\varphi_*} \\
\mathcal{MC}(C')(\mathcal{R}) & \xrightarrow{\psi_*} & \mathcal{MC}(C')(\mathcal{R}')
\end{array}$$

Diagram 9

The commutativity of diagram 9 implies that φ_* is a natural transformation of Maurer–Cartan functors: $\mathcal{MC}(C) \to \mathcal{MC}(C')$. Moreover, lemma 8.3.10 implies that $\varphi_* : \mathcal{MC}(C) \to \mathcal{MC}(C')$ depends only on the homotopy class of φ. The following theorem follows immediately.

Theorem 8.3.4. *If C is homotopy equivalent to C' then the Maurer–Cartan functor $\mathcal{MC}(C)$ is equivalent to $\mathcal{MC}(C')$.*

Remark 8.3.2. Theorem 8.3.4 for differential graded algebras and differential graded Lie algebras is due to Goldman and Milson [37, 38]. Its generalization to A_∞ or L_∞ algebras seems to have been folklore and was quoted by several authors without proof (for example by Kontsevitch [69]). We give its rigorous proof here, assuming theorems 8.3.1, 8.3.2 and 8.3.3 which will be proved in the final version of [33].

Remark 8.3.3. We can state theorem 8.3.4 in a more functorial way as follows. Let $\{A_\infty\,\mathrm{alg.}\,/R\}/$ homotopy be the category whose object is the set of all homotopy equivalence classes of A_∞ algebras over R and whose morphisms are homotopy classes of A_∞ homomorphisms.[10] Let $\mathfrak{Funk}(\{\mathrm{f.\,d.\,Alg.}\,/R\}, \{\mathrm{Sets}\})$ be the category whose objects are the sets of all equivalence classes of covariant functors from {f. d. Alg. $/R$} to {Sets}. Then $\mathcal{MC}$ induces a functor: $\{A_\infty\,\mathrm{alg.}\,/R\}/\,\mathrm{homotopy} \to \mathfrak{Funk}(\{\mathrm{f.\,d.\,Alg.}\,/R\}, \{\mathrm{Sets}\})$.

8.3.3 Canonical model, Kuranishi map and moduli space

In this section, we apply theorem 8.3.4 to construct a versal formal moduli space representing the Maurer–Cartan functor. Theorem 8.3.4 is useful for constructing a moduli space because it enables us to replace a given A_∞ algebra with another one which is homotopy equivalent to the original one but is easier to handle. A good representative of each homotopy class for our purpose is a canonical one, which we now define.

Definition 8.3.13. An A_∞ (or L_∞) algebra $(C, \mathfrak{m}_*)$ is said to be *canonical* if $\mathfrak{m}_1 = 0$.

[10] It seems to be standard notation to say 'set of all homotopy equivalence classes of A_∞ algebras'. We can go round it by introducing a universe in the same way as in [13].

Remark 8.3.4. In [69] Kontsevich called the same notion a 'minimal' L_∞ algebra. Sullivan [100] used 'minimal' for a differential graded algebra for an important notion which differs from definition 8.3.13. This is why we use canonical rather than minimal. The name canonical may be justified by proposition 8.3.1.

We say an A_∞ (or L_∞) homomorphism $\varphi : C \to C'$ is an *isomorphism* if there exists an A_∞ (or L_∞) homomorphism $\varphi' : C' \to C$ such that the compositions $\varphi' \circ \varphi$ and $\varphi \circ \varphi'$ *are equal to* the identity. Here the identity A_∞ homomorphism id_* is defined by $\mathrm{id}_1 = \mathrm{id}$ and $\mathrm{id}_k = 0$ for $k \geq 2$.

Proposition 8.3.1. *A homotopy equivalence between canonical A_∞ (or L_∞) algebras is an isomorphism.*

Proof. The condition $\mathfrak{m}_1 = 0$ implies that the $\mathfrak{m}_1$ cohomology of C is isomorphic to C itself. Since homotopy equivalence φ_* induces an isomorphism on an $\mathfrak{m}_1$ cohomology it follows that $\varphi_1; C[1] \to C[1]$ is an isomorphism. We can then easily prove that $\hat{\varphi} : BC[1] \to BC[1]$ is an isomorphisms. The converse of it is a cochain map which is a coalgebra map. Hence, there exists ψ_k such that $\hat{\psi} = \hat{\varphi}^{-1}$. ψ_k is the inverse of φ, as required. □

Hereafter in this section we assume that R is a field of characteristic 0. We also assume that $H(C, \mathfrak{m}_1)$ is finite dimensional.

Theorem 8.3.5. *There exists a canonical A_∞ algebra C_{can} homotopy equivalent to a given A_∞ algebra C. The same holds for L_∞ algebras.*

Remark 8.3.5. Theorem 8.3.5 was first proved in [54], see also [46, 77]. There might be some others who have found it independently (for example I heard of a talk by A Polishchuk discussing the same theorem in 1998 January at the Winter School held in Harvard University). The proof here is similar to one by Kontsevich and Soibelman [70], and also to [33] (2000 December version) section 8.A6.

Proof. The argument is similar to one explained in section 8.1.5. We prove the L_∞ algebra case to minimize the overlap with the argument in section 8.1.5. Let C be an L_∞ algebra. We put $C^k_{\mathrm{can}} = H^k(C, \mathfrak{m}_1)$. We first need an analogue of theorem 8.2.5. Here we need to use the fact that our coefficient ring is a field.

Lemma 8.3.11. *There exists a linear subspaces $H^k \subseteq C^k$, projections $\Pi_{H^k} : C^k \to H^k$ and R linear maps $G_k : C^k \to C^{k-1}$ such that*

$$G_{k+1} \circ \mathfrak{m}_1 + \mathfrak{m}_1 \circ G_k = 1 - \Pi_{H^k}. \tag{8.47}$$

Proof. We put $Z_k = \mathrm{Ker}\, \mathfrak{m}_1 : C^k \to C^{k+1}$. Let $H^k \subseteq Z_k$ be a linear subspace such that the restriction of the projection: $Z_k \to C^k_{\mathrm{can}} = H^k(C, \mathfrak{m}_1)$ to H^k is an isomorphism. We put $B^k = \mathrm{Im}\, \mathfrak{m}_1 : C^{k-1} \to C^k$. Then $H^k \oplus B^k = Z_k$. We also choose $I^k \subseteq C^k$ such that $Z_k \oplus I^k = C^k$. It is easy to see that $\mathfrak{m}_1$ induces an isomorphism $I^{k-1} \to B^k$. Let $G^k : B^k \to I^{k-1}$ be an inverse of it. We extend G^k to C^k so that $G^k = 0$ on H^k and on I^k. It is easy to check (8.47). □

We call G_k a *propagator*. We remark that in section 8.1.4 we constructed a similar operator (see (8.13)).

Now we consider the set of trees which satisfies a slightly milder condition than condition 8.2.1, i.e. we consider the following condition.

Condition 8.3.1. Γ satisfies (1), (3), (4), (5) of condition 8.2.1 and

(2)′ If $v \in \mathrm{Vertex}_{\int}(\Gamma)$ then $\sharp\partial_{\mathrm{target}}^{-1}(v) \geq 2$, $\sharp\partial_{\mathrm{source}}^{-1}(v) = 1$.

We denote by Gr_k the set of all oriented graphs Γ satisfying condition 8.3.1. (We do not take a ribbon structure here, since we are studying an L_∞ algebra and not an A_∞ algebra.) We use the notation from section 8.1.4.

Remark 8.3.6. The relation of trees to A_∞ or L_∞ algebras has been known to algebraic topologists for a long time, see, for example, [11].

We first put $\varphi_1 = \mathrm{id}$ and $\overline{\mathfrak{m}}_1 = 0$ and are going to define

$$\overline{\mathfrak{m}}_\Gamma : E_k C_{\mathrm{can}}[1] \to C_{\mathrm{can}}[1] \qquad \varphi_\Gamma : E_k C_{\mathrm{can}}[1] \to C[1] \tag{8.48}$$

for $\Gamma \in Gr_k$ inductively on k. We will then put : $\overline{\mathfrak{m}}_k = \sum_{\gamma \in Gr_k} \overline{\mathfrak{m}}_\Gamma$, $\varphi_k = \sum_{\gamma \in Gr_k} \varphi_\Gamma$.

Now let us assume that (8.48) is defined for $\Gamma \in Gr_\ell$, $\ell < k$. Let $\Gamma \in Gr_k$. Let v_{last} be its last vertex. Let e_{last} be the unique edge such that $\partial_{\mathrm{target}}(e_{\mathrm{last}}) = v_{\mathrm{last}}$. We remove $[0,1]_{e_{\mathrm{last}}}$ together with its two vertices from $|\Gamma|$. Then $|\Gamma| \backslash [0,1]_{e_{\mathrm{last}}}$ is a union $\cup|\Gamma_i|$ of several elements $\Gamma_i \in Gr_{k_i}$ with $\sum_{i=1}^{\ell} k_i = k$. We note that since we are using a graph Γ which does not have a particular ribbon structure, there is no canonical way to order $\Gamma_1, \dots, \Gamma_\ell$. So the construction here should be independent of the order.

Let $x_i \in C_{\mathrm{can}}[1]$. Let $\mathfrak{S}_k$ be the group of all permutation of k numbers $\{1, \dots, k\}$. We put

$$y_{1,\sigma} = \overline{\varphi}_{\Gamma_1}(x_{\sigma(1)}, \dots, x_{\sigma(k_1)}), \dots, y_{\ell,\sigma} = \overline{\varphi}_{\Gamma_\ell}(x_{\sigma(k-k_\ell+1)}, \dots, x_{\sigma(k)})$$

and then

$$\overline{\varphi}_\Gamma(x_1 \otimes \cdots \otimes x_k) = -\sum_{\sigma \in \mathfrak{S}_k} \pm \frac{1}{k_1! \dots k_\ell!} G(\mathfrak{m}_\ell(y_{1,\sigma}, y_{2,\sigma}, \dots, y_{\ell,\sigma}))$$

$$\overline{\mathfrak{m}}_\Gamma(x_1 \otimes \cdots \otimes x_k) = -\sum_{\sigma \in \mathfrak{S}_k} \pm \frac{1}{k_1! \dots k_\ell!} \Pi_H(\mathfrak{m}_\ell(y_{1,\sigma}, y_{2,\sigma}, \dots, y_{\ell,\sigma}))$$

where $\pm$ is as in (8.31) and G is the homomorphism in lemma 8.3.11.

We now calculate

$$
\begin{aligned}
&\sum_{\Gamma\in Gr_k} \mathfrak{m}_1(\varphi_\Gamma(x_1\otimes\cdots\otimes x_k))\\
&\qquad = -\sum_{\Gamma\in Gr_k}\sum_{\sigma\in\mathfrak{S}_k} \pm\mathfrak{m}_1(G(\mathfrak{m}_\ell(y_{1,\sigma},y_{2,\sigma},\ldots,y_{\ell,\sigma})))\\
&\qquad = \sum_{\sigma\in\mathfrak{S}_k} \pm\mathfrak{m}_\ell(y_{1,\sigma},y_{2,\sigma},\ldots,y_{\ell,\sigma})\\
&\qquad\quad -\sum_{\Gamma\in Gr_k}\sum_{\sigma\in\mathfrak{S}_k} \pm G(\mathfrak{m}_1(\mathfrak{m}_\ell(y_{1,\sigma},y_{2,\sigma},\ldots,y_{\ell,\sigma}))).
\end{aligned}
\tag{8.49}
$$

We are going to calculate the last line of (8.49) using the L_∞ relation of $\mathfrak{m}$. We have

$$
\begin{aligned}
&-\mathfrak{m}_1(\mathfrak{m}_\ell(y_{1,\sigma},y_{2,\sigma},\ldots,y_{\ell,\sigma}))\\
&\quad =\sum_{m>1}\sum_{n_1+\cdots+n_m=\ell}\sum_{\mu\in\mathfrak{S}_m} \pm\mathfrak{m}_m(\mathfrak{m}_{n_1}(y_{\mu(1),\sigma},\ldots),\ldots,\mathfrak{m}_{n_m}(\ldots,y_{\mu(\ell),\sigma}))).
\end{aligned}
\tag{8.50}
$$

We let $\hat\varphi : EC_{\mathrm{can}}[1]\to EC[1]$ be the coalgebra homomorphism induced by φ_k and $\overline{\delta} : EC_{\mathrm{can}}[1]\to EC_{\mathrm{can}}[1]$ be the coderivation induced by $\overline{\mathfrak{m}}_k$. The coderivation $\delta : EC[1]\to EC[1]$ is induced by $\mathfrak{m}_*$. We now prove the following lemma.

Lemma 8.3.12. *We have* $\delta\circ\hat\varphi=\hat\varphi\circ\overline{\delta}$ *and* $\overline{\delta}\circ\overline{\delta}=0$.

Proof. We prove the equalities on $\mathfrak{G}^k EC_{\mathrm{can}}[1]=\oplus_{i\le k}E_iC_{\mathrm{can}}[1]$ by induction on k. The case $k=1$ is obvious. Let $\mathbf{x}\in E_kC_{\mathrm{can}}[1]$. By (8.49), (8.50) we have

$$
\mathfrak{m}_1(\varphi_k(\mathbf{x})) = -\sum_{\ell>2}\overline{\mathfrak{m}}_\ell(\hat\varphi(\mathbf{x})) + (G\circ(\mathfrak{m}-\mathfrak{m}_1)\circ\delta\circ\hat\varphi)(\mathbf{x}). \tag{8.51}
$$

Here $\mathfrak{m}-\mathfrak{m}_1 : EC[1]\to C[1]$ is an operator which is zero on $E_1C[1]$ and is $\mathfrak{m}_\ell$ on $E_\ell C[1]$, $\ell>2$. We want to apply the induction hypothesis to calculate $(\delta\circ\hat\varphi)(\mathbf{x})$. We provide the following lemma.

Sublemma 8.3.1. *If lemma* 8.3.12 *holds on* $\mathfrak{G}^{k-1}EC_{\mathrm{can}}[1]$ *then* $\delta\circ\hat\varphi=\hat\varphi\circ\overline{\delta}$ *as an equality of homomorphisms:* $\mathfrak{G}^kEC_{\mathrm{can}}[1]/\mathfrak{G}^1EC_{\mathrm{can}}[1]\to\mathfrak{G}^kEC[1]/\mathfrak{G}^1EC[1]$.

The proof of the sublemma is easy and is omitted.

Since $\mathfrak{m}-\mathfrak{m}_1$ is zero on $\mathfrak{G}^1EC[1]$ it follows from the sublemma and the induction hypothesis that

$$
(G\circ(\mathfrak{m}-\mathfrak{m}_1)\circ\delta\circ\hat\varphi)(\mathbf{x}) = (G\circ(\mathfrak{m}-\mathfrak{m}_1)\circ\hat\varphi\circ\overline{\delta})(\mathbf{x}).
$$

It follows from the definition that $G\circ(\mathfrak{m}-\mathfrak{m}_1)\circ\hat{\varphi} = \hat{\varphi}$. We thus obtain $\delta\circ\hat{\varphi} = \hat{\varphi}\circ\overline{\delta}$ on $\mathfrak{G}^k EC_{\mathrm{can}}[1]$. Thus by induction $\delta \circ \hat{\varphi} = \hat{\varphi} \circ \overline{\delta}$ holds.

The second formula $\overline{\delta} \circ \overline{\delta} = 0$ follows from the first one as follows. We have

$$\hat{\varphi} \circ \overline{\delta} \circ \overline{\delta} = \delta \circ \delta \circ \hat{\varphi} = 0. \tag{8.52}$$

Since φ_1 is an isomorphism and $\hat{\varphi}$ preserves the filtration $\mathfrak{G}$ it follows that $\hat{\varphi}$ is injective. Hence (8.52) implies $\overline{\delta} \circ \overline{\delta} = 0$. □

We thus constructed $\overline{\mathfrak{m}}_k : E_k C_{\mathrm{can}}[1] \to C_{\mathrm{can}}[1]$, $\varphi_k : E_k C_{\mathrm{can}}[1] \to C[1]$. It is immediate from definition that φ_1 induces an isomorphism on $\mathfrak{m}_1$ cohomology. Therefore, by theorem 8.3.2, φ_k is a homotopy equivalence of L_∞ algebras. The proof of theorem 8.3.5 is now complete. □

We next use theorem 8.3.5 (and theorem 8.3.4) to construct a versal family of deformations. We assume that the cohomology group $H^*(C, \mathfrak{m}_1)$ is finite dimensional. (We recall that we assumed R to be a field of characteristic zero.) We replace C by a canonical one C_{can} using theorem 8.3.5. Since the $\mathfrak{m}_1$ cohomology of C_{can} is isomorphic to C_{can} itself, it follows that C^k_{can} is finite dimensional. Let $\mathbf{e}_i$, $i = 1, \ldots, b_1$ be a basis of $C^1_{\mathrm{can}}[1]$ and $\mathbf{f}_i$, $i = 1, \ldots, b_2$ be a basis of $C^2_{\mathrm{can}}[1]$. We define elements $P_i \in R[[X_1, \ldots, X_{b_1}]]$, $i = 1, \ldots, b_2$ by

$$\sum_{i=1}^{b_2} P_i(X_1, \ldots, X_k)\mathbf{f}_i = \mathfrak{m}(\exp(X_1\mathbf{e}_1 + \cdots + X_{b_1}\mathbf{e}_{b_1})). \tag{8.53}$$

Here $\exp(X_1\mathbf{e}_1 + \cdots + X_{b_1}\mathbf{e}_{b_1})$ is as in (8.41). It is easy to see that P_i is well defined as a formal power series.

Definition 8.3.14. We define a pro {f. d. Alg. $/R$} object $\mathfrak{K}_{C_{\mathrm{can}}}$ by

$$\mathfrak{K}_{C_{\mathrm{can}}} \cong \frac{R[[X_1, \ldots, X_{b_1}]]}{(P_1, \ldots, P_{b_2})}.$$

We call $P_1, \ldots, P_{b_2}$ the *formal Kuranishi map*.

Lemma 8.3.13. *If two A_∞ (or L_∞) algebras C_{can} and C'_{can} are homotopy equivalent then $\mathfrak{K}_{C_{\mathrm{can}}}$ is isomorphic to $\mathfrak{K}_{C'_{\mathrm{can}}}$ as R algebras.*

Proof. Let $\mathbf{e}'_i$ and $\mathbf{f}'_j$ be the basis of $C'^0_{\mathrm{can}}[1]$, $C'^1_{\mathrm{can}}[1]$ respectively. We define $F_j(X_1, \ldots, X_{b_1})$, $j = 1, \ldots, b_1$ by

$$F_1(X_1, \ldots, X_{b_1})\mathbf{e}'_1 + \cdots + F_{b_1}(X_1, \ldots, X_{b_1})\mathbf{e}'_{b_1} = \varphi_*(X_1\mathbf{e}_1 + \cdots + X_{b_1}\mathbf{e}_{b_1}). \tag{8.54}$$

Here φ_* is defined by $e^{\varphi_* x} = \hat{\varphi}(e^x)$. Since $\varphi_1 : C^1_{\mathrm{can}}[1] \to C'^1_{\mathrm{can}}[1]$ is a linear isomorphism we may choose $\mathbf{f}'_j$ so that $\mathbf{f}'_j = \varphi_1(\mathbf{f}_j)$. It is easy to see that F^* induces an R algebra isomorphism: $R[[X'_1, \ldots, X'_{b_1}]]/(P'_1, \ldots, P'_{b_2}) \to R[[X_1, \ldots, X_{b_1}]]/(P_1, \ldots, P_{b_2})$. □

Therefore, by theorem 8.3.5 we can define $\mathfrak{K}_C = \mathfrak{K}_{C_{\mathrm{can}}}$ for any A_∞ (or L_∞) algebra C with finite dimensional $\mathfrak{m}_1$ cohomology.

We put $(C_{\mathrm{can}})_{\mathfrak{K}_{C_{\mathrm{can}}}} = C^0_{\mathrm{can}}[1] \otimes_R \mathfrak{K}_{C_{\mathrm{can}}}$ and define $\mathfrak{b} = \sum X_i \mathbf{e}_i \in (C^0_{\mathrm{can}})_{\mathfrak{K}_{C_{\mathrm{can}}}}$. Then we have $\mathfrak{m}(e^{\mathfrak{b}}) = 0 \in (C^1_{\mathrm{can}})_{\mathfrak{K}_{C_{\mathrm{can}}}}$. Here we remark that $\mathfrak{m}(e^{\mathfrak{b}})$ is an infinite sum $\sum_k \mathfrak{m}_k(\mathfrak{b}, \dots, \mathfrak{b})$ (in the L_∞ case we divide the terms by $k!$). But it converges in $(C^0_{\mathrm{can}})_{\mathfrak{K}_{C_{\mathrm{can}}}}$ with respect to the topology induced by the non-Archimedean valuation on the formal power series ring. (This means nothing but the infinite sum $\sum_k \mathfrak{m}_k(\mathfrak{b}, \dots, \mathfrak{b})$ makes sense as a formal power series.)

Thus we obtain an A_∞ (or L_∞) algebra $((C_{\mathrm{can}})_{\mathfrak{K}_{C_{\mathrm{can}}}}, \mathfrak{m}^{\mathfrak{b}})$.[11] The next lemma implies that this deformation is complete.

Lemma 8.3.14. *Let $\mathcal{R}$ be a finite dimensional local R algebra and $b \in \widetilde{\mathcal{MC}}(C_{\mathrm{can}})(\mathcal{R})$. Then there exists an R algebra homomorphism $\psi : \mathfrak{K}_{C_{\mathrm{can}}} \to \mathcal{R}$ such that $\psi(\mathfrak{b}) = b$.*

Proof. We have polynomials $R_1(Y_1, \dots, Y_m), \dots, R_N(Y_1, \dots, Y_m)$ of m variables such that $\mathcal{R} \cong R[Y_1, \dots, Y_m]/(R_1, \dots, R_N) \cong R[[Y_1, \dots, Y_m]]/(R_1, \dots, R_N)$. Let us denote by $\mathfrak{a}$ the ideal generated by $R_1, \dots, R_N$. We write

$$b \equiv F_1(Y_1, \dots, Y_m)\mathbf{e}_1 + \cdots + F_{b_1}(Y_1, \dots, Y_m)\mathbf{e}_{b_1} \mod \mathfrak{a}.$$

Here the $\mathbf{e}_i$ are the basis of C^1_{can}. We define $\tilde{\psi} : R[[X_1, \dots, X_{b_1}]] \to R[[Y_1, \dots, Y_m]]$ by $\tilde{\psi}(X_i) = F_i(Y_1, \dots, Y_m)$. Then $\tilde{\psi}(X_1\mathbf{e}_1 + \cdots + X_{b_1}\mathbf{e}_{b_1}) = b \mod \mathfrak{a}$. Therefore

$$\tilde{\psi}(\delta(\exp(X_1\mathbf{e}_1 + \cdots + X_{b_1}\mathbf{e}_{b_1}))) = \delta(e^b) = 0.$$

Hence, by definition, $\tilde{\psi}$ induces a homomorphism $\psi : \mathfrak{K}_{C_{\mathrm{can}}} \to \mathcal{R}$. □

We note that the formal Kuranishi map P_i has no term of degree ≤ 1. It follows that the Zariski tangent space $T_0 \operatorname{Spec} \mathfrak{K}_C$ can be identified with $C^1_{\mathrm{can}} = H^1(C; \mathfrak{m}_1)$. The KS map is then an identity. This fact together with lemma 8.3.14 is a formal scheme analogue of theorem 8.2.6.

We now prove theorem 8.2.6. Let us take $C^* = \Omega^{0,*}(M; \mathrm{End}(E))$ and $\mathfrak{m}_k$ is induced by $\overline{\partial}_{\mathcal{E}}$ and $\circ$ by lemma 8.3.3. Then by theorem 8.3.5 we have $\overline{\mathfrak{m}}_k : B_k C_{\mathrm{can}}[1] \to C_{\mathrm{can}}[1]$, $\varphi_k : B_k C_{\mathrm{can}}[1] \to C[1]$, where $C^k_{\mathrm{can}} = \mathrm{Ext}^k(\mathcal{E}, \mathcal{E})$.

Proposition 8.3.2.

$$\|\overline{\mathfrak{m}}_k(x_1, \dots, x_k)\| \le C^k \|x_1\| \dots \|x_k\| \tag{8.55}$$

$$\|\varphi_k(x_1, \dots, x_k)\| \le C^k \|x_1\| \dots \|x_k\| \tag{8.56}$$

where C is independent of k.

[11] In other words we have a deformation of C parametrized by a formal scheme $\operatorname{Spec} \mathfrak{K}_{C_{\mathrm{can}}}$.

Here the norm in the right-hand side of (8.56) is L^2_ℓ norm for any fixed ℓ. (The constant C in (8.56) may depend on ℓ.) The norm on C_{can} can be defined uniquely up to equivalence since C_{can} is finite dimensional.

We omit the proof of this proposition since it is straightforward to check using properties (A), (B) stated during the proof of lemma 8.2.4 and the definition of $\overline{\mathfrak{m}}_k$ and φ_k given during the proof of theorem 8.3.5. (In other words the proof is an analogue of the proof of lemma 8.2.4.)

Proposition 8.3.2 implies that the formal Kuranishi map $P : \mathrm{Ext}^1(\mathcal{E},\mathcal{E}) \to \mathrm{Ext}^2(\mathcal{E},\mathcal{E})$ actually converges in a neighbourhood of the origin. We put $\mathfrak{X}_{\mathrm{Kura}} = P^{-1}(0)$ in the sense of example 8.2.3. Then $\tilde{b}$ determines deformations parametrized by $\mathfrak{X}_{\mathrm{Kura}}$ as follows. We define

$$\begin{aligned} B_{\mathrm{Kura}}(X_1,\dots,X_\ell) &= \varphi(\exp(\tilde{b})) \\ &= \sum_k \varphi_k(X_1\mathbf{e}_1+\cdots+X_{b_1}\mathbf{e}_{b_1},\dots,X_1\mathbf{e}_1+\cdots+X_{b_1}\mathbf{e}_{b_1}). \end{aligned} \tag{8.57}$$

It is a formal power series of X_i with values in $\Omega^{0,1}(M,\mathrm{End}(E))$. By using (8.56) we find that there exists an open neighbourhood $\mathcal{U}_\ell$ of 0 in $\mathrm{Ext}^1(\mathcal{E},\mathcal{E})$ where (8.57) converges in L^2_ℓ sense for $(X_1,\dots,X_{b_1}) \in \mathcal{U}$. It is easy to see that $B_{\mathrm{Kura}} : \mathcal{U} \to \Omega^{0,1}(M;\mathrm{End}(E))$ defines a deformation of $\mathcal{E}$.[12] We have already shown that the KS map is an isomorphism for this family. Let us prove the completeness. The proof is similar to the proof of lemma 8.3.14, except that we need to discuss convergence. Let $\psi_k : B_k\Omega^{0,*}(M,\mathrm{End}(E))[1] \to \mathrm{Ext}^*(\mathcal{E},\mathcal{E})$ be a homotopy inverse of $\varphi_k : B_k C_{\mathrm{can}}[1] \to C[1]$. We recall that the existence of ψ_k follows from theorem 8.3.2. We use (8.56), and check a proof of theorem 8.3.2 carefully and obtain an estimate

$$\|\psi_k(x_1,\dots,x_k)\| \le C^k\|x_1\|\dots\|x_k\|. \tag{8.58}$$

Now let $B : \mathcal{U} \to \Omega^{0,1}(M;\mathrm{End}(E))$ be a holomorphic map which defines a deformation of $\mathcal{E}$ parametrized by $\mathfrak{X}$. Here $\mathfrak{X}$ is a germ of an analytic subspace. We consider $\psi_* \circ B : \mathcal{U} \to C^1_{\mathrm{can}} = \mathrm{Ext}^1(\mathcal{E},\mathcal{E})$. (Here ψ_* is defined by $\hat{\psi}(e^x) = e^{\psi_*(x)}$.) The map $\psi_* \circ B$ is defined first as a formal power series. We then use (8.58) and the fact that B is a convergent power series to show that $\psi_* \circ B$ converges in a small neighbourhood of the origin. By replacing $\mathcal{U}$ if necessary we may assume that it converges on $\mathcal{U}$.

Now, in the same way as in the proof of lemma 8.3.14, we find that composition of $\psi_* \circ B$ and Kuranishi map $P : \mathrm{Ext}^1(\mathcal{E},\mathcal{E}) \to \mathrm{Ext}^2(\mathcal{E},\mathcal{E})$ vanishes. Hence we obtain a ring homomorphism $\Phi : \mathcal{O}_{0,+}/(P_1,\dots,P_{b_2}) \to \mathfrak{I}_{\mathfrak{X},0}$ by $f \mapsto f \circ \psi_* \circ B$. By definition (definition 8.2.18), Φ is a morphism $\mathfrak{X} \to \mathfrak{X}_{\mathrm{Kura}}$.

To complete the proof of theorem 8.2.6, we need to show that the pullback of the deformation $B_{\mathrm{Kura}} : \mathcal{U} \to \Omega^{0,1}(M;\mathrm{End}(E))$ by Φ is isomorphic to B. The pullback of B_{Kura} by Φ is $\phi_* \circ \psi_* \circ B : \mathcal{U} \to \Omega^{0,1}(M;\mathrm{End}(E))$.

[12] We can modify the family so that B converges in C^∞ topology, see [53].

We use the fact that $\phi \circ \psi$ is homotopic to identity to show that $\phi_* \circ \psi_* \circ B$ is gauge equivalent to B as follows. By definition of homotopy (definition 8.3.8), there exists $H : \Omega^{0,*}(M; \mathrm{End}(E)) \to \Omega^{0,*}(M; \mathrm{End}(E)) \otimes \mathbb{C}[t, dt]$ such that $\mathrm{Eval}_{t=0} \circ H = \mathrm{id}$, $\mathrm{Eval}_{t=1} \circ H = \phi \circ \psi$. We can check the proof of theorem 8.3.2 carefully again to find that an estimate similar to (8.54) holds for H. Then we have a map $H_* \circ B : \mathcal{U} \to (\Omega^{0,*}(M; \mathrm{End}(E)) \otimes \mathbb{C}[t, dt])^1$, after shrinking $\mathcal{U}$ if necessary. We put $H_* \circ B = b(t,x) + c(t,x)\,dt$, where $x \in \mathcal{U}$, $b(t,x) \in \Omega^{0,1}(M; \mathrm{End}(E))$, $c(t,x) \in \Gamma(M; \mathrm{End}(E))$. We use lemma 8.3.7 here, i.e. we solve equation (8.46) to obtain $g(t)(x) \in \Gamma(M; End(E))$. We use the estimate of H to show that $g(t)(x)$ converges if x is in a neighbourhood of 0. Now $g(1)$ gives an isomorphism from the deformation B to $\phi_* \circ \psi_* \circ B$. The proof of theorem 8.2.6 is now complete. $\square$

We now turn to the proof of theorem 8.2.4 and its formal analogue. We note that the ring homomorphism $\pi : \mathfrak{K}_C \to R$ induces a homomorphism

$$\pi : H^0(C_{\mathfrak{K}_C}, \mathfrak{m}_1^{\tilde{b}}) \to H^0(C, \mathfrak{m}_1). \tag{8.59}$$

Theorem 8.3.6. *If* (8.59) *is surjective, then the functor* $\mathfrak{F}_{\mathfrak{K}_C} : \{\text{f. d. Alg.}/R\} \to \{\text{Sets}\}$ *is equivalent to the Maurer–Cartan functor* $\mathcal{MC}(C) : \{\text{f. d. Alg.}/R\} \to \{\text{Sets}\}$.

Before proving theorem 8.3.6, let us explain why it is a formal version of theorem 8.2.4. We consider the case when our canonical A_∞ (or L_∞) algebra satisfies (8.55). Then, as in the proof of theorem 8.2.6, we find a germ of analytic subvarieties $\mathfrak{X}_{\mathrm{Kura}}$ in $\mathcal{U} \subseteq \mathbb{C}^N$ defined by Kuranishi map $P_i \in \mathcal{O}_{0,+}$ and $B_{\mathrm{Kura}} = \tilde{b} : \mathcal{U} \to (C_{\mathrm{can}}[1])^0$ defines a holomorphic family of the A_∞ (or L_∞) algebras parametrized by $\mathfrak{X}_{\mathrm{Kura}}$. The operators $\mathfrak{m}_1^{\tilde{b}}$ define a holomorphic family of chain complexes

$$C^0_{\mathrm{can}} \overset{\mathfrak{m}_1^{\tilde{b}}}{\to} C^1_{\mathrm{can}} \overset{\mathfrak{m}_1^{\tilde{b}}}{\to} C^2_{\mathrm{can}} \to \cdots . \tag{8.60}$$

We note that $H^0(C_{\mathfrak{K}_C}, \mathfrak{m}_1^{\tilde{b}})$ is the set of the germs of holomorphic maps $s : \mathcal{U} \to C^0_{\mathrm{can}}$ such that $\mathfrak{m}_1^{\tilde{b}}(s)$ vanishes on $\mathfrak{X}_{\mathrm{Kura}}$. Hence the surjectivity of (8.59) implies that we have a local frame $s_1, \ldots, s_{b_0}$ of the kernel of (8.60) on $\mathfrak{X}_{\mathrm{Kura}}$ in a neighbourhood of 0. We note that the pointwise zeroth cohomology of (8.60) is semi-continuous. Hence the existence of the sections s_i implies that the pointwise zeroth cohomology of (8.60) is of constant rank in a neighbourhood of 0 in $\mathfrak{X}_{\mathrm{Kura}}$. This is the assumption of theorem 8.2.4. It is also easy to see that the assumption of theorem 8.2.4 implies the surjectivity of (8.59).

Note that the discussion here shows that theorem 8.3.6 implies theorem 8.2.4 and its analogue stated in section 8.1.6.

Proof. We now prove theorem 8.3.6. We have already constructed a natural transformation $\mathfrak{F}_{\mathfrak{K}_C} \to \mathcal{MC}(C)$. Lemma 8.3.14 implies that this transformation is surjective. So it suffices to show that it is injective. In other words, it suffices to show the following lemma.

Lemma 8.3.15. *We assume* (8.59) *is surjective. If $\varphi, \varphi' : \mathfrak{K}_C \to \mathcal{R}$ are R algebra homomorphism, and if $\varphi(\tilde{b}) \sim \varphi'(\tilde{b})$ in $\tilde{\mathcal{MC}}(C)(\mathcal{R})$, then $\varphi = \varphi'$.*

Proof. We may assume C is canonical. Then $H^0(C, \mathfrak{m}_1) = C^0$. Let $\mathbf{l}_1, \dots, \mathbf{l}_{b_0}$ be a generator of C^0. By the surjectivity of (8.59) we have its lift $\tilde{\mathbf{l}}_i$ to $C^0_{\mathfrak{K}_C}$ such that $\mathfrak{m}_1^{\tilde{b}}(\tilde{\mathbf{l}}_i) = 0$ in $C^1_{\mathfrak{K}_C}$. It is easy to see that the $\mathfrak{m}_1^{\tilde{b}}$ generate $C^0_{\mathfrak{K}_C}$ as a $\mathfrak{K}_C$ module. Hence $\mathfrak{m}_1^{\tilde{b}} : C^0_{\mathfrak{K}_C} \to C^1_{\mathfrak{K}_C}$ is zero.

Let $\mathbf{e}_i$ be a generator of C^1 and $\mathfrak{a} \subseteq R[[X_1, \dots, X_{b_1}]]$ be the ideal generated by formal Kuranishi maps. We recall $\tilde{b} \equiv \sum X_i \mathbf{e}_i \mod \mathfrak{a}$.

By assumption, we have $\hat{b} \in \tilde{\mathcal{MC}}(C)(\mathcal{R} \otimes R[t, \mathrm{d}t])$ such that $\mathrm{Eval}_{t=0}(\hat{b}) = \varphi(\tilde{b})$, $\mathrm{Eval}_{t=1}(\hat{b}) = \varphi'(\tilde{b})$. We put $\hat{b} = x(t) + y(t)\,\mathrm{d}t$. We remark that $x(t) \in (C \otimes \mathcal{R}[t])^1$ and $\mathfrak{m}(e^{x(t)}) = 0$. Let $x(t) = \sum X_i(t)\mathbf{e}_i$. Then $\tilde{\varphi}(X_i) \mapsto X_i(t)$ defines a ring homomorphism $\tilde{\varphi} : \mathfrak{K}_C \to \mathcal{R}[t]$. Its composition with obvious homomorphism $\mathrm{Eval}_{t=0} : \mathcal{R}[t] \to \mathcal{R}$ and with $\mathrm{Eval}_{t=1} : \mathcal{R}[t] \to \mathcal{R}$ are equal to φ and φ', respectively.

The condition that $\hat{b}$ is a Maurer–Cartan element implies

$$\frac{\mathrm{d}}{\mathrm{d}t} x(t) = -\mathfrak{m}_1^{b(t)}(y(t)). \tag{8.61}$$

Since $b(t) = \tilde{\varphi}(\tilde{b})$ and since $\mathfrak{m}_1^{\tilde{b}} : C^0_{\mathfrak{K}_C} \to C^1_{\mathfrak{K}_C}$ is zero it follows that the right-hand side of (8.61) is zero, i.e. $x(0) = x(t)$.

We note that $\tilde{\varphi}(\sum X_i \mathbf{e}_i) = x(t)$ and X_i generates $\mathfrak{K}_C$. Therefore $\frac{\mathrm{d}}{\mathrm{d}t}\tilde{\varphi}(t) = 0$. Hence $\varphi = \varphi'$. The proofs of lemma 8.3.15 and theorem 8.3.6 are now complete. □

Let us prove theorem 8.2.2 here. Let us take a versal family constructed in the proof of theorem 8.2.6 and write it as $(\mathfrak{K}_C, B_{\mathrm{Kura}})$. We already know that it is complete. Let $(\mathcal{U}, B)$ be another family such that the KS map is surjective. We may take a submanifold $\mathcal{V} \subset \mathcal{U}$ such that the restriction of KS map to $T_0\mathcal{V}$ is an isomorphism. Since $(\mathfrak{K}_C, B_{\mathrm{Kura}})$ is complete, we have a morphism $(\mathfrak{F}, \Phi) : (\mathcal{V}, B) \to (\mathfrak{K}_C, B_{\mathrm{Kura}})$. Using lemma 8.2.2 and the fact the KS maps of both deformations are isomorphisms, we find that $d_0\mathfrak{F} : T_0\mathcal{V} \to T_0\mathfrak{X}$ is an isomorphism. Then, using the fact that $\mathcal{V}$ is a manifold, we can apply the implicit function theorem to prove that $\mathfrak{F}$ is an isomorphism in a neighbourhood of zero. (We remark that we do not need to assume that $\mathfrak{X}$ is a manifold in a neighbourhood of zero to apply the implicit function theorem here.) Hence the $(\mathcal{V}, B)$ is isomorphic to $(\mathfrak{K}_C, B_{\mathrm{Kura}})$ and is complete. Therefore, $(\mathcal{U}, B)$ is complete. □

8.3.4 Superspace and odd vector fields—an alternative formulation of L_∞ algebras

According to [2, 65], we can rewrite the contents of sections 8.2.2, 8.2.3 using the terminology of formal geometry. In the case of L_∞ algebra we need to use a super formal manifold and in the case of the A_∞ algebra we need a kind of 'non-commutative geometry'; [21] uses the formalism of [2,65]. It is more useful in the L_∞ case than the A_∞ case, since the super manifold is closer to normal geometry than non-commutative geometry. So we only discuss the L_∞ case here. Our argument is very brief.

We start by explaining the formal super manifold (only) in the case with which we are concerned. Let us consider $V = \oplus V^k$, a graded vector space. Let V_{ev}, V_{od} be the sum of their even or odd degree parts, respectively. We regard V_{ev} as the 'bosonic' part and V_{od} as the 'fermionic' part. This means nothing other than we regard the ring of functions on it as $\prod_{k,\ell} S_k V_{\mathrm{ev}}^* \otimes \Lambda^\ell V_{\mathrm{od}}^*$, where S_k denotes the kth symmetric power. We remark that $\prod_{k,\ell} S_k V_{\mathrm{ev}}^* \otimes \Lambda^\ell V_{\mathrm{od}}^*$ is a dual to EV. (Note EV is a direct sum hence only a finite sum is allowed. Its dual is a direct product.)

Definition 8.3.15. The ring of functions on the *formal super manifold* $C[1]$ is the dual vector space $EC[1]^*$ of $EC[1]$.

Let $\mathbf{e}_i$ be a basis of $C[1]^{\mathrm{ev}} = C^{\mathrm{od}}$ and $\mathbf{f}_i$ be a basis of $C[1]^{\mathrm{od}} = C^{\mathrm{ev}}$. An element of C is written as a finite sum: $\sum x^i \mathbf{e}_i + \sum y^j \mathbf{f}_j$. Then x^i and y^j are a basis of the dual vector space C^*. Hence an element of $EC[1]^*$ can be written uniquely as

$$h = \sum_k \sum_{j_1 < \cdots < j_k} h_{j_1,\ldots,j_k}(x^1, x^2, \cdots) y^{j_1} \wedge \cdots \wedge y^{j_k} \tag{8.62}$$

where $h_{j_1,\ldots,j_k} \in R[[x^1, x^2, \ldots]]$. The ring structure is determined by $y^i \wedge y^j = -y^j \wedge y^i$ and the ring structure on $R[[x^1, x^2, \ldots]]$.

Definition 8.3.16. A formal vector field on the formal super manifold $C[1]$ can be expressed as

$$\mathfrak{V} = \sum V_i^x \frac{\partial}{\partial x^i} + \sum V_i^y \frac{\partial}{\partial y^i}$$

where $V_i \in EC[1]^*$. For h as in (8.62) and $\mathfrak{V}$ as before, we put

$$\begin{aligned}\mathfrak{V}(h) = &\sum_{k,\ell} \sum_{j_1 < \cdots < j_k} V_\ell^x \frac{\partial h_{j_1,\ldots,j_k}}{\partial x^i} y^{j_1} \wedge \cdots \wedge y^{j_k} \\ &+ \sum_k \sum_{j_1 < \cdots < j_k} \sum_{i=1}^k (-1)^{i-1} V_{j_i}^y h_{j_1,\ldots,j_k} y^{j_1} \wedge \cdots \wedge \hat{y}^{j_i} \wedge \cdots \wedge y^{j_k}.\end{aligned} \tag{8.63}$$

(8.63) is characterized by

$$\frac{\partial}{\partial x^i} x^j = \delta_{ij}$$
$$\frac{\partial}{\partial y^i} x^j = 0$$
$$\frac{\partial}{\partial y^i} y^j = \delta_{ij}$$
$$\frac{\partial}{\partial x^i} y^j = 0$$

and

$$\mathfrak{V}(hh') = \mathfrak{V}(h)h' + (-1)^{\deg \mathfrak{V} \deg h} h\mathfrak{V}(h').$$

Here we define degree by $\deg \frac{\partial}{\partial x^i} = \deg x^i = 0, -\deg \frac{\partial}{\partial y^i} = \deg y^i = 1$.

It is easy to see the following lemma.

Lemma 8.3.16. *If* $\deg \mathfrak{V}$ *and* $\deg \mathfrak{W}$ *are even then there exists a super vector field* $[\mathfrak{V}, \mathfrak{W}]$ *such that* $[\mathfrak{V}, \mathfrak{W}](h) = (\mathfrak{V}\mathfrak{W} - \mathfrak{W}\mathfrak{V})(h)$. *If* $\deg \Theta$ *and* $\deg \Xi$ *are odd then there exists a super vector field* $\{\Theta, \Xi\}$ *such that* $\{\Theta, \Xi\}(h) = (\Theta\Xi + \Xi\Theta)(h)$.

It is obvious that $[\mathfrak{V}, \mathfrak{V}] = 0$ for super vector field $\mathfrak{V}$ of *even* degree. However, in general, $\{\Theta, \Theta\} \neq 0$ for a super vector field Θ of *odd* degree. In fact we have the following lemma.

Lemma 8.3.17. *We assume two is invertible on* R. *Then, the super vector field* Θ *of degree* 1 *satisfying* $\{\Theta, \Theta\} = 0$ *corresponds one to one to the* L_∞ *structure on* $C[1]$.

Proof. Let Θ define a derivation $\Theta : EC[1]^* \to EC[1]^*$. Its dual defines a coderivation $\delta : EC[1] \to EC[1]$. $\delta\delta = 0$ is equivalent to $\{\Theta, \Theta\} = 2\Theta^2 = 0$. The lemma follows. □

Let $(C, \mathfrak{m})$, $(C', \mathfrak{m}')$ be L_∞ algebras and Θ, Θ' be odd vector fields corresponding to them by lemma 8.3.17. Let φ_k be an L_∞ homomorphism. It induces a coalgebra homomorphism $\hat{\varphi} : EC[1] \to EC'[1]$. Hence its dual is an algebra homomorphism $\hat{\varphi}^* : EC'[1]^* \to EC[1]^*$; that is a morphism of formal super manifold $C[1] \to C'[1]$. Since $\hat{\varphi} \circ \delta = \delta \circ \hat{\varphi}$ it follows that $\hat{\varphi}^* \circ \Theta' = \Theta \circ \hat{\varphi}^*$. Thus, an L_∞ homomorphism will become a morphism of the formal super manifold preserving odd vector fields on it. One may continue and translate various other operations of L_∞ algebras to the language of super manifolds. We do not attempt to do it here.

8.4 Application to mirror symmetry

8.4.1 Novikov rings and filtered A_∞, L_∞ algebras

In this chapter, we explain the relations between the discussion of deformation theory in chapters 1 and 2 and mirror symmetry.

We explain a construction in [33] which associates an 'A_∞ algebra' with a Lagrangian submanifold of a symplectic manifold (satisfying some conditions we will explain later). (We reviewed the various notions of symplectic geometry we need at the beginning of section 8.3.2.) In fact what we will associate with a Lagrangian submanifold is a slight modification of the A_∞ algebra, which we call the filtered A_∞ algebra. To define it we first need to define a universal Novikov ring. We will discuss the universal Novikov ring more in section 8.3.5. Novikov introduced a kind of formal power series ring in [80] to study the Morse theory of a closed 1-form. It was applied in [51] and others to the infinite-dimensional situation of Floer homology (which may be regarded as a Morse theory of closed 1-form on loop space), i.e. the Floer homology, which we will discuss in section 8.3.3, is defined as a module over a Novikov ring. In order to use the same ring independently of the symplectic manifold (and of its Lagrangian submanifold) we use a ring which we call the universal Novikov ring. We now define it.

Let R be a commutative ring. We consider the formal sum $x = \sum_i a_i T^{\lambda_i}$ satisfying the following conditions.

Condition 8.4.1. (1) $a_i \in R$, (2) $\lambda_i \in \mathbb{R}$, (3) $\lambda_i < \lambda_{i+1}$, (4) $\lim_{i\to\infty} \lambda_i = \infty$.

Definition 8.4.1. The set of all formal sums $x = \sum_i a_i T^{\lambda_i}$ satisfying conditions 8.4.1 is called the *universal Novikov ring* and is written as $\Lambda_{R,\text{nov}}$. This becomes a ring (R algebra) by an obvious definition of sum and multiplication. We replace condition 8.4.1(2) by $\lambda_i \geq 0$. We then obtain a subring $\Lambda_{R,\text{nov},0}$. We replace condition 8.4.1(2) by $\lambda_i > 0$. We then obtain an ideal $\Lambda_{R,\text{nov},+}$ of $\Lambda_{R,\text{nov}}$. (We omit R to avoid confusion.)

When R is a field, $\Lambda_{R,\text{nov},0}$ is a local ring with maximal ring $\Lambda_{R,\text{nov},+}$.

We define a filtration $\mathfrak{F}$ on Λ_{nov} by

$$\mathfrak{F}^\lambda \Lambda_{\text{nov}} = \{x | x \text{is as in (3.1) satisfying condition 8.4.1 and } \lambda_i \geq \lambda\}.$$

$\mathfrak{F}^\lambda \Lambda_{\text{nov}}$ is a filtration, i.e. $\mathfrak{F}^\lambda \Lambda_{\text{nov}}$ is a sub-Abelian group (with respect to $+$) and $\mathfrak{F}^\lambda \Lambda_{\text{nov}} \cdot \mathfrak{F}^{\lambda'} \Lambda_{\text{nov}} \subseteq \mathfrak{F}^{\lambda+\lambda'} \Lambda_{\text{nov}}$. $\mathfrak{F}$ induces a filtration on $\Lambda_{\text{nov},0}$ and $\Lambda_{\text{nov},+}$.

Remark 8.4.1. $\Lambda_{\text{nov},0}$ is not a Noether ring, since the ascending sequence of ideals $\mathfrak{F}^{1/i} \Lambda_{\text{nov}}$ does not stop. This fact makes it harder to study an algebra or module over it.

A filtered $\Lambda_{\text{nov},0}$ module is a $\Lambda_{\text{nov},0}$ module C together with filtration $\mathfrak{F}^\lambda C$ such that $\mathfrak{F}^\lambda \Lambda_{\text{nov}} \cdot \mathfrak{F}^{\lambda'} C \subseteq \mathfrak{F}^{\lambda+\lambda'} C$. We say that a Λ_{nov} module homomorphism $\varphi : C \to C'$ is a filtered $\Lambda_{\text{nov},0}$ module homomorphism if $\varphi(\mathfrak{F}^\lambda C) \subset \mathfrak{F}^\lambda C'$.

Filtration defines a metric on Λ_{nov} and a module on it by $d(x, y) = \exp(-\inf\{\lambda | x - y \in \mathfrak{F}^\lambda C\})$. Λ_{nov}, $\Lambda_{\text{nov},0}$ and $\Lambda_{\text{nov},+}$ are complete with respect to this metric. From now on we *assume* all filtered $\Lambda_{\text{nov},0}$ modules are complete. We also *assume* that all filtered A_∞ (or L_∞) algebras are the completions of a free $\Lambda_{\text{nov},0}$ module.

If C, C' are filtered $\Lambda_{\mathrm{nov},0}$ modules, we define a filtration on their tensor product $C \otimes_{\Lambda_{\mathrm{nov},0}} C'$ by

$$\mathfrak{F}^{\lambda}(C \otimes_{\Lambda_{\mathrm{nov},0}} C') = \bigcup_{\mu} \mathfrak{F}^{\mu} C \otimes_{\Lambda_{\mathrm{nov},0}} \mathfrak{F}^{\lambda-\mu} C'.$$

The tensor product $C \otimes_{\Lambda_{\mathrm{nov},0}} C'$ is not complete with respect to the metric induced by this filtration. We denote the completion by $C \hat{\otimes}_{\Lambda_{\mathrm{nov},0}} C'$.

A graded filtered $\Lambda_{\mathrm{nov},0}$ module is defined in an obvious way. Let C be a graded filtered $\Lambda_{\mathrm{nov},0}$ module. We consider

$$B_k C[1] = \underbrace{C[1] \hat{\otimes}_{\Lambda_{\mathrm{nov},0}} \cdots \hat{\otimes}_{\Lambda_{\mathrm{nov},0}} C[1]}_{k \text{ times}}$$

Let $\hat{B}C[1]$ be the completion of the direct sum $\oplus_k B_k C[1]$. $E_k C[1]$ and $\hat{E}C[1]$ are defined by taking its submodule which is invariant under the action of $\mathfrak{S}_k$.

They are formal coalgebras. Here formal coalgebra is defined by replacing $\otimes$ with $\hat{\otimes}$ in the definition of a coalgebra. We can define coderivation and cohomomorphism for a formal coalgebra in the same way. (We assume them to be filtered.) The following analogy of lemmata 8.3.2, 8.3.5 holds.

Lemma 8.4.1. *Let $f_k : B_k C[1] \to C[1]$, $k = 0, 1, \ldots$, be a sequence of filtered homomorphisms of degree* 1. *Then there exists a unique coderivation $\delta : \hat{B}C[1] \to \hat{B}C[1]$ whose restriction to $B_k C[1]$ is f_k.*

Let $\varphi_k : B_k C[1] \to C'[1]$, $k = 0, 1, \ldots$, be a sequence of filtered homomorphisms of degree 0. *We assume $\varphi_0(\Lambda_{\mathrm{nov},0}) \subseteq \mathfrak{F}^{\lambda_0} B_1 C'[1]$ for some positive λ_0. Then there exists a coalgebra homomorphism $\hat{\varphi} : \hat{B}C[1] \to \hat{B}C[1]$ whose* $\mathrm{Hom}(B_C k[1], B_1 C'[1])$ *component is φ_k.*

The same statement holds when we replace B by E.

We note that we include f_0 and φ_0 here but not in lemmata 8.3.2 and 8.3.5.

Proof. We prove the A_∞ case only. We put

$$\begin{aligned}
&\hat{f}_0(x_1 \otimes \cdots \otimes x_k) \\
&\qquad = \varphi_0(1) \otimes x_1 \otimes \cdots \otimes x_k + (-1)^{\deg' x_1} x_1 \otimes \varphi_0(1) \otimes x_2 \otimes \cdots \otimes x_k \\
&\qquad\quad + \cdots + (-1)^{\deg' x_1 + \cdots + \deg' x_k} x_1 \otimes \cdots \otimes x_k \otimes \varphi_0(1).
\end{aligned}$$

We define $\hat{f}_k$, $k \geq 1$ in the same way as in the proof of lemma 8.3.2. Then we can prove that $\delta = \hat{f}_0 + \cdots + \hat{f}_k + \cdots$ converges by using the fact that f_i preserves filtration.

We next define $\hat{\varphi}$. We put

$$e^{\varphi(1)} = \sum_k \underbrace{\varphi(1) \otimes \cdots \otimes \varphi(1)}_{k \text{ times}}.$$

By assumption it converges in $\hat{B}C'[1]$. Now we put

$$\hat{\varphi}(\mathbf{x}) = \sum_k \sum_a e^{\varphi(1)} \otimes \overline{\varphi}(\mathbf{x}_a^{k;1}) \otimes e^{\varphi(1)} \otimes \cdots \otimes e^{\varphi(1)} \otimes \overline{\varphi}(\mathbf{x}_a^{k;k}) \otimes e^{\varphi(1)}.$$

Here $\overline{\varphi}$ is φ_k on $B_kC[1]$ $k \neq 0$ and is zero on $B_0C[1]$. It is easy to check that $\hat{\varphi}$ converges and is a coalgebra homomorphism. □

Remark 8.4.2. $\hat{B}C[1]$ has another filtration different from $\mathfrak{F}^\lambda \hat{B}C[1]$, i.e. we put $\mathfrak{G}^k \hat{B}C[1] = \oplus_{i \le k} B_iC[1]$. We call the filtration $\mathfrak{F}$ the *energy filtration* and $\mathfrak{G}$ the *number filtration*. If $f_0 \neq 0$ or $\varphi_0 \neq 0$, $\hat{f}$ or $\hat{\varphi}$ does *not* preserve the number filtration. They preserve the energy filtration. We also note that $\hat{B}C[1]$ is not complete with respect to the number filtration.

Definition 8.4.2. A structure of a *filtered A_∞ algebra* on a filtered graded $\Lambda_{\mathrm{nov},0}$ module C is a series of filtered homomorphisms $\mathfrak{m}_k : B_kC[1] \to C[1]$, $k = 0, 1, \ldots$ of degree 1 such that $\delta\delta = 0$ where $\delta = \sum \hat{\mathfrak{m}}_k : \hat{B}_kC[1] \to C[1]$ is obtained by lemma 8.4.1. We also assume that $\mathfrak{m}_0(\Lambda_{\mathrm{nov},0}) \subseteq \mathfrak{F}^{\lambda_0} B_1C[1]$.

A sequence of homomorphisms $\varphi_k : B_kC[1] \to C'[1]$ is a *filtered A_∞ homomorphism* between filtered A_∞ algebras if $\varphi_0(\Lambda_{\mathrm{nov},0}) \subseteq \mathfrak{F}^{\lambda_0} B_1C'[1]$ and the homomorphism $\hat{\varphi}$ obtained by lemma 8.4.1 satisfies $\delta\hat{\varphi} = \hat{\varphi}\delta$.

L_∞ can be defined in a similar way.

We will explain how to modify the argument in the previous section to our filtered situation later after we have introduced our main example.

8.4.2 Review of a part of global symplectic geometry

In this section, we review several points on global symplectic geometry, especially those related to pseudoholomorphic curves, which we need for our main construction; [52, 75, 76] are standard references for them.

A *symplectic manifold* is a pair (M, ω) where ω is a closed 2-form on M such that it is non-degenerate as an antisymmetric 2-form on T_pM for each $p \in M$. M is automatically even dimensional. Let $2n$ be its dimension. Then ω^n is a nowhere vanishing $2n$ form on M and hence determines an orientation. A *Lagrangian submanifold* of a symplectic manifold (M, ω) is an n-dimensional closed submanifold L such that $\omega|_L = 0$. If the dimension of a submanifold of M is strictly larger than n then the restriction of ω to L cannot vanish.

A typical example of a symplectic manifold is a Kähler manifold. In particular, if M is a projective variety that is a complex submanifold of $\mathbb{C}P^n$ then it is a symplectic manifold, whose symplectic structure is obtained as a pullback of the Fubini–Study form ω on $\mathbb{C}P^n$ which is defined by $\pi^*\omega = -4\sqrt{-1}\partial\overline{\partial}\log(|z^0|^2 + \cdots + |z^n|^2)$ where $\pi : \mathbb{C}^{n+1} \to \mathbb{C}P^n$ is the projection.

Another important example of a symplectic manifold is a cotangent bundle T^*N of any smooth manifold N. The symplectic form on T^*N is given by $\mathrm{d}\theta$.

Here θ is a 1-form on T^*N such that $\theta(X) = u(\pi_*(X))$ where $X \in T_uT^*N$ and $\pi : T^*N \to N$ is the natural projection. If $x^1, \dots, x^n$ is a coordinate of N then elements of T^*N are written as $p_1\,\mathrm{d}x^1 + \cdots + p_n\,\mathrm{d}x^n$. Hence p_i, x^j, $i, j = 1, \dots, n$ is a coordinate of T^*N. Using this coordinate our symplectic form ω on T^*N is $\omega = \mathrm{d}p_1 \wedge \mathrm{d}x^1 + \cdots + \mathrm{d}p_n \wedge \mathrm{d}x^n$.

An example of a Lagrangian submanifold of a projective variety is a submanifold consisting of real valued points. Let $M \subseteq \mathbb{C}P^n$ be a complex submanifold which is preserved by complex conjugation τ. We assume $L = \{x \in M | \tau(x) = x\}$ is a submanifold of dimension $n = \dim M/2$. Then we can show L is a Lagrangian submanifold.

There are also various examples of Lagrangian submanifolds in T^*N. One is a conormal bundle $T^*_K N$ of a submanifold K of N. Here

$$T^*_K N = \{(x, u) \in T^*M | x \in N, u|_{T_x N} = 0\}.$$

In fact, if K is defined by equations $x^{k+1} = \cdots = x^n = 0$ then $T^*_K N$ is defined by equations $x^{k+1} = \cdots = x^n = 0$, $p_1 = \cdots = p_k = 0$. Hence ω is zero on $T^*_K N$.

Another example is a graph of a closed 1-form u which is defined as follows. For a 1-form u on N, we put $G_u = \{(x, u(x)) | x \in N\}$. Let us define a diffeomorphism $i : N \to G_u$ by $i(x) = (x, u(x))$. We can show that the pullback of θ by i is u itself. It follows that $i^*\omega = \mathrm{d}u$. Therefore G_u is a Lagrangian submanifold if and only if u is closed.

Symplectic geometry has a long history. There are many interesting results and applications. However, for a long time, it seems that there were only a few results in symplectic geometry which were really global in nature. For example, the following question was open for a long time. Does a pair of symplectic manifolds (M, ω) and (M, ω') on a same manifold M such that $[\omega] = [\omega'] \in H^2(M; \mathbb{R})$ but there is no diffeomorphism $\varphi : M \to M$ with $\varphi^*\omega' = \omega$ exist. The reason why such results were not known seems to me that there was basically no general technique which could be applied to study global symplectic geometry. Arnold in the 1960s formulated a series of conjectures which are related to the global problem of symplectic geometry (see, for example, [3]). Roughly speaking, those questions ask whether a 'symplectic topology' exists. Around the beginning of the 1980s several works appeared which show that such a 'symplectic topology' does exist and is extremely rich. Among these results, Gromov's in [42] is quite remarkable. In this, Gromov introduced a new technique, the pseudoholomorphic curve, in order to study the global structure of a symplectic manifold. Let us briefly review it here. Let (M, ω) be a symplectic manifold. An almost complex structure $J : TM \to TM$ is a tensor such that $JJ = -1$.

Definition 8.4.3. J is said to be compatible with ω if (1) $\omega(JX, JY) = \omega(X, Y)$, (2) $\omega(X, JX) > 0$ and (3) $\omega(X, JX) = 0$ implies $X = 0$.

It is proved in [42] (see [75, 76]) that a compatible almost complex structure always exists and the set of compatible almost complex structures is contractible.[13]

Gromov's idea was to apply the techniques of complex geometry to an almost complex manifold (M, J) to get information on the symplectic manifold (M, ω).

There are basically two methods in complex geometry: one uses holomorphic functions or holomorphic maps defined on M and the other uses holomorphic maps to M. When (M, J) is almost complex (i.e. J is not integrable), there are not so many holomorphic functions on M. Hence the first method is hard to apply in our case of (M, J).

Gromov's important observation was that, even when (M, J) is not integrable, there are many holomorphic maps to (M, J) from the Riemann surface (a complex one-dimensional manifold). The basic reason for this is that any almost complex structure on a real two-dimensional manifold is automatically integrable.

The method then initiated by Gromov was to study the moduli space of holomorphic maps from the Riemann surface to (M, J) to get information about (M, ω). Gromov called the holomorphic map from the Riemann surface to an almost complex manifold, a *pseudoholomorphic curve*. The contractibility of the set of almost complex structures guarantees that any invariant obtained by using a compatible almost complex structure is an invariant of the symplectic manifold if it is independent of any continuous change in the compatible almost complex structures.

Using the existence of a symplectic structure compatible with J, Gromov proved various compactness results for the moduli space of pseudoholomorphic maps, hence its fundamental cycle, in principle, defines such an invariant. Ruan [88] made this construction (which was somewhat implicit in [42]) more explicit.

This invariant, in turn, was found to be an invariant of the topological σ-model with target space M, which Witten [106] introduced in an informal way. The invariant obtained in this way is now called the *Gromov–Witten invariant*. See [34, 76, 88, 89] for more about it.

Our main concern here is its relative version, i.e. we consider a Lagrangian submanifold L of M and study a map $\varphi : D^2 \to M$ such that the following condition holds.

Condition 8.4.2. (1) φ is pseudoholomorphic. Namely $J \circ \mathrm{d}\varphi = \mathrm{d}\varphi \circ j_{D^2}$. Here j_{D^2} is the standard complex structure of D^2. (2) $\varphi(\partial D^2) \subseteq L$.

Gromov [42] had already studied the moduli space of such φ to obtain information about the Lagrangian submanifolds in $\mathbb{C}^n$ (for example he proved that a simply connected compact Lagrangian submanifold in $\mathbb{C}^n$ does not exist).

[13] It follows that the Chern classes of (the tangent bundle of) a symplectic manifold is well defined.

Floer [22] used a similar idea to study problems of the intersection of Lagrangian submanifolds and, in particular, the following problem due to Arnold [3].

Problem 8.4.1. Let $L \subset M$ be a Lagrangian submanifold, and let $\varphi : M \to M$ be a Hamiltonian diffeomorphism (which we will define later). We assume L is transversal to $\varphi(L)$. Then, under 'some' condition, we have an estimate

$$\sharp L \cap \varphi(L) \geq \sum \operatorname{rank} H_k(L; \mathbb{Z}_2). \tag{8.64}$$

Let us define a Hamiltonian diffeomorphism. Let (M, ω) be a symplectic manifold and f be a function on it. There exists a vector field X_f such that $\omega(X_f, V) = \mathrm{d}f(V)$ holds for any vector V. X_f is called the *Hamiltonian vector field*. We now consider $f : M \times [0, 1] \to \mathbb{R}$. Let $f_t(x) = f(x, t)$. It induces a one-parameter family of vector fields X_{f_t}. We define a family of diffeomorphisms $\varphi_t : M \to M$ by

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi_t(x) = X_{f_t}(\varphi_t(x)) \qquad \varphi_0(x) = x. \tag{8.65}$$

Definition 8.4.4. $\varphi : M \to M$ is called a *Hamiltonian diffeomorphism* if there exists $f_t(x) = f(x, t)$ such that $\varphi_1 = \varphi$, and φ_t is defined by (8.65).

One can prove easily that a Hamiltonian diffeomorphism is a symplectic diffeomorphism, i.e. $\varphi^*\omega = \omega$.

Floer [22] proved (8.64) for $\pi_2(M, L) = 0$. He used a new homology theory which is now called the *Floer homology* for this purpose. Let us briefly explain it here. Let L_1, L_2 be two Lagrangian submanifolds such that $\pi_2(M, L_i) = 0$. We assume that L_1 is transversal to L_2. We consider the $\mathbb{Z}_2$ vector space $CF(L_1, L_2)$ whose basis is identified with the intersection points $p \in L_1 \cap L_2$, i.e. we put

$$CF(L_1, L_2) = \bigoplus_{p \in L_1 \cap L_2} \mathbb{Z}_2[p]. \tag{8.66}$$

Floer defined the degree of each $[p]$ and defined a boundary operator on it as follows. Let $p, q \in L_1 \cap L_2$. We consider maps $\varphi : D^2 \to M$ satisfying the following conditions. We put $\partial D^2_+ = \{z \in \partial D | \operatorname{Im} z > 0\}$, $\partial D^2_- = \{z \in \partial D | \operatorname{Im} z < 0\}$.

Condition 8.4.3.

(1) φ is pseudoholomorphic, i.e. $J \circ \mathrm{d}\varphi = d\varphi \circ j_{D^2}$. Here j_{D^2} is the standard complex structure of D^2.
(2) $\varphi(\partial_+ D^2) \subset L_1$, $\varphi(\partial_- D^2) \subset L_2$.
(3) $\varphi(-1) = p$, $\varphi(1) = q$.

Let $\widetilde{\mathcal{M}}(p, q; L_1, L_2)$ be the moduli space of all such maps φ. The group $\operatorname{Aut}(D^2; \{\pm 1\})$ of biholomorphic maps $D^2 \to D^2$ preserving ± 1 acts on

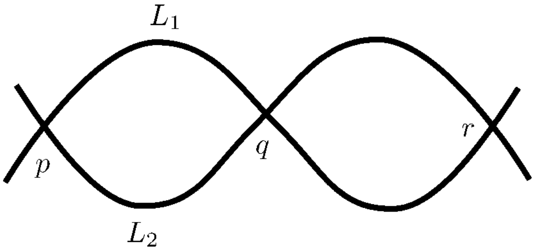

Figure 8.3.

$\widetilde{\mathcal{M}}(p, q; L_1, L_2)$. This group is isomorphic to $\mathbb{R}$. We denote the quotient space by $\mathcal{M}(p, q; L_1, L_2)$. Floer proved the following theorem.

Theorem 8.4.1. *Under the assumption* $\pi_2(M, L_i) = 0$*, there exists* $\mu : L_1 \cap L_2 \to \mathbb{Z}$ *such that the following hold after taking a 'generic perturbation'*[14] *of the pseudoholomorphic curve equation* $J \circ \mathrm{d}\varphi = \mathrm{d}\varphi \circ j_{D^2}$.

(1) $\mathcal{M}(p, q; L_1, L_2)$ *is a smooth manifold of dimension* $\mu(q) - \mu(p) - 1$.
(2) *If* $\mu(q) - \mu(p) - 1 = 0$*, then* $\mathcal{M}(p, q; L_1, L_2)$ *consists of finitely many points.*
(3) *If* $\mu(q) - \mu(p) - 1 = 1$*, then* $\mathcal{M}(p, q; L_1, L_2)$ *can be compactified to* $\mathcal{CM}(p, q; L_1, L_2)$ *which is a one-dimensional manifold with boundary.*
(4) *In* (3) *the boundary of* $\mathcal{CM}(p, q; L_1, L_2)$ *is identified with*

$$\bigcup_{r \in L_1 \cap L_2, \mu(r) = \mu(q)+1} \mathcal{M}(p, r; L_1, L_2) \times \mathcal{M}(r, q; L_1, L_2). \tag{8.67}$$

(Note that (8.67) is of finite order by (2).) We will not discuss the proof of this theorem which is now a classic. μ is called the *Maslov–Viterbo index.*

(4) can be explained by figure 8.3.

Now the definition of a Floer homology is as follows. We put $\deg[p] = \mu(p)$ hence $CF(L_1, L_2)$ is a graded $\mathbb{Z}_2$ vector space. We next define

$$\langle \delta[p], [q] \rangle \equiv \sharp \mathcal{M}(p, q; L_1, L_2) \mod 2$$

for $\mu(q) - \mu(p) - 1 = 0$ and put

$$\delta[p] = \sum \langle \delta[p], [q] \rangle [q].$$

δ is an operator of degree 1. We show the following corollary.

Corollary 8.4.1. $\delta \circ \delta = 0$.

[14] We do not explain its precise meaning in this article.

Proof. Let us calculate the coefficient of $[q]$ in $\delta\delta([p])$. We may write it as

$$\sum_r \langle \delta[p], [r]\rangle \langle \delta[r], [q]\rangle.$$

It suffices to consider the case $\mu(q) - \mu(p) - 1 = 1$, i.e. we can apply (4) of theorem 8.4.1. We then have

$$\begin{aligned}
&\sum_r \langle \delta[p], [r]\rangle \langle \delta[r], [q]\rangle \\
&\quad = \bigcup_{r \in L_1 \cap L_2, \mu(r) = \mu(q)+1} \sharp(\mathcal{M}(p, r; L_1, L_2) \times \mathcal{M}(r, q; L_1, L_2)) \\
&\quad = \sharp\partial\mathcal{M}(p, q; L_1, L_2) \equiv 0 \mod 2
\end{aligned}$$

since the order of the boundary of a one-dimensional compact manifold is even. □

We thus define a *Floer cohomology* by

$$HF(L_1, L_2) = H(CF(L_1, L_2), \delta).$$

Floer proved the following two properties of it.

Theorem 8.4.2. *We assume $\pi_2(M, L_i) = 0$. If φ_i are Hamiltonian diffeomorphisms then $HF(L_1, L_2) \cong HF(\varphi_1 L_1, \varphi_2 L_2)$.*

Theorem 8.4.3. *We assume $\pi_2(M, L) = 0$. Then $HF(L, L) \cong H(L; \mathbb{Z}_2)$.*

It is easy, from definition, to see that

$$\operatorname{rank} HF(L_1, L_2) \geq \sharp L_1 \cap L_2. \tag{8.68}$$

(8.64) follows from theorems 8.4.2, 8.4.3 and (8.68).

Our discussion so far has assumed $\pi_2(M, L) = 0$. After Floer, Oh [81] relaxed the condition $\pi_2(M, L) = 0$. His assumption is that the Lagrangian submanifold is monotonic and its minimal Maslov number is ≥ 3. We do not explain this condition here.

However, there is an example where (8.64) does not hold in general.

Example 8.4.1. Let us consider S^2. Any one-dimensional submanifold of it is a Lagrangian submanifold. Let L be a circle which is in a small neighbourhood of the north pole. We can easily find a Hamiltonian diffeomorphism φ such that $\varphi(L) \cap L = \emptyset$. However, $H_*(L) = H_*(S^1) \neq 0$.

Therefore, something should go wrong if we try to generalize Floer's theory to more general Lagrangian submanifolds. For example, in the case of example 8.4.1 it turns out that the Floer homology is not defined for such a Lagrangian submanifold.

Thus finding a good condition for defining a Floer homology is important for its application to global symplectic geometry. Later it was found that the same problem is also closely related to mirror symmetry. This is the main point of our article. So before stating the results (and conjectures) precisely, we first explain the rough outline.

Oh, Ohta, Ono[15] and myself have developed an obstruction theory for the well-definedness of the Floer homology of a Lagrangian submanifold [33]. (Our project started around 1997 just after the necessary analytic machinery had been completed in [34].) We found that there is a series of obstructions which take values in the cohomology of L such that if they all vanish then the Floer homology is well defined.

However, a Floer homology thus defined actually *depends* on the various choices involved; i.e. there exists a moduli space associated with a Lagrangian submanifold and the Floer homology is well defined as a family parametrized by this moduli space. The condition that the obstruction vanishes is equivalent to the condition that this moduli space is non-empty. The algebraic machinery we need to establish it is the one we developed in the previous section (its filtered version precisely), i.e. the moduli space parametrizing the Floer homology is a moduli space representing the appropriate Maurer–Cartan functor.

However, I proposed generalizing the Lagrangian intersection Floer homology to the case in which there are three or more Lagrangian submanifolds.[16] Then, in the early 1990s I found that an A_∞ structure appears [23]. Then Kontsevich conjectured that the A_∞ structure on a Floer homology of Lagrangian submanifold should be a 'mirror' of a similar A_∞ structure on a sheaf cohomology of a complex manifold (i.e. the A_∞ algebra related to the complex $\Omega^{0,*}(\mathrm{End}(E))$ introduced in section 8.1 and 8.2), i.e. Kontsevich proposed a *homological mirror symmetry conjecture* in [66, 69]. It roughly states that there are pairs of symplectic manifolds M and complex manifolds $M^\wedge$ such that the Lagrangian submanifolds of M correspond to coherent sheafs on $M^\wedge$ and the Floer homology of Lagrangian submanifolds in M corresponds to the sheaf cohomology on $M^\wedge$. Moreover, the A_∞ structure on the Floer homology on M corresponds to the Yoneda and Massey–Yoneda product on the sheaf cohomology on $M^\wedge$.

The homological mirror symmetry conjecture was developed before the 'second string theory revolution'. Later, Brane theory became important. The homological mirror symmetry conjecture can then be regarded naturally as corresponding to branes and as part of various dualities. After that and after Strominger *et al*'s important proposal [99] to construct a mirror manifold by using a dual torus fibration, several people began to be interested in homological mirror symmetry. Among them, Polishchuk and Zaslow [85] proved part of the theory for an elliptic curve. (Kontsevich [66] had discussed the case of elliptic curve

[15] Kontsevich gave us an important suggestion to start this research.

[16] I was inspired by an idea from Donaldson and Segal when I started this project.

earlier.) Just after that I generalized it partially to a complex torus of higher dimension [27]. The main tool used in [27] is a result from [32] which calculates the A_∞ structure of a Floer homology for a cotangent bundle. Some more papers on Abelian mirror symmetry appeared after that. There are other important works by Seidel and his coauthors (such as those in [57, 93–95]) which are also related to the homological mirror symmetry conjecture.

While studying complex tori, I found that various interesting and delicate phenomena happen in the Floer homology of Lagrangian submanifolds, its product (A_∞) structure, and also its family version [31]. The study of Floer homology is thus tied more with homological algebra and with deformation theory, which we explained in sections 8.1 and 8.2. The homological mirror symmetry conjecture has now become more precise since it has been involved in homological algebra and deformation theory. For example, we conjecture the coincidence of the two moduli spaces, one for deformation of the Floer homology of Lagrangian submanifolds and the other for deformation of coherent sheaves or vector bundles.

8.4.3 From Lagrangian submanifold to A_∞ algebra

Now, after this brief explanation of its history, let us discuss the construction of filtered A_∞ algebras associated with Lagrangian submanifolds. Here we consider the case in which we have only one Lagrangian submanifold L rather than a pair of Lagrangian submanifolds as in Floer's case. (Two or more Lagrangian submanifolds are mentioned at the end of section 8.3.4 and discussed in detail in [30, 33].)

The condition $\pi_2(M, L) = 0$ which we assumed before implies that there are no maps $\varphi : (D^2, \partial D^2) \to (M, L)$ satisfying condition 8.4.2. In fact, since φ is zero homotopic it follows that $\int_{D^2} \varphi^*\omega = 0$. We can easily show that if φ is pseudoholomorphic and is non-constant then $\int_{D^2} \varphi^*\omega > 0$. This is the basic reason why theorem 8.4.2 holds. In other words, the presence of pseudoholomorphic disc $\varphi : (D^2, \partial D^2) \to (M, L)$ deforms the usual homology group $H(L)$ to the Floer homology group $HF(L)$ (see section 8.3.4).

Let us discuss the moduli space of pseudoholomorphic discs satisfying condition 8.4.2. Let $\beta \in \pi_2(M, L)$. We use the following moduli space.

Definition 8.4.5. The moduli space $\mathcal{M}_{k+1}(L; \beta)$ is the set of all $\sim$ equivalence classes of pairs $(\varphi, \vec{z})$ where

(1) $\varphi : (D^2, \partial D^2) \to (M, L)$ satisfies condition 8.4.2;
(2) he homotopy class of φ is β; and
(3) $\vec{z} = (z_0, \dots, z_k)$ where $z_i \in \partial D^2$.

We assume that $z_0, \dots, z_k$ respects the cyclic order of ∂D^2.

We say $(\varphi, \vec{z}) \sim (\varphi', \vec{z}')$ if there exists a biholomorphic automorphism $u : D^2 \to D^2$ such that $\varphi' = \varphi \circ u$, $z_i = u(z'_i)$.

The basic task we need to carry out in order to apply the moduli space $\mathcal{M}_{k+1}(L;\beta)$ to various problems is:

(A) find an appropriate compactification $\mathcal{CM}_{k+1}(L;\beta)$;
(B) find an appropriate perturbation of the pseudoholomorphic curve equation $J \circ d\varphi = d\varphi \circ j_{D^2}$, so that the moduli space $\mathcal{CM}_{k+1}(L;\beta)$ will become a 'smooth manifold' after perturbation (transversality);
(C) calculate the dimension of $\mathcal{CM}_{k+1}(L;\beta)$ (index theory); and
(D) find a condition on M, L under which $\mathcal{CM}_{k+1}(L;\beta)$ is oriented (orientation).

These are the package of results one needs to establish topological field theory by a nonlinear partial differential equation and the moduli space of its solutions. (Donaldson [18] first used such a package to establish the invariants of 4-manifolds (Donaldson invariants). Gromov applied Donaldson's idea to the moduli space of pseudoholomorphic curves (from a closed Riemann surface).)

For point (A), we can easily modify Kontsevich's [67] notion of a stable map so that it can be applied to the case of a Riemann surface with boundary (disc), see [33]. For point (B), there is now a general theory (developed in [34]) which can be applied to various situations in a uniform way. Hence basically there is nothing new to work out but we can just apply [34].[17] The key notion we use to carry out (B) is a *space with a Kuranishi structure*. We will explain this informally later. In fact, it is a smooth analogue of the notion of complex analytic space discussed in section 1.6.

Now the package (A), (B), (C), (D) in our situation can be stated as follows.

Theorem 8.4.4. [33] *There exists $\mu : \pi_2(M; L) \to \mathbb{Z}$ (the Maslov index) with the following properties.*

(1) $\mathcal{CM}_{k+1}(L;\beta)$ *is a compact space with a Kuranishi structure (with corners), of dimension $\mu(\beta) + n + k - 1$.*
(2) $\mathcal{CM}_{k+1}(L;\beta)$ *is oriented if L is relatively spin, in the sense defined later. The relative spin structure determines the orientation of $\mathcal{CM}_{k+1}(L;\beta)$.*
(3) *The boundary of $\mathcal{CM}_{k+1}(L;\beta)$ is described by fibre products of various $\mathcal{CM}_{k'+1}(L;\beta')$ where $k' \le k$ and $\omega \cap \beta' \le \omega \cap \beta$.*

Explanation of the notion used in the statement of theorem 8.4.4 will follow (after several remarks). Statement (3) is a bit vague since we do not mention which fibre product appears. We do not explain this since the main focus of this article is on algebraic formalism and we want to minimize the explanation of the geometric analysis.

[17] There is one point to clarify about the transversality of $\mathcal{CM}_{k+1}(L;\beta)$ which is not included in the general theory: that is the problem of transversality at diagonal $\subset L^{k+1}$. This is point (2) mentioned in the discussion after theorem 8.4.5, and is handled in [33].

Remark 8.4.3. We explain two more properties of the Maslov index μ given in theorem 8.4.4. Let $\pi_2(M) \to \pi_2(M, L) \to \pi_1(L)$ be part of a homotopy exact sequence of the pair M, L. Then the composition $\pi_2(M) \to \pi_2(M, L) \xrightarrow{\mu} \mathbb{Z}$ coincides with $2c^1(M)$. Here $c^1(M) \in H^2(M; \mathbb{Z})$ determines $\pi_2(M) \to \mathbb{Z}$. Also $\pi_2(M) \to \pi_2(L) \xrightarrow{w^1} \mathbb{Z}_2$ coincides with μ modulo 2. Here $w^1 : \pi_1(L) \to \{0, 1\}$ is the first Stiefel–Whitney class of L, i.e. $w^1(\gamma) = 0$ if the orientation of TL is preserved along the loop γ. In particular, μ is even valued if L is oriented.

Let us now explain the notions used in theorem 8.4.4. We will explain two notions: the relative spin structure and a space with Kuranishi structure with corners. To define the relative spin structure, let us recall some well-known facts about vector bundles on manifolds. First, for any vector bundle E, there exists a characteristic class $w^2(E) \in H^2(L; \mathbb{Z}_2)$, the Stiefel–Whitney class, such that the structure group of E is reduced to a spinor group if and only if $w^2(L) = 0$. Second, a real vector bundle E on a 3-manifold is trivial if it is oriented and spin (i.e. $w_1(E) = w_2(E) = 0$).

Definition 8.4.6. Let L be a submanifold of M. Then L is said to be relatively spin if L is oriented and if there exists a cohomology class $st \in H^2(M; \mathbb{Z}_2)$ such that $w^2(TL) = i^*st$.

We take an oriented vector bundle E on M such that $w^2(E) = st$. Then the *relative spin structure* of L is a trivialization of $TL \oplus E$ on the two skeleton of L.

In the case when L is spin, we may take $st = 0$. Hence the relative spin structure is in one-to-one correspondence with the spin structure of L. We refer the reader to [33] for more detail on the orientation of our moduli space.

We next briefly explain a Kuranishi structure with corners. We consider an open neighbourhood $\mathcal{U} \subset \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}_{\geq 0}$ of 0. Let $F = (f^1, \ldots, f^m) : \mathcal{U} \to \mathbb{R}^m$ be a smooth map. We consider the set $F^{-1}(0)$. As in an analytic subset and analytic subspace (which we discussed in section 1.6), the subset $F^{-1}(0)$ of $\mathcal{U}$ does not contain enough information if $\mathrm{d}f^1, \ldots, \mathrm{d}f^m$ are not linearly independent. For example, in the simplest situation, we need to count the order of such a space (that is a moduli space of pseudoholomorphic curves). Let us suppose that $n_1 = 1, n_2 = 0, m = 1$ and $f(x) = x^2$. (Here $x \in \mathbb{R}$.) Set theoretically, $f^{-1}(0)$ consists of one point. But if we perturb the equation $x^2 = 0$ slightly and consider $x^2 = \epsilon$, then the number of solutions is zero (if we count them by sign). Hence we need to 'remember' the additional information from the equation $F(x) = 0$ itself. So we need something more than a subset of $\mathcal{U}$. When F is a complex analytic function, this is exactly the analytic space explained in section 8.1.6. When F is a polynomial, this is the idea of a scheme. What we need here is its C^∞ analogue. (Working in the C^∞ category is inevitable in studying moduli space of pseudoholomorphic curves, especially with a Lagrangian boundary condition, since our problem is strictly a real one and it is impossible to assume the complex analyticity of the equation.) Working in the C^∞ category, considering the ring of germs at 0 and dividing it by ideals generated by f^i does not seem to work. This is because the ring of smooth functions does not have some nice properties which are enjoyed by the ring of holomorphic functions. So instead of considering rings (or ringed spaces), we regard the pair $(\mathcal{U}, F)$ itself as a *chart* of our 'space'. We then define the appropriate notion of coordinate change (or, equivalently, define a way to glue charts). We thus obtain a space with an approximate Kuranishi structure.

More precisely, we need to include one more feature; i.e. in general, our moduli space is not given as $F^{-1}(0)$ but as $F^{-1}(0)/\Gamma$ locally. Let us explain the notation. Γ is a finite group. We assume that there are linear actions of Γ on $\mathbb{R}^{n_1}$ and on $\mathbb{R}^m$. We assume that this action preserves $\mathcal{U}$. Moreover, we assume that F is Γ equivariant. Hence the zero point set $F^{-1}(0)$ has a Γ action and we may consider the quotient space $F^{-1}(0)/\Gamma$. By the Kuranishi structure in theorem 8.4.4, we mean an object gluing such triples $(\mathcal{U}, F, \Gamma)$ in an appropriate sense. (Hence those objects are a smooth analogue of a Deligne and Mumford stack,[18] see [15].) We do not try to define what we mean by gluing $(\mathcal{U}, F, \Gamma)$ (see [34]).

We said that our Kuranishi structure is one with corners, since $\mathcal{U}$ is an open subset of $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}_{\geq 0}$. We can define a boundary of a space with Kuranishi structure with corners.

Once we have obtained an oriented Kuranishi structure with corners, we can define its fundamental chain. It is a chain not a cycle, in general, since we are studying an analogy of a manifold with boundary or corners. We note that, already in theorem 8.4.1(3), the moduli space $\mathcal{CM}(p, q; L_1, L_2)$ was a one-dimensional manifold *with a boundary*. Studying its boundary is the main part of the proof of the basic equality $\delta\delta = 0$. The same situation will occur in our more general setting. This is the main difference between the pseudoholomorphic curve from a closed Riemann surface (where everything can be discussed at the level of homology) and our case of a pseudoholomorphic curve from a disc (where we need to work in the chain level). This causes various technical problems which are treated in [33].

Now that we have finished our brief explanation of the statement of theorem 8.4.4 we apply it to construct an A_∞ algebra.

Let L be a Lagrangian submanifold of M. We assume that L is relatively spin and fix a relative spin structure. Then the orientation of the moduli spaces $\mathcal{CM}_{k+1}(L; \beta)$ is induced. We next assume that the Maslov index $\mu : \pi_2(M, L) \to \mathbb{Z}$ is zero. Then, the dimension of our moduli space $\mathcal{CM}_{k+1}(L; \beta)$ is $n + k - 1$ and is independent of β.

Remark 8.4.4. This, in particular, implies that $c^1(M)$ is zero on $\pi_2(M)$ by remark 8.4.3. The Calabi–Yau manifold has this property. We also note that if L is a special Lagrangian submanifold (see [99]) of the Calabi–Yau manifold then the Maslov index $\mu : \pi_2(M, L) \to \mathbb{Z}$ is zero (see, for example, [33] for its

[18] In some other situation like gauge theory, we need to consider the case when Γ is of positive dimension. It will then be an analogy of an Artin stack.

proof). The special Lagrangian submanifold in a Calabi–Yau manifold is believed to be the most important case in mirror symmetry.

Definition 8.4.7. The evaluation map $ev = (ev_0, \ldots, ev_k) : \mathcal{CM}_{k+1}(L;\beta) \to L^{k+1}$ is defined by $ev[\varphi, \vec{z}] = (\varphi(z_0), \ldots, \varphi(z_k))$ where $\vec{z} = (z_0, \ldots, z_k)$.

As we have already mentioned, theorem 3.3.1 and the general theory of Kuranishi structure developed in [34] imply that a fundamental chain $[\mathcal{CM}_{k+1}(L;\beta)] \in S_{n+k-1}(\mathcal{CM}_{k+1}(L;\beta))$ exists. Here we may regard it as a singular chain. We use it to define an operator $\mathfrak{m}_k$.

We first take a countably generated complex of singular chains of L over $\mathbb{Q}$. (The choice of this subcomplex is based on a delicate technical argument which we do not mention here.) We write it as $S(L)$. We consider the tensor product $C^*(L) = S_{n-*}(L)\hat{\otimes}_{\mathbb{Q}}\Lambda_{\mathbb{Q},\mathrm{nov},0}$. Here $\hat{\otimes}_{\mathbb{Q}}$ is a completion with respect to the metric induced by the filtration on $\Lambda_{\mathbb{Q},\mathrm{nov},0}$. We are going to define a structure of filtered A_∞ algebra on it.

Now we define $\mathfrak{m}_k$ as follows. Let us take an element P_i of $C^{d_i}(L)$. It is a chain of degree $n - d_i$. We now take the fibre product

$$\mathcal{CM}_{k+1}(L;\beta) \times_{(ev_1,\ldots,ev_k)} (P_1 \times \cdots \times P_k).$$

It is a $\mathbb{Q}$ chain of dimension $n + k - 2 - \sum d_i$. We use the evaluation map ev_0 to regard it as a chain in L. We thus obtain

$$ev_*(\mathcal{CM}_{k+1}(L;\beta) \times_{(ev_1,\ldots,ev_k)} (P_1 \times \cdots \times P_k)) \in S_{n+k-2-\sum d_i}(L). \tag{8.69}$$

We regard (8.69) as an element of $C^{\sum d_i+2-k}(L)$ and write it as $\mathfrak{m}_{k,\beta}(P_1, \ldots, P_k)$. When $\beta = 0$ we need to define $\mathfrak{m}_{k,\beta}$ in a slightly different way. Roughly speaking we 'put'

$$\mathfrak{m}_{1,0}(P) = \partial P \qquad \mathfrak{m}_{2,0}(P_1, P_2) = P_1 \cap P_2. \tag{8.70}$$

However, (8.70) itself is not correct as we will mention later.

Definition 8.4.8. We put

$$\mathfrak{m}_k(P_1, \ldots, P_k) = \sum_\beta T^{[\omega]\cap\beta}\mathfrak{m}_{k,\beta}(P_1, \ldots, P_k) \tag{8.71}$$

and extends it to a $\Lambda_{\mathbb{Q},\mathrm{nov},0}$ module homomorphism.

Then the main theorem of [33] is as follows.

Theorem 8.4.5. $\mathfrak{m}_k$ *defines a structure of filtered* A_∞ *algebra on* $C^*(L)$.

Before mentioning various serious and delicate points in the rigorous argument to justify definition 8.4.8 and proving theorem 8.4.5, let us explain parts of the arguments which are easier to explain.

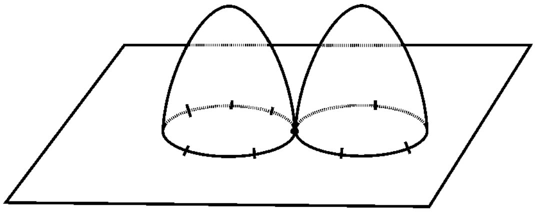

Figure 8.4.

(A) We first need to show that (8.71) converges in $C^*(L)$. This is non-trivial since there are infinitely many terms involved. However, we can prove it by using Gromov compactness; i.e. Gromov compactness implies that, for each E, there is only a finitely many β such that $\int_\beta \omega < E$ and that $\mathcal{CM}_{k+1}(L;\beta)$ is non-empty. This means that modulo $\mathfrak{F}^E\mathcal{CM}_{k+1}(L;\beta)$ there is only finitely many terms in (8.71), where $\mathfrak{F}^E\mathcal{CM}_{k+1}(L;\beta)$ is the energy filtration. This implies the convergence of (8.71).

(B) We next check that the degree is correct. The degree of P_i is $d_i - 1$ after it has shifted. However, the degree of the right-hand side of (8.71) is $\sum d_i - k + 1$ after it has shifted. Hence the degree of $\mathfrak{m}_k$ (after shifting) is 1 as required.

(C) Next we check the condition that $\mathfrak{m}_0 \equiv 0$ modulo $\Lambda_{\mathbb{Q},\mathrm{nov},0}$. This is immediate from $\mathfrak{m}_{0,\beta} = 0$ if $\beta = 0$.

(D) The proof of the fact that $\mathfrak{m}_k$ satisfies the A_∞ relation is based on theorem 8.4.4(3) and is roughly as follows. We study $\mathfrak{m}_{1,0} \circ \mathfrak{m}_{k,\beta}$. Since $\mathfrak{m}_{1,0}$ is the usual boundary operator by (8.70), it follows that this composition is obtained by using the boundary of the moduli space $\mathcal{CM}_{k+1}(L;\beta)$. Theorem 8.4.4(3) asserts that the boundary of $\mathcal{CM}_{k+1}(L;\beta)$ is described as a fibre product of various similar moduli spaces. Taking the fibre product of the moduli spaces corresponds to taking a composition of the operators obtained by it. Hence by looking at which kinds of fibre product appear in the compactification, we find the A_∞ relation. Roughly the boundary of the moduli space $\mathcal{CM}_{k+1}(L;\beta)$ is described by figure 8.4.

Now we mention more delicate parts of the argument to justify definition 8.4.8 and prove theorem 8.4.5. We discuss them only briefly since these points are not our main concern in this article.[19]

(1) We need to specify the orientation to define the right-hand side of (8.69) as a $\mathbb{Q}$ chain. Basically the relative spin structure gives a way to define orientation of $\mathcal{CM}_{k+1}(L;\beta)$. Moreover, P_i (which is Poincaré dual to a cochain over $\mathbb{Q}$) is co-oriented. So we obtain a co-orientation of the fibre product. However

[19] However, I would emphasize that these 'technical details' (which took myself together with Oh, Ohta, Ono almost 5 years to work out) are the main part of the theory and asserting results without working this kind of detail out is extremely dangerous.

it is rather a delicate problem to handle the orientation of the fibre product and check that the A_∞ formula is correct with sign. Actually we needed more than 60 pages in [33] for this purpose. It is preferable to find a simpler argument.

(2) Formula (8.70) itself cannot actually be justified. The reason is that P is not transversal to P so we can not put $\mathfrak{m}_{2,0}(P, P) = P \cap P$. This is the problem of transversality at the diagonal. The way to overcome this is as follows. We perturb the diagonal and define operations $\mathfrak{m}_{k,0}$ inductively on k so that we can define it at the chain level on a countably generated subcompex $S(L)$ of singular chain complex and it is an A_∞ algebra homotopy equivalent to the De-Rham complex. In other words, we need to use an A_∞ algebra for $\mathfrak{m}_{k,0}$, which corresponds to the rational homotopy theory.[20]

(3) We need to take a complex $S(L)$ carefully so that it is countable (otherwise we cannot use Baire's category theorem to achieve transversality) and the right-hand side of (8.69) is again contained in the same complex.

After this brief explanation of the proof of theorem 8.4.5, we continue our story. Our next task is to state that the A_∞ algebra in theorem 8.4.5 is independent of the various choices involved. Precisely it is invariant up to homotopy equivalence. Let us define the homotopy equivalence of filtered A_∞ algebra. It is similar to the usual A_∞ algebra described in section 8.2. A few points need modification, which we explain here.

Let C be a filtered A_∞ algebra. We first define a filtered A_∞ algebra $C[1]\hat{\otimes}_{\Lambda_{R,\mathrm{nov}}} \Lambda_{R,\mathrm{nov}}[t, \mathrm{d}t]$. The definition is almost the same as definition 8.3.7. One important difference is that we take completion of the tensor product here; i.e. the element of $C[1]\hat{\otimes}_{\Lambda_{R,\mathrm{nov}}} \Lambda_{R,\mathrm{nov}}[t, \mathrm{d}t]$ is written as $P(t) + Q(t)\,\mathrm{d}t$ where $P(t)$, $Q(t)$ are the *infinite* sums $P(t) = \sum t^i P_i$, $Q(t) = \sum t^i Q_i$ where $P_i, Q_i \in \mathfrak{F}^{\lambda_i} C$ with $\lambda_i \to \infty$. The operation $\mathfrak{m}_k$ on $C[1]\hat{\otimes}_{\Lambda_{R,\mathrm{nov}}} \Lambda_{R,\mathrm{nov}}[t, \mathrm{d}t]$ is defined in the same way as definition 8.3.7.

We use $C[1][t, \mathrm{d}t]$ in place of $C[1]\hat{\otimes}_{\Lambda_{R,\mathrm{nov}}} \Lambda_{R,\mathrm{nov}}[t, \mathrm{d}t]$ hereafter.

We can define a filtered A_∞ homomorphism $\mathrm{Eval}_{t+t_0} : C[1][t, \mathrm{d}t] \to C[1]$. In the same way as in (8.38); let $P(t) + Q(t)\,\mathrm{d}t \in \Lambda_{R,\mathrm{nov}}[t, \mathrm{d}t]$. We put

$$\mathrm{Eval}_{t+t_0}(P(t) + Q(t)\,\mathrm{d}t) = \sum t_0^i P_i \tag{8.72}$$

(8.72) is an infinite sum but it converges in $C[1]$.

Now we can define a homotopy between filtered A_∞ (or L_∞) algebras in the same way as in definition 8.3.8. Theorem 8.3.1 holds in the case of a filtered A_∞ (or L_∞) algebra. The definition of homotopy equivalence is also the same as in definition 8.3.9.

Now we have the following theorem.

[20] It has been realized by several specialists in surgery theory that transversality at the diagonal is one of the most essential points in differential topology.

Theorem 8.4.6. *The A_∞ algebra is independent of the various choices involved (for example to compatible almost complex structures) up to homotopy equivalence. If φ is a symplectic diffeomorphism then $(C^*(L), \mathfrak{m}_*)$ is homotopy equivalent to $(C^*(\varphi(L)), \mathfrak{m}_*)$.*

We omit the proof which is in [33].[21]

8.4.4 Maurer–Cartan equation for filtered A_∞ algebras

In this section, we explain the way to modify the argument of section 8.2.3 for a filtered A_∞ algebra, especially the filtered A_∞ algebra $(C^*(L), \mathfrak{m}_*)$ in section 8.3.3. (We only discuss the case of filtered A_∞ algebra C to save notation. All the arguments are parallel for filtered L_∞ algebras.) The Maurer–Cartan equation is

$$\delta(e^b) = 0 \qquad \text{where } b \in \cup_{\lambda>0}\mathfrak{F}^\lambda C^1. \tag{8.73}$$

Since $e^b = \sum_{k=0}^\infty \underbrace{b \otimes \cdots \otimes b}_{k \text{ times}}$ converges as an element of $\hat{B}C[1]$, the left-hand side of (8.73) makes sense. This point is different from section 8.2.3.

In the same way as in section 8.2.3, the solution of (8.73) defines a filtered A_∞ algebra $(C, \mathfrak{m}^b)$ by

$$\mathfrak{m}_k^b(x_1, \ldots, x_k) = \mathfrak{m}(e^b, x_1, e^b, \ldots, e^b, x_k, e^b).$$

$(C, \mathfrak{m}^b)$ is a filtered A_∞ algebra if $b \in \cup_{\lambda>0}\mathfrak{F}^\lambda C^1$, without assuming $\delta(e^b) = 0$. The Maurer–Cartan equation $\delta(e^b) = 0$ is equivalent to $\mathfrak{m}_0^b = 0$, i.e. it is equivalent to the condition that $(C, \mathfrak{m}^b)$ is an A_∞ algebra.

Let us elaborate the assumption $b \in \cup_{\lambda>0}\mathfrak{F}^\lambda C^1$. Let us reduce the coefficient ring of our filtered A_∞ algebra C to $R = \Lambda_{\mathrm{nov},0}/\Lambda_{\mathrm{nov},+}$ and obtain $\overline{C} = C \otimes_{\Lambda_{\mathrm{nov},0}} R$. Then $\overline{C}$ together with induced operations $\overline{\mathfrak{m}}_k$ is an A_∞ algebra over R. (Note $\overline{\mathfrak{m}}_0 = 0$ by our assumption that $\mathfrak{m}_0 \equiv 0 \mod \Lambda_{\mathrm{nov},+}$.)

Hence $(C, \mathfrak{m})$ is a 'deformation' over $\Lambda_{\mathrm{nov},0}$ in the sense similar to that in definition 8.2.26. (However, since $\mathfrak{m}_0 \neq 0$ it is not strictly so.) The condition $b \in \cup_{\lambda>0}\mathfrak{F}^\lambda C^1$ implies $(C, \mathfrak{m}^b) \otimes_{\Lambda_{\mathrm{nov},0}} R \cong (\overline{C}, \overline{\mathfrak{m}})$, i.e. $(C, \mathfrak{m}^b)$ is a deformation of $(\overline{C}, \overline{\mathfrak{m}})$. When b satisfies the Maurer–Cartan equation, $\mathfrak{m}_0^b = 0$, it is a deformation of $(\overline{C}, \overline{\mathfrak{m}})$ in the sense of definition 8.2.26 strictly.

Thus when we study the set of solutions of (8.73), we are studying the moduli space of deformations, i.e. deformations of deformations. We will explain, in the next section, that studying these is natural in mirror symmetry.

We expand equation (3.10) and obtain

$$\mathfrak{m}_0(1) + \mathfrak{m}_1(b) + \mathfrak{m}_2(b, b) + \cdots = 0. \tag{8.74}$$

[21] The definition of homotopy equivalence we gave in the December 2000 version of [33] looks different from the one we gave here. We proved theorem 8.4.6 in [33] using the definition there. We will rewrite the proof of it and prove the homotopy equivalence in the sense defined here in the final version of [33]. The two definitions are actually equivalent to each other.

Another important difference between (8.74) and (8.39) is that (8.74) is inhomogeneous (i.e. there is a term $\mathfrak{m}_0(1)$). As a consequence $b = 0$ is not a solution of (8.74). Actually there are cases where (8.74) has no solutions.

Definition 8.4.9. We say that filtered A_∞ algebra C is *unobstructed* if (8.74) has a solution. We say that b is a *bounding chain* (or Maurer–Cartan element) of C if (8.73) is satisfied.

We now define the gauge equivalence of solutions of (8.73) as follows.

Definition 8.4.10. Let b, b' be bounding chains of C. Then b is said to be gauge equivalent to b' (and is written as $b \sim b'$), if there exists $\tilde{b}$ a bounding chain of $C[t, \mathrm{d}t]$ such that $\mathrm{Eval}_{t=0}(\tilde{b}) = b$, $\mathrm{Eval}_{t=1}(\tilde{b}) = b'$.

We can prove that $b \sim b'$, $b' \sim b''$ imply $b \sim b''$ in a similar way to the proof of theorem 8.3.3 (see the final version of [33]). Lemma 8.3.10 can be generalized to our situation in the same way.

Definition 8.4.11. We denote by $\widetilde{\mathcal{MC}}(C)$ the sets of all bounding chains of filtered A_∞ algebra C. The set of the gauge equivalence class of bounding chains is denoted by $\mathcal{MC}(C)$.

Theorem 8.4.7 follows from what have we already explained. Let us consider a category {filtered A_∞ alg./R} whose object is a filtered A_∞ algebra over R and whose morphism is a filtered A_∞ homomorphism. We consider its quotient category[22] {filtered A_∞ alg./R}/homotopy whose object is a homotopy equivalence class of filtered A_∞ algebra over R and whose morphism is a homotopy class of an A_∞ homomorphism.

Theorem 8.4.7. *$C \mapsto \mathcal{MC}(C)$ induces a functor:* {filtered A_∞ alg./R}/homotopy $\to$ {Sets}.

Theorem 8.4.7 implies that the set $\mathcal{MC}(C)$ is a homotopy type invariant of C. However, 'invariant as sets' does not mean very much. This is the reason we state theorem 8.4.7 using the quotient category as before. A better way to state the homotopy invariance of $\mathcal{MC}(C)$ is given later in proposition 8.4.2.

To study $\mathcal{MC}(C)$ we use the canonical model as in section 8.2.3.

Definition 8.4.12. A filtered A_∞ algebra $(C, \mathfrak{m})$ is said to be *canonical* if $\mathfrak{m}_0 = 0$ and $\mathfrak{m}_1 \equiv 0 \mod \Lambda_{\mathrm{nov},+}$.

To generalize theorem 8.3.5 we assume a kind of finiteness condition for our A_∞ algebra C.

Definition 8.4.13. A filtered A_∞ algebra C is said to be *weakly finite* if it is homotopy equivalent to C' which is finitely generated as a $\Lambda_{\mathrm{nov},0}$ module.

[22] See, for example, [35, 47, 55] for its definition.

We need one more assumption. We recall that we have assumed that C is a completion of a free $\Lambda_{\mathrm{nov},0}$ module (as a $\Lambda_{\mathrm{nov},0}$ module). So, as the $\Lambda_{\mathrm{nov},0}$ module, we have $C \cong \overline{C} \otimes \Lambda_{\mathrm{nov},0}$. Hence, an R module homomorphism $\overline{f}_k : B_k\overline{C}[1] \to \overline{C}[1]$ induces a filtered $\Lambda_{\mathrm{nov},0}$ module homomorphism f_k : $B_kC[1] \to C[1]$.

Definition 8.4.14. A filtered A_∞ algebra C is said to be strongly gapped if there exists λ_i and $\overline{\mathfrak{m}}_{k,i} : B_k\overline{C}[1] \to \overline{C}[1]$ such that

$$\mathfrak{m}_k = \sum_i T^{\lambda_i}\mathfrak{m}_{k,i} \qquad \lim_{i\to\infty} \lambda_i = \infty. \tag{8.75}$$

Our main example $(C(L), \mathfrak{m}_k)$ is strongly gapped by definition. We proved in [33, section 8.A4], that it is weakly finite.

Theorem 8.4.8. *For any weakly finite, strongly gapped, and unobstructed filtered A_∞ algebra, there exists a canonical filtered A_∞ algebra homotopy equivalent to it.*

We note that if we require $\mathfrak{m}_1 = 0$ in the definition of a canonical filtered A_∞ algebra, then we cannot prove theorem 8.4.8. This is because we need an analogue of lemma 8.3.11, which does not hold over $\Lambda_{\mathrm{nov},0}$ but only over a field.[23]

Proof (sketch). We consider $\mathfrak{m}_{1,0} = \overline{\mathfrak{m}}_0 \otimes 1$ as in (8.75) (here we put $\lambda_0 = 0$). We note that $\mathfrak{m}_{1,0} \circ \mathfrak{m}_{1,0} = 0$ follows from $\mathfrak{m}_0 \equiv 0 \mod \Lambda_{\mathrm{nov},+}$. We use the assumption that R is a field to obtain a decomposition $\mathrm{Ker}\,\mathfrak{m}_{1,0} = \mathrm{Im}\,\mathfrak{m}_{1,0} \oplus H(C; \mathfrak{m}_{1,0})$. We put $C^*_{\mathrm{can}} = H^*(C; \mathfrak{m}_{1,0})$. We then obtain a propagator G_k such that $G_{k+1} \circ \mathfrak{m}_{1,0} + \mathfrak{m}_{1,0} \circ G_k = 1 - \Pi_{C^k_{\mathrm{can}}}$.

The construction of the structure $\mathfrak{m}^{\mathrm{can}}_k$ of filtered A_∞ algebra on C_{can} and of filtered A_∞ homomorphism $\varphi_k : B_kC_{\mathrm{can}} \to C$ then goes in a similar way to that in the proof of theorem 8.3.5.

There are two differences, however: first, we use graphs such that the interior vertex may have one or two edges (in condition 8.3.1, we have assumed the interior vertices have at least three edges); and second, we assign λ_i to each vertex.

Each interior vertex of our graph then corresponds to a term $\mathfrak{m}_{k,\lambda_i}$ in (8.75) with $(k, \lambda_i) \neq (1, 0)$. The rest of the construction of $\mathfrak{m}^{\mathrm{can}}_k$, φ_k is similar to the proof of theorem 8.3.5 and is omitted (see the final version of [33]).

To complete the proof, we need the following two results. Let $\varphi : C \to C'$ be a filtered A_∞ homomorphism of a weakly finite strongly gapped A_∞ algebra. We assume C is unobstructed and let $b \in C$ be a bounding chain.

Theorem 8.4.9. *If $\varphi_1 : H(C, \mathfrak{m}^b_1) \to H(C', \mathfrak{m}^{\varphi_* b}_1)$ is an isomorphism, then φ is a homotopy equivalence.*

[23] However, if we use the field Λ_{nov} as a coefficient ring, we can prove a lemma similar to lemma 8.3.11. However, the propagator G obtained over the Λ_{nov} coefficient does not preserve energy filtration. As a consequence, if we try to define operators $\overline{\mathfrak{m}}_k$ using G, it does not converge.

Proposition 8.4.1. *If* $\overline{\varphi}_1 : H(\overline{C}, \overline{\mathfrak{m}}_1^b) \to H(\overline{C}', \overline{\mathfrak{m}}_1^{\overline{\varphi}_* b})$ *is an isomorphism, then* $\varphi_1 : H(C, \mathfrak{m}_1^b) \to H(C', \mathfrak{m}_1^{\varphi_* b})$ *is an isomorphism.*

Theorem 8.4.9 is an analogue of theorem 8.3.3 and can be proved in the same way. Proposition 8.4.1 will follow from a spectral sequence which we explain later (theorem 8.4.11). It is easy to see that $\varphi_k : B_k C_{\text{can}} \to C$ satisfies the assumption of proposition 8.4.1. Hence it is a homotopy equivalence. □

We can show that homotopy equivalences between canonical A_∞ algebras are isomorphisms in the same way as in proposition 8.3.1.

We now use theorem 8.4.8 to define a formal Kuranishi map as follows. Let C be a canonical A_∞ algebra. Let $\mathbf{e}_i$ and $\mathbf{f}_i$ be the bases of C^1 and C^2, respectively. We put b_i equal to $\text{rank}_R \overline{C}^i$. We take formal parameters $X_1, \ldots, X_{b_1}, X_1^\epsilon, \ldots, X_{b_1}^\epsilon$ and put

$$\Lambda_{\text{nov},0}[[X_1, \ldots, X_{b_1}]] = \Lambda_{\text{nov},0} \hat{\otimes}_R R[[X_1, \ldots, X_{b_1}]]$$
$$\Lambda_{\text{nov},0}\langle X_1^\epsilon, \ldots, X_{b_1}^\epsilon \rangle = \Lambda_{\text{nov},0} \hat{\otimes}_R R[X_1^\epsilon, \ldots, X_{b_1}^\epsilon].$$

We note that $\Lambda_{\text{nov},0}\langle X_1^\epsilon, \cdots, X_{b_1}^\epsilon \rangle$ consists of the elements

$$\sum a_{i_1,\ldots,i_{b_1}} X_1^{i_1} \ldots X_{b_1}^{i_{b_1}}$$

such that $a_{i_1,\ldots,i_{b_1}} \in \mathfrak{F}^{\lambda_{i_1,\ldots,i_{b_1}}} \Lambda_{\text{nov},0}$ with

$$\lim_{\min\{i_1,\ldots,i_{b_1}\} \to \infty} \lambda_{i_1,\ldots,i_{b_1}} = \infty.$$

$\Lambda_{\text{nov},0}\langle X_1^\epsilon, \ldots, X_{b_1}^\epsilon \rangle$ is called a *strictly convergent ring* in rigid analytic geometry (see [12]).

We define $P_j \in \Lambda_{\text{nov},0}[[X_1, \ldots, X_{b_1}]]$, $P_j^\epsilon \in \Lambda_{\text{nov},0}\langle X_1^\epsilon, \ldots, X_{b_1}^\epsilon \rangle$ by

$$\sum P_j(X_1, \ldots, X_{b_1}) \mathbf{f}_j = \mathfrak{m}(\exp(X_1 \mathbf{e}_1 + \cdots + X_{b_1} \mathbf{e}_{b_1}))$$
$$\sum P_j^\epsilon(X_1^\epsilon, \ldots, X_{b_1}^\epsilon) \mathbf{f}_j, = \mathfrak{m}(\exp(T^\epsilon X_1^\epsilon \mathbf{e}_1 + \cdots + T^\epsilon X_{b_1}^\epsilon \mathbf{e}_{b_1})).$$

Definition 8.4.15. We put $\mathfrak{K}_C = \Lambda_{\text{nov},0}[[X_1, \ldots, X_{b_1}]]/(P_1, \ldots, P_{b_2})$. We also put: $\mathfrak{K}_C^\epsilon = \Lambda_{\text{nov},0}\langle X_1^\epsilon, \ldots, X_{b_1}^\epsilon \rangle/(P_1^\epsilon, \ldots, P_{b_2}^\epsilon)$.

For $\epsilon < \delta$ we define a homomorphism $\pi_{\epsilon,\delta} : \Lambda_{\text{nov},0}\langle X_1^\epsilon, \ldots, X_{b_1}^\epsilon \rangle \to \Lambda_{\text{nov},0}\langle X_1^\delta, \ldots, X_{b_1}^\delta \rangle$ by $\pi_{\epsilon,\delta}(X_i^\epsilon) = T^{\delta-\epsilon} X_i^\delta$. It induces $\pi_{\epsilon,\delta} : \mathfrak{K}_C^\epsilon \to \mathfrak{K}_C^\delta$. We then define $\mathfrak{K}_C^+ = \varprojlim \mathfrak{K}_C^\epsilon$.

We can easily prove the following proposition.

Proposition 8.4.2. *The isomorphism classes (as* $\Lambda_{\text{nov},0}$ *algebras) of* $\mathfrak{K}_C$, $\mathfrak{K}_C^\epsilon$, $\mathfrak{K}_C^+$ *are independent of the homotopy equivalence of* C.

We define $\tilde b = \sum X_i \mathbf{e}_i$, $\tilde b^\epsilon = \sum T^\epsilon X_i^\epsilon \mathbf{e}_i$. $\tilde b^\epsilon$ defines $\tilde b^+ \in \mathfrak{K}_C^+$. They satisfy Maurer–Cartan equation (8.73). Therefore $(C_{\mathfrak{K}_C}, \mathfrak{m}^{\tilde b})$, $(C_{\mathfrak{K}_C^\epsilon}, \mathfrak{m}^{\tilde b^\epsilon})$, $(C_{\mathfrak{K}_C^+}, \mathfrak{m}^{\tilde b^+})$ are A_∞ algebras. (Here we put $C_{\mathfrak{K}_C} = C\hat\otimes_{\Lambda_{\mathrm{nov},0}}\mathfrak{K}_C$ etc.)

$\mathfrak{K}_C$ is a complete local ring whose maximal ideal is generated by T^λ (for all $\lambda > 0$) and X_i. Hence it parametrizes a deformation of $\overline{C}$ which is an A_∞ algebra over R.

However, $\mathfrak{K}_C^\epsilon$ is not a local ring, i.e. its spectrum has many (closed) points. This is equivalent to the fact that we can define $f_{a_1,\dots,a_{b_1}} : \Lambda_{\mathrm{nov},0}\langle X_1^\epsilon, \dots, X_{b_1}^\epsilon\rangle \to \Lambda_{\mathrm{nov},0}$ by $f_{a_1,\dots,a_{b_1}}(P(X_1^\epsilon, \dots, X_{b_1}^\epsilon)) = P(a_1, \dots, a_{b_1})$. (In other words $\mathrm{Spec}(\mathfrak{K}_C^\epsilon)$ is infinitesimally small in T direction but is of positive size in the X_i direction.) Hence $\mathfrak{K}_C^+$ is not a local ring either. We consider the ideal $\Lambda_{\mathrm{nov},+}\mathfrak{K}_C^+$ of it and put $\overline{\mathfrak{K}}_C^+ = \mathfrak{K}_C^+/(\Lambda_{\mathrm{nov},+}\mathfrak{K}_C^+)$. We then consider a deformation $C_{\overline{\mathfrak{K}}_C^+} = (C_{\mathfrak{K}_C^+}, \mathfrak{m}^{\tilde b^+}) \otimes_{\mathfrak{K}_C^+} \overline{\mathfrak{K}}_C^+$. $C_{\overline{\mathfrak{K}}_C^+}$ is a deformation of $\overline{C}$ and is the restriction of the family $(C_{\mathfrak{K}_C^+}, \mathfrak{m}^{\tilde b^+})$ to its 'sub space' defined by $T = 0$. We can easily see that $C_{\overline{\mathfrak{K}}_C^+}$ is a trivial deformation, i.e. it is isomorphic $(\overline{C}, \overline{\mathfrak{m}}) \otimes_R \overline{\mathfrak{K}}_C^+$. Thus, $(C_{\mathfrak{K}_C^+}, \mathfrak{m}^{\tilde b^+})$ is a deformation of C whose restriction to $T = 0$ is trivial.

$(C_{\mathfrak{K}_C}, \mathfrak{m}^{\tilde b})$, $(C_{\mathfrak{K}_C^+}, \mathfrak{m}^{\tilde b^+})$ have the following completeness properties (lemma 8.4.2) similar to those in lemma 8.3.14. To state the lemma, we need some notation.

Let {filtered complete Artin local$/\Lambda_{\mathrm{nov},0}$} be the category of filtered complete Artin local $\Lambda_{\mathrm{nov},0}$ algebras. {filtered complete$/\Lambda_{\mathrm{nov},0}$} is the category of filtered complete $\Lambda_{\mathrm{nov},0}$ algebras. (They can be defined in the same way as in definition 8.2.29.)

To each filtered A_∞ algebra C (which may not be canonical), we define functors $\mathcal{MC}(C)$: {filtered complete Artin local$/\Lambda_{\mathrm{nov},0}$} $\to$ {Sets}, $\mathcal{MC}^+(C)$: {filtered complete$/\Lambda_{\mathrm{nov},0}$} $\to$ {Sets} as follows.

If $\mathfrak{R}$ is a filtered complete Artin local $\Lambda_{\mathrm{nov},0}$ algebra, then $\mathcal{MC}(C)(\mathfrak{R})$ is the set of all gauge equivalence classes of $b^+ \in C_{\mathfrak{R}}^+$ such that $b \equiv 0 \mod \mathfrak{R}_+$ and that b satisfies the Maurer–Cartan equation. (Here $\mathfrak{R}_+$ is a maximal ideal of $\mathfrak{R}$.) If $\mathcal{R}$ is a filtered complete $\Lambda_{\mathrm{nov},0}$ algebra, then $\mathcal{MC}^+(C)(\mathcal{R})$ is a set of all gauge equivalence classes of b such that $b \equiv 0 \mod \Lambda_{\mathrm{nov},+}\mathcal{R}$ and b satisfies the Maurer–Cartan equation.

Lemma 8.4.2. *If $b \in \mathcal{MC}(C)(\mathfrak{R})$, then there exists a $\Lambda_{\mathrm{nov},0}$ algebra homomorphism $\varphi : \mathfrak{K}_C \to \mathfrak{R}$ such that $\varphi(\tilde b) = b$. If $b \in \mathcal{MC}^+(C)(\mathcal{R})$, then there exists a $\Lambda_{\mathrm{nov},0}$ algebra homomorphism $\psi : \mathfrak{K}_C^+ \to \mathcal{R}$ such that $\psi(\tilde b^+) = b$.*

The proof of lemma 8.4.2 is similar to the proof of lemma 8.3.14. We next discuss universality. We need a condition similar to the one in theorem 8.3.6. We consider the homomorphisms:

$$\pi : H^0(C_{\mathfrak{K}_C}, \mathfrak{m}_1^{\tilde{b}}) \to H^0(\overline{C}, \overline{\mathfrak{m}}_1) \tag{8.76a}$$

$$\pi : H^0(C_{\mathfrak{K}_C^+}, \mathfrak{m}_1^{\tilde{b}^+}) \to H^0(\overline{C}, \overline{\mathfrak{m}}_1). \tag{8.76b}$$

Lemma 8.4.3. *If* (8.76) *is surjective then the homomorphisms φ and ψ in lemma* 8.4.2 *are unique.*

The proof is the same as the proof of theorem 8.3.6.

These two lemmata immediately imply the following theorem (8.4.10). We need some more notation to state it. We define another functor $\mathfrak{F}_{\mathfrak{K}_C}$: {filtered complete Artin local/$\Lambda_{\mathrm{nov},0}$} $\to$ {Sets} so that $\mathfrak{F}_{\mathfrak{K}_C}(\mathfrak{R})$ is the set of all $\Lambda_{\mathrm{nov},0}$ algebra homomorphisms $\mathfrak{K}_C \to \mathfrak{R}$. We define $\mathfrak{F}_{\mathfrak{K}_C^+}$: {filtered complete/$\Lambda_{\mathrm{nov},0}$} $\to$ {Sets} so that $\mathfrak{F}_{\mathfrak{K}_C^+}(\mathcal{R})$ is the set of all $\Lambda_{\mathrm{nov},0}$ algebra homomorphisms $\psi : \mathfrak{K}_C^+ \to \mathcal{R}$.

Theorem 8.4.10. *If* (8.76) *is surjective then the functor $\mathcal{MC}(C)$ is equivalent to $\mathfrak{F}_{\mathfrak{K}_C}$ and $\mathcal{MC}^+(C)$ is equivalent to $\mathfrak{F}_{\mathfrak{K}_C^+}$.*

Let us consider the filtered A_∞ algebra $C(L)$ of a Lagrangian submanifold defined in the last section. We assume that L is connected and $C(L)$ is unobstructed. The A_∞ algebra $\overline{C}(L) = C(L)/(\Lambda_{\mathrm{nov},+} \cdot C(L))$ is one by a rational homotopy. In particular, $H^0(\overline{C}(L); \overline{\mathfrak{m}}_1) = R$ since L is connected. Its generator is the fundamental cycle $[L]$. We proved in [33] that $[L]$ gives a non-zero element in $H^0(C_{\mathfrak{K}_C}, \mathfrak{m}_1^{\tilde{b}})$, $H^0(C_{\mathfrak{K}_C^+}, \mathfrak{m}_1^{\tilde{b}})$,[24] Therefore (8.76) is surjective in this case. Hence $\mathfrak{F}_{\mathfrak{K}_C}, \mathfrak{F}_{\mathfrak{K}_C^+}$ are universal moduli spaces of the appropriate Maurer–Cartan functors. This fact may be related to the stability of the mirror object in the complex side (compare [103]).

We recall that $\mathfrak{K}_C^+$ represents the moduli functor of deformations of C such that its restriction to $T = 0$ is trivial. This moduli functor is not an infinitesimal one, since it is a functor from {filtered complete/$\Lambda_{\mathrm{nov},0}$} whose object is not necessarily Artin or local. So it makes sense to talk about its points. (However, $\mathfrak{K}_C$ represents a moduli functor of *infinitesimal* deformation of C.)

Theorem 8.4.10 implies $\mathcal{MC}(C) = \mathrm{Hom}_{\Lambda_{\mathrm{nov},0}}(\mathfrak{K}_C^+, \Lambda_{\mathrm{nov},0})$ if (8.76) is surjective. (Here the right-hand side is the set of all $\Lambda_{\mathrm{nov},0}$ algebra homomorphisms which is continuous with respect to the $\Lambda_{\mathrm{nov},+}$ adic topology.) In other words, $\mathcal{MC}(C)$ is the set of R valued points[25] of $\mathfrak{K}_C^+$.

We next explain a spectral sequence which describes the relation of $(C, \mathfrak{m}_1^b)$ to $(\overline{C}, \overline{\mathfrak{m}}_1)$. In a filtered A_∞ algebra $C(L)$ of a Lagrangian submanifold, it gives the relation between the cohomology of L and the Floer cohomology

[24] We proved in [33] that it is a (homotopy) unit of our A_∞ algebra $C(L)$. In particular, it gives a non-trivial element for the cohomology.

[25] See, for example, [48] for its definition.

$H(C, \mathfrak{m}_1^b)$. When $\pi_2(M, L) = 0$, theorem 8.4.3 asserts that the Floer cohomology is isomorphic to the usual cohomology. The reason for this was that there is no holomorphic disc when $\pi_2(M, L) = 0$. If this assumption is not satisfied, then the Floer cohomology may not be equal to the usual cohomology of the Lagrangian submanifold. The spectral sequence we discuss here describes the procedure by which they are deformed.

Remark 8.4.5. A translation of this phenomenon into the language of physics might be: 'the instanton effect changes the dimension of the moduli space of the vacuum since it changes the mass of some particle from zero to a positive number'.

Let b be a bounding chain of C. In general, if there exists a filtration on a chain complex, we obtain a spectral sequence (see any textbook on homological algebra). Filtered A_∞ algebra C has a filtration (energy filtration) and hence $(C, \mathfrak{m}_1^b)$ is a filtered complex. However, the filtration is parametrized by a real number and not by an integer. So we fix a sufficiently small $\lambda_0 > 0$ and use the filtration $\mathcal{F}^k C = \mathfrak{F}^{k\lambda_0} C$.

The other problem is that the ring $\Lambda_{\mathrm{nov},0}$ is not noetherian. It causes serious trouble when proving the convergence of the spectral sequence. This problem is overcome in [33, section 8.A4]. We then obtain the following theorem.

Theorem 8.4.11. *We assume that C is weakly finite and strongly gapped. Let b be a bounding chain of it. Then, there exists a spectral sequence $E_r^{p,q}$ with the following properties.*

(1) $E_2^{p,q} \cong H(\overline{C}, \overline{\mathfrak{m}}_1) \otimes_R \mathcal{F}^q \Lambda_{\mathrm{nov},0} / \mathcal{F}^{q+1} \Lambda_{\mathrm{nov},0}$.
(2) *There exists a filtration $F^q H^p(C, \mathfrak{m}_1^b)$ on $H^p(C, \mathfrak{m}_1^b)$ and r_0 such that* $E_{r_0}^{p,q} \cong E_{r_0+1}^{p,q} \cong \cdots \cong E_\infty^{p,q} \cong F^q H^p(C, \mathfrak{m}_1^b)/F^{q+1} H^p(C, \mathfrak{m}_1^b)$.

Theorem 8.4.11 was proved by Oh [82] for monotonic Lagrangian submanifolds with minimal Maslov number ≥ 3 (see [33, section 8.A4] for the proof of theorem 8.4.11). We note that proposition 8.4.1 follows from theorem 8.4.11.

Before going to the next section, we explain very briefly the case when there are more than one Lagrangian submanifold. See [33] for the case when there are two Lagrangian submanifolds and [30] for three or more Lagrangian submanifolds.

Let L_1, L_2 be two Lagrangian submanifolds. We *assume* that their Maslov indexes are zero. We also assume that they are relatively spin, i.e. we assume that there exists $st \in H^2(M; \mathbb{Z}_2)$ which reduces to the second Stiefel–Whitney class of L_i. We assume that we can take the same st for both of the Lagrangian submanifolds. We then obtain filtered A_∞ algebras $(C(L_i), \mathfrak{m}_i)$. The Lagrangian intersection Floer homology is then a filtered A_∞ bimodule $CF(L_1, L_2)$, i.e. it is a left $(C(L_1), \mathfrak{m}_i)$ and right $(C(L_2), \mathfrak{m}_i)$ module. We do not define A_∞ bimodule here (see [33] for its definition).

When there are three or more Lagrangian submanifolds L_i, then we can define a product operation.

We can use the ring $\mathfrak{K}^+_{C(L_i)}$ defined in definition 8.4.15 to rewrite Lagrangian intersection Floer cohomology and their product structures as follows.

Theorem 8.4.12. *Let L_i be a countable set of mutually transversal Lagrangian submanifolds. We assume that their Maslov indexes are all zero. We also assume that there exists $st \in H^2(M;\mathbb{Z}_2)$ which restricts to $w^2(L_i)$ for any i. We fix relative spin structure for each L_i.*

Then for each i, j finitely generated, there exist $\mathfrak{K}^+_{C(L_i)}$ $\mathfrak{K}^+_{C(L_j)}$ differential graded bimodule $(\mathcal{D}(L_i, L_j), \mathfrak{m}_1)$ and operations

$$\mathfrak{m}_k : \mathcal{D}(L_{i_1}, L_{i_2})\hat{\otimes}_{\mathfrak{K}_{C(L_2)}}\mathcal{D}(L_{i_2}, L_{i_3})\hat{\otimes}_{\mathfrak{K}^+_{C(L_3)}} \cdots$$
$$\cdots\hat{\otimes}_{\mathfrak{K}^+_{C(L_{k-1})}}\mathcal{D}(L_{i_{k-1}}, L_{i_k}) \to \mathcal{D}(L_{i_1}, L_{i_k})$$

which satisfy A_∞ formula.

See [31, 33]) for the proof of theorem 8.4.12. As $\mathfrak{K}_C(L_i)\hat{\otimes}_{\Lambda_{\mathrm{nov},0}}\mathfrak{K}_C(L_j)$ module, $\mathcal{D}(L_i, L_j)$ is:

$$\mathcal{D}(L_i, L_j) \cong \bigoplus_{p\in L_1\cap L_2} \mathfrak{K}_C(L_i)\hat{\otimes}_{\Lambda_{\mathrm{nov},0}}\Lambda_{\mathrm{nov},0}[p]\hat{\otimes}_{\Lambda_{\mathrm{nov},0}}\mathfrak{K}_C(L_j).$$

We note that $\mathrm{Spec}(\mathfrak{K}^+_{C(L_i)})$ is a moduli space parametrizing a deformation of A_∞ algebra $C(L_i)$. The bimodule over $\mathfrak{K}^+_{C(L_i)}$, $\mathfrak{K}^+_{C(L_j)}$ is regarded as a coherent sheaf over the product $\mathrm{Spec}(\mathfrak{K}^+_{C(L_i)}) \times_{\mathrm{Spec}\,\Lambda_{\mathrm{nov},0}} \mathrm{Spec}(\mathfrak{K}^+_{C(L_j)})$. Hence its cohomology sheaf (which is an object of the derived category of coherent sheaves on $\mathrm{Spec}(\mathfrak{K}^+_{C(L_i)}) \times_{\mathrm{Spec}\,\Lambda_{\mathrm{nov},0}} \mathrm{Spec}(\mathfrak{K}^+_{C(L_j)})$) is a family of Floer homologies. (This is only a local family. To study global family we need more. See [31].) We will discuss the mirror object of one constructed in theorem 8.4.11.

Remark 8.4.6. The bimodule $\mathcal{D}(L_i, L_j)$ and operations in theorem 8.4.11 are invariant with respect to various choices involved, for example the choice of compatible almost complex structure J and of various perturbations. However, it is not independent of Hamiltonian diffeomorphisms, i.e. $\mathcal{D}(L_1, L_2) \neq \mathcal{D}(\varphi_1(L_1), \varphi_2(L_2))$ in general. However, the Floer homology coincides with it if we change the coefficient ring to Λ_{nov}. See [33] for the proof.

Remark 8.4.7. We have assumed that the Maslov index $\pi_2(M, L) \to \mathbb{Z}$ vanishes in this section. The reason we need it is that otherwise the operator $\mathfrak{m}_k$ does not preserve degree, because the dimension of the moduli space $\mathcal{M}_{k+1}(L;\beta)$ depends on the cohomology class β. This assumption is not used anywhere else. So by considering a $\mathbb{Z}_2$ graded Floer homology we may remove this assumption without difficulty. To apply this to symplectic geometry, it is necessary to study the general case. However, homomorphism (8.76) may not then be surjective.

Including the case when the Maslov index is non-zero leads us to the notion of *extended moduli space* (see [86]). In complex geometry, we may consider the $\mathbb{Z}_2$ graded chain complex of coherent sheaves. Including such objects also leads us to extended moduli space.

Remark 8.4.8. Usually in mirror symmetry we include a flat $U(1)$ bundle on L. In fact such a parameter is already included in our story, i.e. our parameter T may be regarded as a complex number. Its real part is related to $H^1(L;\mathbb{R})$ by the spectral sequence in theorem 8.4.11. $H^1(L;\mathbb{R})$ parametrizes a deformation of our Lagrangian submanifold L. Then the imaginary part $H^1(L;\sqrt{-1}\mathbb{R})$ corresponds to the deformation of the trivial bundle on L to a flat (non-trivial) $U(1)$ bundle.

8.4.5 Homological mirror symmetry

We now return to the complex geometry side of the story and explain more what is expected to be a mirror of the construction in sections 8.3.2, 8.3.3 and 8.3.4. I am not an expert in the complex geometry part of this story. There is much deep mathematics involved, some of which I do not know enough about. I am afraid that there might be some error in this section; however, I dare to write this section because it seems almost impossible to find anyone with sufficient knowledge of all the many aspects of mirror symmetry. For example it is rare to find anyone with enough background in both the symplectic and complex parts of the story.

I have been much influenced by Kontsevich and Soibelman [70] in writing this section.

We first introduce some more Novikov rings. Let $\lozenge$ be a sub semigroup of $\mathbb{R}$ (i.e. $\lozenge \subseteq \mathbb{R}$ such that $a, b \in \lozenge$ implies $a+b \in \lozenge$). We put $\Lambda_{R,\mathrm{nov}}^{\lozenge} = \{\sum a_i T^{\lambda_i} | \lambda_i \in \lozenge\}$. For example $\Lambda_{R,\mathrm{nov}}^{\mathbb{R}_{\geq 0}} = \Lambda_{R,\mathrm{nov},0}$, $\Lambda_{R,\mathrm{nov}}^{\mathbb{R}_{>0}} = \Lambda_{R,\mathrm{nov},+}$. However, $\Lambda_{R,\mathrm{nov}}^{\mathbb{Z}_{\geq 0}}$ is the formal power series ring $R[[T]]$ and $\Lambda_{R,\mathrm{nov}}^{\mathbb{Z}}$ is the Laurent polynomial ring $R[[T]][T^{-1}]$.

The example which appeared in the symplectic geometry of the Lagrangian submanifold is the case when $\lozenge$ is the semigroup generated by the symplectic areas (i.e. the symplectic integration form) of pseudoholomorphic discs. In the case studied by Novikov himself, i.e. Morse theory of the closed 1-form θ, the semigroup $\lozenge$ is the set of all $\theta \cap \ell$ for $\ell \in \pi_1(M)$.

In mirror symmetry, we consider the case $\Lambda_{R,\mathrm{nov}}^{\mathbb{Q}_{\geq 0}}$, and its maximal ideal $\Lambda_{R,\mathrm{nov},+}^{\mathbb{Q}_{\geq 0}} = \Lambda_{R,\mathrm{nov}}^{\mathbb{Q}_{>0}}$ and $\Lambda_{R,\mathrm{nov}}^{\mathbb{Q}}$. One may also take

$$\Lambda_{R,\mathrm{nov}}^{\mathbb{Q},0} = \bigcup_m \Lambda_{R,\mathrm{nov}}^{\mathbb{Z}[1/m]} \qquad \Lambda_{R,\mathrm{nov}}^{\mathbb{Q}_{\geq 0},0} = \Lambda_{R,\mathrm{nov}}^{\mathbb{Q},0} \cap \Lambda_{R,\mathrm{nov}}^{\mathbb{Q}_{\geq 0}}$$

$$\Lambda_{R,\mathrm{nov},+}^{\mathbb{Q}_{\geq 0},0} = \Lambda_{R,\mathrm{nov}}^{\mathbb{Q},0} \cap \Lambda_{R,\mathrm{nov}}^{\mathbb{Q}_{>0}}.$$

Exercise 8.4.1. Prove that $\Lambda_{R,\mathrm{nov}}^{\mathbb{Q},0}$ is the algebraic closure of $R[[T]][T^{-1}]$ if R is an algebraically closed field.

Elements of $\Lambda_{\mathbb{C},\mathrm{nov}}^{\mathbb{Q}_{\geq 0},0}$ is the 'formal Puiseux series' $\sum_{k=0}^{\infty} a_k T^{k/n}$. The ring $\Lambda_{R,\mathrm{nov}}^{\mathbb{Q}}$ is a completion of $\Lambda_{R,\mathrm{nov}}^{\mathbb{Q},0}$.

The geometric meaning of these rings is as follows. The ring $\mathbb{C}[[T]]$ is a ring of functions of one variable. Since we are considering *formal* power series we may say it is a 'ring of holomorphic functions on $D^2(0)$, a disc of radius zero'. The elements of $\mathbb{C}[[T]][T^{-1}]$ is a meromorphic function that is a function which is defined outside the origin but only has a pole at 0. Hence $\mathbb{C}[[T]][T^{-1}]$ may be regarded as a 'ring of holomorphic functions on $D^2(0)\backslash\{0\}$'.

Considering $\Lambda_{R,\mathrm{nov}}^{\mathbb{Z}[1/n]}$ corresponds to taking an n-fold Galois cover $D_n^2(0) \to D^2(0)$, the elements of its sum $\Lambda_{R,\mathrm{nov}}^{\mathbb{Q},0}$ may be regarded as an inductive limit $\varinjlim \mathcal{O}(D_n^2(0)\backslash\{0\})$. In algebraic geometry, it is impossible to consider a universal cover in the usual sense, so one considers a system of finite covers and takes its limit. In this sense $\Lambda_{R,\mathrm{nov}}^{\mathbb{Q},0}$ is a ring of functions of the 'algebraic universal cover' of $D^2(0)\backslash\{0\}$.

The universal Novikov ring $\Lambda_{\mathrm{nov},0}$ we used in sections 8.3.3, 8.3.4 is more transcendental in nature and may be regarded as a ring of holomorphic functions of the 'usual universal cover' of $D^2(0)\backslash\{0\}$.

Now let us explain how they appear in the complex geometry side of mirror symmetry. Let us start with a symplectic manifold (M, ω). Mirror symmetry predicts that there exists a complex manifold $(M^\vee, J)$ which is a mirror to (M, ω) under some assumption. (It is not conjectured that any symplectic manifold has a mirror, however.)

To be more precise, we have to modify it slightly. First in mirror symmetry, one usually includes a closed 2-form B on M, which is called a B field. The sum $\Omega = \omega + \sqrt{-1}B$ is called a complexified symplectic structure and (M, Ω) is expected to correspond to a complex manifold $(M^\vee, J) = (M, \Omega)^\vee$. We need to include B since the moduli space of complex structure is a complex analytic object so its mirror (a moduli space of a symplectic manifold) should be modified so that it becomes a complex object.

Furthermore, when we have a symplectic manifold (M, ω) we actually have a family of them, i.e. a family $(M, z\omega)$ where z is a complex number such that $\mathrm{Re}\, z > 0$ (then $(\mathrm{Re}\, z)\omega$ becomes the symplectic form and $(\mathrm{Im}\, z)\omega$ becomes a B field). Here I would like to state two points which seem to be widely accepted among researchers in mirror symmetry.

($\star$) The mirror $(M, \omega)^\vee$ exists only if $\omega \in H^2(M; \mathbb{Q})$, i.e. ω is a rational homology class.

($\star\star$) If $\omega \in H^2(M; \mathbb{Z})$ and $k \in \mathbb{Z}$ then $(M, z\omega)^\vee = (M, (z + k\sqrt{-1})\omega)^\vee$.

I will not try to explain the reason why these are believed. Let us assume them. We may then assume $\omega \in H^2(M; \mathbb{Z})$ since we can replace ω by $n\omega$ ($n \in \mathbb{Z}$) if necessary. We put $q = e^{-2\pi z}$.

Remark 8.4.9. In definition 8.4.8, T appeared as $T^{\omega\cap\beta}$. Hence if we put $T = e^{-1}$

then $T^{z\omega\cap\beta}$ will be $q^{\omega\cap\beta}$, the same form as in definition 8.4.8. Thus by *redefining* $T = q$, the formal parameter T in sections 8.3.3 and 8.3.4 may be regarded as a coordinate of the disc parametrizing the mirror family (see [31] for more details about this point).

($\star\star$) implies that we have a family of complex manifolds $(M^\vee, J_q) = (M, z\omega)^\vee$ parametrized by $q \in D^2(1)\backslash\{0\}$. Another conjecture (see [44, 70]) predicts that such a family is a maximal degenerate family of Calabi–Yau manifolds. Here we recall the definition of a maximal degenerate family briefly (see [70, 73] for more detail).

Let $\pi : \hat{M} \to D^2\backslash\{0\}$ be a family of complex manifolds parametrized by a unit disc D^2 minus the origin. We are interested in the case when the fibres are smooth. We assume that it extends to a flat family $\pi^+ : \overline{\hat{M}} \to D^2$ over D^2 but assume the fibre of the origin $(\pi^+)^{-1}(0)$ is singular.

The theory here is related to the theory of variation in Hodge structures (see [39]). We have a fibre bundle $\mathcal{H}^k(M) \to D^2\backslash\{0\}$ whose fibre at $q \in D^2\backslash\{0\}$ is the cohomology group of $(\pi^+)^{-1}(q) = (M, J_q)$. The bundle $\mathcal{H}^k(M)$ is a flat bundle and the flat connection is the well-known Gauss–Manin connection. Let us denote its monodromy on $H^n(M)$ by $\rho : H^n(M) \to H^n(M)$. (Here $n = \dim_{\mathbb{C}} M$.) (Since $\pi_1(D^2\backslash\{0\}) = \mathbb{Z}$ we only need to consider the generator.) It is known that the eigenvalues of ρ are all roots of unity, i.e. $\rho^N - 1$ is nilpotent for some N. It is also known that $(\rho^N - 1)^{n+1} = 0$.

Definition 8.4.16. The family $\pi : \hat{M} \to D^2\backslash\{0\}$ is said to be a maximal degenerate family if $(\rho^N - 1)^n \neq 0$, $(\rho^N - 1)^{n+1} = 0$

In [44, 70], an interesting conjecture was proposed about the behaviour of the Calabi–Yau metric on (M, J_q) when $q \to 0$ for a maximal degenerate family $\hat{M} \to D^2\backslash\{0\}$, by using Gromov–Hausdorff convergence (see [43]) of Riemannian manifolds.

We note that the maximal degenerate family is the extreme opposite to the case when monodromy is given by a Dehn twist along symplectic sphere, which is studied in detail in [57, 93–95].

Now let us return to our situation of $\pi : \hat{M} \to D^2\backslash\{0\}$. We obtain a differential graded algebra $\Omega^{0,*}(\hat{M}/D^2\backslash\{0\})$ over $\mathcal{O}(D^2\backslash\{0\})$, in the same way as explained at the beginning of section 8.1.7. The ring $\mathcal{O}(D^2\backslash\{0\})$ is different from the ring of meromorphic functions on D^2, i.e. $\mathcal{O}(D^2\backslash\{0\})$ contains a function which has an essential singularity at 0. However, using an extension of our family to 0, one can find a differential graded algebra $\hat{C}$ over $\mathcal{O}(D^2)$ so that $\hat{C} \otimes_{\mathcal{O}(D^2)} \mathcal{O}(D^2\backslash\{0\})$ is homotopy equivalent to $\Omega^{0,*}(\hat{M}/D^2\backslash\{0\})$. Moreover $\hat{C} \otimes_{\mathcal{O}(D^2)} \mathbb{C}[[T]][T^{-1}]$ is independent of the extension but depends only on the family over $D^2\backslash\{0\}$.

However, the differential graded algebra $\hat{C}$ over $\mathbb{C}[[T]]$ depends on the choice of the extension of our family at 0. The choice of such an extension is sometimes called the choice of model. It seems that no canonical choice of such model is known in the general situation and this seems to be related to a deep problem in algebraic geometry. A natural choice of a model is known in some cases, for example in the case of elliptic curves [59, 79].

The discussion on the Novikov ring at the beginning of this section suggested that it is important to consider the family parametrized not only by $\mathbb{C}[[T]]$ but also by $\Lambda_{\mathbb{C},\mathrm{nov}}^{\mathbb{Q}_{\geq 0},0}$ or $\Lambda_{\mathbb{C},\mathrm{nov}}^{\mathbb{Q}_{\geq 0}}$. To obtain such a family, we first take an n-fold cover of the base $D^2\backslash\{0\}$, pullback the family and consider the 'limit' when $n \to \infty$. It seems that what Kontsevich and Soibelman [70] suggested about the relation of mirror symmetry and rigid analytic geometry is somehow related to this point. Let me mention just one example to show this relation.

Example 8.4.2. Let us consider a (real) 2-torus T^2. Its symplectic form ω is unique up to a constant. Its mirror $(T^2, z\omega)^\vee$ is $\mathbb{C}/(\mathbb{Z} \oplus \sqrt{-1}z\mathbb{Z})$. This gives a standard family of elliptic curves parametrized by $q = e^{-2\pi z} \in D^2\backslash\{0\}$. The monodromy matrix is $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. In this case there is a canonical choice of model, i.e. we put a type I singular fibre (in the classification of Kodaira [59]) over the origin. Now we replace q by $q^{1/n}$. Then the monodromy matrix will become $\left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right)$. Hence the the singular fibre will become type I_n.

Now, what happens when $n \to \infty$? We will have a dense set of singular points which consists of all rational points of S^1 and its completion (that is S^1) may be regarded as a limit. This seems to be the picture of rigid analytic geometry [9, 12].

This S^1, in turn, will be the Gromov–Hausdorff limit of the Riemannian manifold $\mathbb{C}/(\mathbb{Z} \oplus \sqrt{-1}z\mathbb{Z})$ equipped with a Calabi–Yau metric (which is nothing other than the flat metric in this case) with its diameter normalized to 1.

This seems to be the simplest case of the picture proposed by Kontsevich and Soibelman [70].

Now let us try to formulate the homological mirror symmetry conjecture. Suppose we have a Lagrangian submanifold L of a Calabi–Yau manifold (M, ω) which is relatively spin and Maslov index $\pi_2(M, L) \to \mathbb{Z}$ is zero.

Definition 8.4.17. L is said to be *rational* if $\int_{D^1} \varphi^*\omega \in \mathbb{Q}$ for any $[\varphi] \in \pi_2(M, L)$.

Remark 8.4.10. The definition of rationality here is a tentative one. For example let us consider the case of symplectic 2-torus (T^2, ω) (2 is the real dimension). In $T^2 = \mathbb{R}^2/(\mathbb{Z} \oplus \mathbb{Z})$, we consider Lagrangian submanifolds $\mathbb{R}/\mathbb{Z} \times \{a\}$, where $a \in \mathbb{R}/\mathbb{Z}$. I would rather like to call it rational only when $a \in \mathbb{Q}/\mathbb{Z}$ but, in the sense of definition 8.4.17, it is always rational. Such a problem might disappear in the case when M is simply connected.

With a rational Lagrangian submanifold L, we can associate an A_∞ algebra $(C(L), \mathfrak{m})$ over $\Lambda_{\mathrm{nov},0}^{\mathbb{Q},0}$. Actually, using the fact that $\pi_2(M, L)$ is finitely generated, we can define it over $\Lambda_{\mathrm{nov},0}^{\mathbb{Z}[1/m]}$ for some $m \in \mathbb{Z}_{>0}$.

We suppose that we have a mirror family $(M, z\omega)^\vee$ parametrized by $q = e^{-z}$.[26] By Remark 8.4.9 we may identify q with T in sections 8.3.3 and 8.3.4. We take an m-fold cover of our mirror family and get a family $\pi_m : \hat{M}_m \to D^2\backslash\{0\}$. The mirror object is expected to be a family of holomorphic vector bundles over this family. But we need a slightly more general object than a vector bundle, i.e. an object of the derived category of coherent sheaves. In our situation, this may be regarded as a complex

$$\hat{\mathcal{E}}_1 \xrightarrow{\delta_1} \hat{\mathcal{E}}_2 \xrightarrow{\delta_2} \hat{\mathcal{E}}_3 \xrightarrow{\delta_3} \cdots \xrightarrow{\delta_{N-1}} \hat{\mathcal{E}}_N \tag{8.77}$$

here $\hat{\mathcal{E}}_i$ is a family of holomorphic vector bundles over $\pi_m : \hat{M}_m \to D^2\backslash\{0\}$ and δ_i a holomorphic section of $\mathrm{Hom}(\hat{\mathcal{E}}_i, \hat{\mathcal{E}}_{i+1})$ over $\hat{M}_m$. We write such an object as $\mathbb{E}$.

One can define a differential graded algebra describing a deformation of this object. For example we can proceed as follows.[27] We put

$$\Omega^k(\mathbb{E}, \mathbb{E}) = \bigoplus_{i,j:i\le j\le i+k} \Omega^{0,k+i-j}(\hat{M}_m/\mathcal{O}(D^2\backslash\{0\}); \mathrm{Hom}(\mathcal{E}_i, \mathcal{E}_j)). \tag{8.78}$$

For $\varphi \in \Omega^k(\mathbb{E}, \mathbb{E})$ we denote its $\Omega^{0,p-\ell}(\hat{M}_m/\mathcal{O}(D^2\backslash\{0\}); \mathrm{Hom}(\mathcal{E}_i, \mathcal{E}_{i+\ell}))$ component by $\varphi_{i,\ell}$. We then put

$$(\mathrm{d}\varphi)_{i,j} = \pm\overline{\partial}_{\mathcal{E}_j} \circ \varphi_{i,j} \pm \varphi_{i,j} \circ \overline{\partial}_{\mathcal{E}_i} \pm \delta_j \circ \varphi_{i,j-1} \pm \varphi_{i+1,j} \circ \delta_i.$$

We also put

$$(\varphi \circ \phi)_{i,j} = \sum_\ell \pm\varphi_{\ell,j} \circ \phi_{i,\ell}$$

(see [27, chapter 4] for sign). We can check $(\Omega^*(\mathbb{E}, \mathbb{E}), d, \circ)$ is a differential graded algebra and hence an A_∞ algebra over $\mathcal{O}(D^2\backslash\{0\})$ (where $T^{1/m}$ is the coordinate). If we can extend $\mathbb{E}$ to a family over D^2 then we have a differential graded algebra over $\mathcal{O}(D^2)$. We formalize it as in section 8.1.7 and obtain a differential graded algebra over $\Lambda_{\mathrm{nov},0}^{\mathbb{Z}[1/m]}$ or $\Lambda_{\mathrm{nov}}^{\mathbb{Z}[1/m]}$. We note that a differential graded algebra over $\Lambda_{\mathrm{nov},0}^{\mathbb{Z}[1/m]}$ depends on the choice of the extension (model).

Now a part of homological mirror symmetry conjecture is stated as follows.

Conjecture 8.4.1. *Let L be a rational Lagrangian submanifold which is relatively spin and its Maslov index is zero. We assume that* $(C(L), \mathfrak{m})$ *is unobstructed.*

Then there exists an object $\mathbb{E}$ *as in* (8.78) *together with its extension to* $0 \in D^2_m$, *such that* $(C(L), \mathfrak{m})$ *is homotopy equivalent to* $(\Omega^*(\mathbb{E}, \mathbb{E}), d, \circ)$ *as an* A_∞ *algebra over* $\Lambda_{\mathrm{nov},0}^{\mathbb{Z}[1/m]}$.

[26] The construction of a mirror family has now been studied extensively by several people. In this article I do not try to discuss it.

[27] Compare [27, chapter 4]. There we developed a similar but more general construction in the case of a twisted complex.

We proceed to the case corresponding to theorem 8.4.11 as follows. Let us consider the case when we have two objects $\mathbb{E}$ and $\mathbb{F}$ of a derived category of coherent sheaves on $\hat{M}_m \to D^2\backslash\{0\}$. Then we can define $\Omega^*(\mathbb{E}, \mathbb{F})$ in a way similar to that for (8.78). It is a differential graded bimodule over $(\Omega^*(\mathbb{E}, \mathbb{E}), d, \circ)$ and $(\Omega^*(\mathbb{F}, \mathbb{F}), d, \circ)$. It induces a $\mathfrak{K}_{\Omega^*(\mathbb{E},\mathbb{E})}$, $\mathfrak{K}_{\Omega^*(\mathbb{F},\mathbb{F})}$ differential graded bimodule $\mathcal{D}^*(\mathbb{E}, \mathbb{F})$.

Conjecture 8.4.2. *If Lagrangian submanifold* L_i *corresponds to* $\mathbb{E}_i$ *by conjecture* 8.4.1 *then the differential graded bimodule* $\mathcal{D}^*(\mathbb{E}_i, \mathbb{E}_j)$ *is chain homotopy equivalent to the differential graded bimodule* $\mathcal{D}^*(L_i, L_j)$.

Note conjecture 8.4.1 *implies* $\mathfrak{K}_{C(L_i)} \cong \mathfrak{K}_{\Omega(\mathbb{E}_i,\mathbb{E}_i)}$ *and hence* $\mathcal{D}^*(\mathbb{E}_i, \mathbb{E}_j)$, $\mathcal{D}^*(L_i, L_j)$ *are both* $\mathfrak{K}_{C(L_i)}$ $\mathfrak{K}_{C(L_j)}$ *differential graded bimodule.*

We can continue and state the coincidence of the product structures. In the complex side we have the Massey–Yoneda product induced by the obvious composition operators $\circ : \Omega^k(\mathbb{E}_1, \mathbb{E}_2) \otimes_{\Lambda_{\mathrm{nov},0}^{\mathbb{Z}[1/m]}} \Omega^\ell(\mathbb{E}_2, \mathbb{E}_3) \to \Omega^{k+\ell}(\mathbb{E}_1, \mathbb{E}_3)$. The most natural way to do this is to use the notion of filtered A_∞ category defined in [30]. We add two remarks.

Remark 8.4.11. In conjecture 8.4.1 we took $\Lambda_{\mathrm{nov},0}^{\mathbb{Z}[1/m]}$ with large m as a coefficient ring. However, the number m depends on $\mathbb{E}$, so to have a better statement it is natural to use $\Lambda_{\mathrm{nov},0}^{\mathbb{Q},0}$ or $\Lambda_{\mathrm{nov}}^{\mathbb{Q},0}$ in place of $\Lambda_{\mathrm{nov},0}^{\mathbb{Z}[1/m]}$ by taking the limit. To go to the limit however, we have to clarify the following point. Let us denote the mth branched cover of D^2 by D^2_m (D^2_m is actually D^2 but its coordinate is $T^{1/m}$). Let $\hat{M}_m \to D^2_m\backslash\{0\}$. Let $\hat{M}_m^+ \to D^2_m$ be a model of $\hat{M}_m$ (that is an extension of it to the origin). Let us consider $\hat{M}_{mm'} \to D^2_{mm'}\backslash\{0\}$. We want to find $\hat{M}^+_{mm'} \to D^2_{mm'}$ together with $\hat{M}^+_{mm'} \to \hat{M}^+_m$. The naive choice, that is the fibre product $D^2_{mm'} \times_{D^2_m} \hat{M}^+_m$, does not seem to be a good choice. For example this is not the correct choice for example 8.4.2. In the case of Abelian variety, [79] seems to give an appropriate choice. I do not not know whether the choice of such systems $\hat{M}^+_m \to D^2_m$ together with maps $\hat{M}^+_{mm'} \to \hat{M}^+_m$ are known in the general case of, say, Calabi–Yau manifolds.

We suppose that such a choice is given. Then we may consider a derived category of coherent sheaves over $\hat{M}$ equipped with a kind of Etal topology. Then, the A_∞ category we obtain is defined over $\Lambda_{\mathrm{nov}}^{\mathbb{Q},0}$. The coherent sheaves on Berkovich spectra [9] (as discussed in [70]) might be related to such objects. I do not have enough knowledge to discuss them at the time of writing this article.

We finally explain a mirror to example 8.2.6. Let us first consider $T^2 = \mathbb{R}^2/\mathbb{Z}^2$. Let x, y be coordinates of $\mathbb{R}^2$. We take a Lagrangian submanifold L_k of T^2 defined by $y = -kx$, where $k \in \mathbb{Z}_{\geq 0}$. We use the mirror symmetry of the elliptic curve from [85] then L_0 will become the trivial bundle on mirror $\mathbb{T}$ (elliptic curve) and L_k becomes a complex line bundle $\mathcal{L}$.

Note in this case Floer's condition $\pi_2(T^2, L_k) = 0$ is satisfied. Hence, Floer homology is defined as in section 8.3.3. We find that the intersections of L_0 and

L_k consists of k points. We can calculate its Maslov–Viterbo index and find that it is 1. Hence $HF^0(L_0, L_k) = 0$, $HF^1(L_0, L_k) = \mathbb{C}^k$. We are thus in a situation which mirrors example 8.2.6.

Now we take two Lagrangian submanifolds $L_{(1)} = L_k \times L_0$ and $L_{(2)} = L_0 \times L_k$ on the direct product $T^2 \times T^2$. We take the symplectic form ω on it such that $(T^2 \times \{0\}) \cap \omega = (\{0\} \times \omega) \cap \omega$. Then the mirror family is a product $\mathbb{T}_q \times \mathbb{T}_q$. When $q = e^{-z}$ converges to 0 the two factors will degenerate as in example 8.4.2. It is easy to see that this family is a maximally degenerate family.

Now it is easy to see that $L_{(1)}$, $L_{(2)}$ are mirrors of $\mathrm{pr}_1^* \mathcal{L}$ and $\mathrm{pr}_2^* \mathcal{L}$, respectively. The mirror of $\mathcal{E} = \mathrm{pr}_1^* \mathcal{L} \oplus \mathrm{pr}_2^* \mathcal{L}$ should be the union $L_{(1)} \cup L_{(2)}$. It is immersed, however, so the construction in sections 8.3.3 and 8.3.4 does not apply directly. However, we can modify it as follows. Take $C(L_1) \oplus C(L_2)$ and add two generators $[p^{12}]$, $[p^{21}]$ to each intersection point $p \in L_1 \cap L_2$. There are k^2 intersection points $p \in L_1 \cap L_2$ which we write as p_{ij} $i, j = 1, \dots, k$. Now

$$C(L_1) \oplus C(L_2) \oplus \bigoplus_{ij} (\Lambda_{\mathrm{nov},0}[p_{ij}^{12}] \oplus \Lambda_{\mathrm{nov},0}[p_{ij}^{21}])$$

is our complex (see [1]). The boundary operator $\mathfrak{m}_1$ is non-trivial on $C(L_i)$ and hence $\mathfrak{m}_1$ cohomology is

$$H^*(L_1) \oplus H^*(L_2) \oplus \bigoplus_{ij} (\Lambda_{\mathrm{nov},0}[p_{ij}^{12}] \oplus \Lambda_{\mathrm{nov},0}[p_{ij}^{21}]). \tag{8.79}$$

We can define a structure of A_∞ algebra on it. To calculate the formal Kuranishi map we need to calculate the product structure. Since there is no holomorphic disc which bounds L_i, it follows that the operator $\mathfrak{m}_2$ is equal to the usual cup product and $\mathfrak{m}_3$ and higher are zero, on the first two components. There is also no pseudoholomorphic disc bounding the union of the two Lagrangian submanifolds L_1, L_2, other than the trivial one. The trivial disc contributes

$$\mathfrak{m}_2(p_{ij}^{12}, p_{ij}^{21}) = [p_{ij;1}] \qquad \mathfrak{m}_2(p_{ij}^{21}, p_{ij}^{12}) = -[p_{ij;2}] \tag{8.80}$$

where $p_{ij;1}$ is a singular 0 chain $p_{ij} \in L_{(1)}$ and $p_{ij;2}$ is a singular 0 chain $p_{ij} \in L_{(2)}$. All other products are zero. We note that $H^2(L_i) = H_0(L_i)$. Hence the degree 2 part of (8.79) is $\mathbb{C}^2$ and $\mathfrak{m}_2(p_{ij}^{12}, p_{ij}^{21})$ $\mathfrak{m}_2(p_{ij}^{21}, p_{ij}^{12})$ will be the first and second factor of $H^2(L_i) \oplus H^2(L_j) = \mathbb{C}^2$, respectively. We thus find that

$$\sum_{ij} x_{ij}[p_{ij;1}] + y_{ij}[p_{ij;2}] \mapsto \left(\sum_{ij} x_{ij} y_{ij}, -\sum_{ij} x_{ij} y_{ij} \right)$$

is the (non-zero part of) Kuranishi map. This map coincides with the one in example 8.2.6.

In this example, the operator is independent of $q = e^{-z}$. We can find an example that the operator actually depends on q (and is a theta function of it) in the case $M^\vee = \mathbb{T}^3$ (see chapter 4 of [27] and chapter 7 of (2000 December version of) [33]). More examples seem to be available in the Physics literature.

References

[1] Akaho M 1997 An intersection theory of Lagrangian immersion *PhD Thesis* Kyoto University

[2] Alexandrov M, Schwarz A, Kontsevich M and Zaboronsky O 1997 The geometry of the master equation and topological quantum field theory *Int. J. Mod. Phys.* A **12** 1405–29

[3] Arnol'd V 1965 Sur une propriété topologique des applications globalement canoniques de la mécanique classique *C. R. Acad. Sci., Paris* **261** 3719–22

[4] Artin M 1976 *Deformation of Singularities* Tata Institute

[5] Axerlod S and Singer I 1991 Chern–Simons perturbation theory I *Proc. XXth Int. Conf. on Differential Theoretic method in Theoretical Physics* ed S Catto and A Rocha (Singapore: World Scientific) pp 3–45

[6] Axerlod S and Singer I 1991 Chern–Simons perturbation theory II *J. Diff. Geom.* **39** 173–213

[7] Barannikov S and Kontsevich M 1998 Frobenius manifolds and formality of Lie algebras of polyvector fields *Int. Math. Res. Notices* **4** 201–15

[8] Behrend K and Fantechi B 1997 The intrinsic normal cone *Inv. Math.* **128**(1) 45–88

[9] Berkovich V 1990 *Spectral Theory and Analytic Geometry over Non-Archimedean Fields* (Providence, RI: American Mathematical Society)

[10] Bershadsky M, Ceccoti M, Ooguri S and Vafa C 1994 Kodaira–Spencer theory of gravity and exact results for quantum string amplitude *Commun. Math. Phys.* **165** 311–427

[11] Boardman J M and Vogt R M 1973 *Homotopy Invariant Algebraic Structures on Topological Spaces (Lecture Notes in Mathematics 347)* (Berlin: Springer)

[12] Bosch S, Güntzer U and Remmert R 1984 Non-Archimedean analysis *Grundlehren der Mathematizchen Wissenschaften* (Berlin: Springer)

[13] Bourbaki N 1970 Universe *Théorie des Topos et Cohomologie Etale des Schémas (Lecture Notes in Mathematics 269)* (Berlin: Springer) pp 185–217

[14] Ciocan-Fontanine I and Kapranov M 2001 Derived Quot schemes *Ann. Sci. École Norm. Sup. (4)* **34**(3) 403–40

[15] Deligne P and Mumford D 1969 The irreducibility of the space of curves of given genus *Inst. Hautes Études Sci. Publ. Math.* **36** 75–109

[16] Donaldson S 1986 Connection, cohomology, and intersection forms of 4-manifolds *J. Diff. Geom.* **24** 275–341

[17] Donaldson S 1987 Irrationality and the h-cobordism conjecture *J. Diff. Geom.* **26**(1) 141–68

[18] Donaldson S 1990 Polynomial invariants for smooth four-manifolds *Topology* **29**(3) 257–315

[19] Douady A 1966 Le problém de modules pour les sou-espaces analytiques compact d'un espace analytique donné *Ann. Inst. Fourier* **16** 1–95

[20] Douady A, Grothendieck A and Houzel C 1960/61 *Familles d'Espace Complexes et Fondements de La Geométrié Analytique (Séminaire Henri Cartan 13)* (Secrétariat mathématique)

[21] Eliashberg Y, Givental A and Hofer H 2000 Introduction to symplectic field theory *Geom. Funct. Anal.*, Special Volume, Part I, pp 560–673 GAFA 2000 (Tel Aviv 1999)

[22] Floer A 1988 Morse theory for Lagrangian intersections. *J. Diff. Geom.* **28** 513–47

[23] Fukaya K 1993 Morse homotopy, A^{∞}-categories, and Floer homologies *Proc. 1993 Garc Workshop on Geometry and Topology (Lecture Notes Series 18)* ed H J Kim (Seoul National University) pp 1–102 http://www.kusm.kyoto-u.ac.jp/∼fukaya/fukaya.html

[24] Fukaya K 1996 Morse homotopy and Chern–Simons perturbation theory *Commun. Math. Phys.* **181**(1) 37–90

[25] Fukaya K 1997 Floer homology for 3-manifolds with boundary I *Preprint* never to appear, http://www.kusm.kyoto-u.ac.jp/∼fukaya fukaya.html

[26] Fukaya K 1998 Anti-self-dual equation on 4-manifolds with degenerate metric *Geom. Funct. Anal.* **8**(3) 466–528

[27] Fukaya K 2002 Mirror symmetry of Abelian variety and multi theta functions *J. Alg. Geom.* **11** 393–512

[28] Fukaya K 2001 Floer homology and mirror symmetry I *Winter School on Mirror Symmetry, Vector Bundles and Lagrangian Submanifolds* ed C Vafu and S-T Yau (Providence, RI: American Mathematical Society/International Press)

[29] Fukaya K 1999 *Symplectic Geometry* (Tokyo: Iwanami Shoten Publisher) in Japanese

[30] Fukaya K 2001 Floer homology and mirror symmetry II *Adv. Stud. Pure Math. Minimal Surf., Geom. Anal. Symplectic Geom.* to appear

[31] Fukaya K 2001 Floer homology for families—report of a project in progress *Contemp. Math.* to appear

[32] Fukaya K and Oh Y G 1997 Zero-loop open strings in the cotangent bundle and Morse homotopy *Asian J. Math.* **1** 99–180

[33] Fukaya K, Oh Y G, Ohta H and Ono K 2000 Lagrangian intersection Floer theory—anomaly and obstruction *Preprint* http://www.kusm.kyoto-u.ac.jp/∼fukaya/fukaya.html

[34] Fukaya K and Ono K 1999 Arnold conjecture and Gromov–Witten invariants *Topology* **38** 933–1048

[35] Gelfand S and Manin Y 1996 *Methods of Homological Algebra* (Berlin: Springer) translated from the 1988 Russian original

[36] Gilberg D and Trudinger N 1970 *Elliptic Partial Differential Equations of Second Order* vol 224 (Berlin: Springer)

[37] Goldman W and Millson J 1988 The deformation theory of representation of fundamental groups in compact Kähler manifolds. *Publ. Math. I.H.E.S.* **67** 43–96

[38] Goldman W and Millson J 1990 The homotopy invariance of Kuranishi space *Illinois J. Math.* **34** 337–67

[39] Griffiths P (ed) 1984 *Topics in Transcendental Algebraic Geometry* (Princeton, NJ: Princeton University Press)

[40] Griffiths P and Harris J 1978 *Principles of Algebraic Geometry (Pure and Applied Mathematics)* (New York: Wiley–Interscience)

[41] Griffiths P and Morgan J 1981 *Rational Homotopy Theory and Differential Form (Progress in Mathematics 16)* (Boston, MA: Birkhäuser)

[42] Gromov M 1985 Pseudoholomorphic curves in symplectic manifolds *Inv. Math.* **82**(2) 307–47

[43] Gromov M 1999 *Metric Structures for Riemannian and Non-Riemannian Spaces* (Boston, MA: Birkhäuser Boston) based on the 1981 French original [MR 85e:53051], with appendices by M Katz, P Pansu and S Semmes, translated from the French by Sean Michael Bates

[44] Gross M and Wilson P 2000 Large complex structure limits of K3 surfaces *J. Diff. Geom.* **55** 475–546

[45] Grothendieck A and Houzel C 1962 *Fondements de la Géometrie Algébraique* (Secrétariat mathématique)

[46] Gugenheim V and Stasheff J 1987 On perturbations and A_∞-structures *Bull. Soc. Math. Belg. Sér.* A **38** 237–46

[47] Hartshorne R 1966 *Residues and Duality (Lecture notes in Mathematics 20)* (Berlin: Springer)

[48] Hartshorne R 1977 *Algebraic Geometry (Graduate Texts in Mathematics 52)* (New York: Springer)

[49] Hinich V and Schechtman V 1997 Deformation theory and Lie algebra homology. I *Algebra Colloq.* **4**(2) 213–40

[50] Hinich V and Schechtman V 1997 Deformation theory and Lie algebra homology. II *Algebra Colloq.* **4**(3) 291–316

[51] Hofer H and Salamon D 1995 Floer homology and Novikov rings *The Floer Memorial Volume* (Basel: Birkhäuser) pp 483–524

[52] Hofer H and Zehnder E 1994 *Symplectic Invariants and Hamiltonian Dynamics* (Basel: Birkhäuser)

[53] Itagaki Y 1999 Convergence of the normalized solution of Maurer–Cartan equation in the Baranikov–Kontsevich construction *Master Thesis* Kyoto University

[54] Kadeishvili T V 1983 The algebraic structure in the homology of an $A(\infty)$-algebra *Soobshch. Akad. Nauk Gruzin. SSR* **108**(2) 249–52

[55] Kashiwara M and Shapira P 1990 *Sheaves on Manifolds (Grund. der Math. 292)* (Berlin: Springer)

[56] Keller B 2001 Introduction to A-infinity algebras and modules *Homology, Homotopy and Applications* **3** 1–35

[57] Khovanov M and Seidel P 2000 Quivers, Floer cohomology, and braid group actions *Preprint* MathQA/0006056

[58] Kobayashi S 1987 *Differential Geometry of Complex Vector Bundles* (Princeton, NJ: Princeton University Press)

[59] Kodaira K 1963 On compact analytic surfaces II *Ann. Math.* **77** 523–626

[60] Kodaira K, Nirenberg L and Spencer D C 1958 On the existence of deformations of complex analytic structures *Ann. Math.* **68** 450–9

[61] Kodaira K and Spencer D C 1958 On deformations of complex analytic structures I, II *Ann. Math.* **67** 328–466

[62] Kodaira K and Spencer D C 1958 A theorem of completeness for complex analytic structures *Acta Math.* **100** 281–94

[63] Kodaira K and Spencer D C 1960 On deformations of complex analytic structures III. *Ann. Math.* **71** 43–76

[64] Kontsevich M 1993 A_∞ algebras in mirror symmetry *MPI Arbeitstagung Talk*

[65] Kontsevich M 1993 Formal (non)commutative symplectic geometry *The Gelfand Mathematical Seminars, 1990–1992* (Boston, MA: Birkhäuser)

[66] Kontsevich M 1994 Feynman diagram and low dimensional topology. *Proceeding of the First European Congress of Mathematics* (Boston, MA: Birkhäuser) pp 97–122

[67] Kontsevich M 1995 Enumeration of rational curves via torus actions *The Moduli Space of Curves (Texel Island, 1994)* (Boston, MA: Birkhäuser) pp 335–68

[68] Kontsevich M 1995 Homological algebra of mirror symmetry *Proc. Int. Congress of Mathematicians, Zürich* vol I (Boston, MA: Birkhäuser) pp 120–39

[69] Kontsevich M 1997 Deformation quantization of poisson manifolds *i Preprint* q-alg/9709040

[70] Kontsevich M and Soibelman Y 2001 Homological mirror symmetry and torus fibrations *Symplectic Geometry and Mirror Symmetry* (Singapore: World Scientific)

[71] Kuranishi M 1962 On the locally complete families of complex analytic structures *Ann. Math.* **75** 536–77

[72] Li J and Tian G 1998 Virtual moduli cycles and Gromov-Witten invariants of algebraic varieties *J. Am. Math. Soc.* **11**(1) 119–74

[73] Lian B, Todorov A and Yau S T 2000 Maximal unipotent monodromy for complete intersection CY manifolds *Preprint* math.AG/0008061

[74] Manin Y 1999 *Frobenius Manifolds, Quantum Cohomology, and Moduli Spaces* (Providence, RI: American Mathematical Society)

[75] McDuff D and Salamon D 1994 *J-Holomorphic Curves and Quantum Cohomology (University Lecture Series 6)* (Providence, RI: American Mathematical Society)

[76] McDuff D and Salamon D 1995 *Introduction to Symplectic Topology* (Oxford: Oxford Science)

[77] Merkulov S A 1999 Strongly homotopy algebras of a Kähler manifold *Int. Math. Res. Notices* **3** 153–64

[78] Mumford D 1965 *Geometric Invariant Theory* (Berlin: Springer)

[79] Néron A 1964 Modèles minimaux des variétés abéliennes sur les corps locaux et globaux *Inst. Hautes Études Sci. Publ. Math. No.* **21** 128

[80] Novikov S 1981 Multivalued functions and functional—an analogue of the Morse theory *Sov. Math. Dokl.* **24** 222–5

[81] Oh Y G 1993 Floer cohomology of Lagrangian intersections and pseudo-holomorphic discs I, II *Comment. Pure Appl. Math.* **46** 949–94, 995–1012

[82] Oh Y G 1996 Floer cohomology, spectral sequences, and the Maslov class of Lagrangian embeddings *Int. Math. Res. Notices* **7** 305–46

[83] Ohta H 2001 Obstraction to and deformation of Lagrangian intersection Floer cohomology *Symplectic Geometry and Mirror Symmetry* (Singapore: World Scientific)

[84] Paramodov V P 1976 Deformation of complex spaces *Russ. Math. Surv.* **31–33** 129–97

[85] Polishchuk A and Zaslow E 1998 Categorical mirror symmetry: the elliptic curve *Adv. Theor. Math. Phys.* **2** 443–70

[86] Ran Z 1997 Thickening Calabi–Yau moduli spaces *Mirror Symmetry* vol II (Providence, RI: American Mathematical Society) pp 393–400

[87] Ray D B and Singer I M 1971 R-torsion and the Laplacian on Riemannian manifolds *Adv. Math.* **7** 145–210

[88] Ruan Y 1996 Topological sigma model and Donaldson-type invariants in Gromov theory *Duke Math. J.* **83**(2) 461–500

[89] Ruan Y and Tian G 1995 A mathematical theory of quantum cohomology *J. Diff. Geom.* **42**(2) 259–367

[90] Saito K 1983 Period mapping associated to a primitive form *Publ. Res. Inst. Math. Sci.* **19**(3) 1231–64

[91] Schlessinger M and Stasheff J 1985 The Lie algebra structure of tangent cohomology and deformation theory *J. Pure Appl. Algebra* **38**(2–3) 313–22

[92] Schlessinger V P 1968 Functor from Artin rings *Trans. Am. Math. Soc.* **130** 208–22

[93] Seidel P 2001 More about vanishing cycles and mutations *Symplectic Geometry and Mirror Symmetry* (Singapore: World Scientific)

[94] Seidel P 2002 Vanishing cycles and mutations *Proc. 3rd European Congress* (Boston, MA: Birkhauser)

[95] Seidel P and Thomas R 2001 Braid group actions on derived categories of coherent sheaves *Duke Math. J.* **108**(1) 37–108

[96] Smirnov V 2001 *Simplicial and Operad Methods in Algebraic Topology* (Providence, RI: American Mathematical Society) translated from the Russian manuscript by G L Rybnikov

[97] Stasheff J 1963 Homotopy associativity of H-spaces. I, II *Trans. Am. Math. Soc.* **108** 275–92, 293–312

[98] Stasheff J 1997 Deformation theory and the Batalin–Vilkovisky master equation *Deformation Theory and Symplectic Geometry (Ascona, 1996)* (Dordrecht: Kluwer) pp 271–84

[99] Strominger A, Yau S T and Zaslow E 1996 Mirror symmetry is T-duality *Nucl. Phys.* B **479**(1–2) 243–59

[100] Sullivan D 1978 Infinitesimal calculations in topology *Publ. I.H.E.S.* **47** 269–331

[101] Taubes C 1984 Self-dual connections on 4-manifolds with indefinite intersection matrix *J. Diff. Geom.* **19** 517–60

[102] Thomas R 1999 An obstructed bundle on a Calabi–Yau 3-fold *Adv. Theor. Math. Phys.* **3**(3) 567–76

[103] Thomas R 2001 Moment map, monodromy and mirror manifolds *Symplectic Geometry and Mirror Symmetry* (Singapore: World Scientific)

[104] Tyurina G N 1969 Locally semiuniversal flat deformations of isolated singularities of complex spaces *Izv. Akad. Nauk. SSSR Ser. Math.* **33** 1026–58

[105] Warner F 1971 *Foundations of Differential Manifolds and Lie Groups* vol 94 (Berlin: Springer)

[106] Witten E 1988 Topological sigma models *Commun. Math. Phys.* **118**(3) 411–49

Chapter 9

Large N dualities and transitions in geometry

Antonella Grassi[1] *and Michele Rossi*[2]
[1] ***University of Pennsylvania Philadelphia, PA, USA and Institute for Advanced Study, Princeton, NJ USA***
[2] ***Dipartimento di Matematica, Università di Torino, 10123 Torino***

This chapter is based on lectures given by the first author in May 2001 in Como. The second author attended the lectures and volunteered to help write the notes; in the end MR completely wrote sections 9.2.2, 9.2.3 and the appendices, which were only sketched in the lectures.

The lectures, hence this chapter, were prepared for an audience of beginning graduate students, in mathematics and physics, whom we hoped to get interested in this subject. Because most of the material presented here comes from the physics literature, we aim to build a bridge for the mathematicians towards the physics papers on the subject.

On one hand, we have tried to make this chapter self-contained and have not assumed much knowledge beyond first/second year courses. On the other hand, we thought it was important to outline links between this chapter and other research topics in string theory and mathematics, even when these were not essential to the main motif. In these cases, we have just given statements, without necessarily defining all the terms involved.

In 1974 't Hooft conjectured that large N-gauge theories are dual to closed-string theory. In 1998, Gopakumar and Vafa conjectured that $SU(N)$ Chern–Simons theory on S^3 is dual to IIA string theory (with fluxes) compactified on a certain local Calabi–Yau manifold Y, where the geometry of Y is the key to the duality.

It is, in fact, possible to do topological surgery on Y (a birational contraction followed by a complex deformation in algebraic geometry) to obtain another Calabi–Yau manifold $\widehat{Y}$; it turns out that $\widehat{Y} \cong T^*S^3$. Y and $\widehat{Y}$ are said to be related by a '*geometric conifold transition*'. From previous work by Witten, Chern–Simons theory on S^3 is equivalent to IIA on $\widehat{Y}$, with $SU(N)$ D-branes wrapped on S^3.

Evidence for the conjecture comes by comparing the partition function for the Chern–Simons theory on S^3 and the partition function for IIA on Y. The corresponding mathematical quantities are certain topological invariants of S^3 and Gromow–Witten invariants on Y; knot invariants on S^3 and 'open Gromow–Witten invariants' on Y. These invariants should 'count' maps of Riemann surfaces with boundary to Y. We use quotation marks as the 'open Gromow–Witten invariants' are not yet rigorously defined; yet, in this particular case it is possible to make some working definitions. There is still an ambiguity but, as it turns out, there is also an ambiguity on the Chern–Simons side (due to the choice of the framing of the knot) and the ambiguity on both sides match.

The topic of the last lecture in Como was the strategy to prove the conjecture, proposed by Atiyah, Maldacena and Vafa, by lifting the IIA theories to M-theory compactified on seven-dimensional manifolds with G_2 holonomy.

In the first section, after fixing some notation we describe in details the geometry of the conifold transition, because the local geometry is the key to the duality. We include two sections on transitions between Calabi–Yau threefolds and their significance in algebraic geometry and the physics of string theory. The manifolds are local Calabi–Yau, so we start with a definition of Calabi–Yau. In the second section we present some background on Chern–Simons theory and, in section 3, the evidence for the conjectures. In the last section we present the strategy of Atiyah Maldacena and Vafa and include some basics on spaces with G_2 holonomy.

We gloss over the notion of D-branes wrapped on Lagrangian submanifolds, as these were discussed in A Lerda's lectures, as well as many aspects of conformal field theory, the topic of Y Stanev's lectures. There is no discussion of IIA theory itself, partly because of time constraints, partly because IIA, IIB theories and Gromow–Witten invariants have recently been in the spotlight, thanks to the celebrated 'mirror symmetry'.

Many of the results presented in these lectures appeared in preprint form, or were announced, while the lectures were prepared and given. Other related papers appeared afterwards; we do not discuss these papers, as this chapter closely follows the lectures.

The first author would like to thank the organizers of the Conference, for the opportunity to give these talks and S Katz, J Maldacena, and N Seiberg for kindly explaining their work. Thanks are also due to R Donagi, D Harbater, P Horja, K Karu, D Morrison, B Ovrut, I Zharkov and especially L Traynor, for many useful conversations. AG is much indebted to D E Diaconescu, who patiently answered all her questions on various topics concerning these lectures. MR would like also to thank S Garbiero for useful conversations.

These lectures were prepared while the first author was a member at the Institute for Advanced Study in Princeton, NJ. AG was supported in part by the National Science Foundation grants DMS-9706707, DMS-0074980, DMS-9729992; MR by Italian MIUR's grants.

9.1 Geometry and topology of transitions

In this section we describe in detail the geometry of the conifold transition between two varieties $\widehat{Y} \supset S^3$ and Y. The local geometry is, in fact, the key to the duality between $SU(N)$ Chern–Simons on S^3 and IIA on Y, for large N. We also include two sections on the transitions between Calabi–Yau threefolds and their significance in algebraic geometry and the physics of string theory.

Y and $\widehat{Y}$ are local Calabi–Yau, i.e. open neighbourhoods in Calabi–Yau manifolds. The Calabi–Yau condition preserves the supersymmetry needed by the IIA string theory:

Definition 9.1.1. *A Calabi–Yau manifold is a smooth n-dimensional complex algebraic manifold with a trivial canonical bundle, i.e.* $\Omega^n_Y \cong \mathcal{O}_Y$ *and such that*

$$H^j(\mathcal{O}_Y) = 0\,\forall j \qquad 0 < j < n.$$

It can be verified that hypersurfaces of degree d in $\mathbb{P}^{d+1}$ are $d-1$ Calabi–Yau manifolds. Elliptic curves and $K3$ surfaces are the one- and two-dimensional Calabi–Yau manifolds.

This definition of *Calabi–Yau variety* is the most common in the algebraic geometry literature: it is the natural generalization of that of a $K3$ surface. It is worthwhile keeping in mind that there are other, non-equivalent, definitions of a Calabi–Yau threefold; we will discuss a definition, which is relevant in the physics context and its equivalence to the following one in (9.4.3), section 9.4. Note also that the current definition of $K3$ is different from the one originally used by Weil (see, for example, Barth *et al* 1984). For a nice presentation of some of the different definitions and implications among them, see Joyce (2000).

In the three-dimensional case it is first possible to have *transitions* between topologically different Calabi–Yau manifolds:

Definition 9.1.2 (Cox and Katz 1999, Morrison 1999). *Let Y be a Calabi–Yau threefold and $\phi : Y \longrightarrow \overline{Y}$ be a bimeromorphic contraction onto a normal variety. If there exists a complex deformation (smoothing) of $\overline{Y}$ to a smooth Calabi–Yau threefold $\widehat{Y}$ then the process from Y to $\widehat{Y}$ is called a transition.*

This concept plays an important role both in algebraic geometry and in superstring theory as we will see later. The following transition, the *conifold transition*, is the focus of the work of Vafa and collaborators and of these lectures; in definition 9.1.2 we briefly discuss other transitions of Calabi–Yau manifolds. This example is based on a Clemens' construction (Clemens 1983) and reported in Greene *et al* (1995) (see also Cox and Katz 1999, example 6.2.4.1).

Example 9.1.3 (Conifold transition). *Let $\overline{Y}$ be the generic quintic threefold in $\mathbb{P}^4(x_0 : \ldots : x_4)$ containing the plane π defined by $x_3 = x_4 = 0$. It is the hypersurface defined by the equation*

$$x_3 g(x_0, \ldots, x_4) + x_4 h(x_0, \ldots, x_4) = 0$$

where g, h are generic homogeneous polynomial of degree 4 (sections in $H^0(\mathcal{O}_{\mathbb{P}^4}(4))$). $\overline{Y}$ is singular precisely at the 16 points defined by the equations:

$$x_3 = x_4 = g = h = 0.$$

We will see in definition 9.1.1 that the topology of the variety around each singular point is that of a real cone, hence the name, conifold. The local equation defining each singularity is that of a node (see also appendix 9.5 in the proof of theorem 9.1.5):

$$z_1 z_3 + z_2 z_4 = 0 \subset \mathbb{C}^4. \tag{9.1}$$

Now consider the threefold $Y \subset \mathbb{P}^4 \times \mathbb{P}^1$ defined by the equations:

$$\begin{cases} y_0 g(x_0, \ldots, x_4) + y_1 h(x_0, \ldots, x_4) = 0 \\ y_0 x_4 - y_1 x_3 = 0 \end{cases} \tag{9.2}$$

with $[y_0, y_1] \in \mathbb{P}^1$. It can be directly verified that Y is smooth (or use Bertini's theorem); then $\phi : Y \longrightarrow \overline{Y}$ is an isomorphism outside the 16 nodes of $\overline{Y}$, and their inverse images in Y, which are 16 copies of $\mathbb{P}^1$s. Y is the birational resolution of $\overline{Y}$ (see appendix 9.5); ϕ is also called a 'small blow up' of Y, because the inverse image of points are complex curves and not complex surfaces. In particular $K_Y \sim \phi^*(K_{\overline{Y}}) \sim \mathcal{O}_Y$, that is ϕ is a crepant resolution (see 9.5). Moreover

$$h^{1,0}(Y) = h^{2,0}(Y) = h^{1,0}(\overline{Y}) = h^{2,0}(\overline{Y}) = 0$$

then Y is a Calabi–Yau threefold with

$$h^{1,1}(Y) = h^{1,1}(\overline{Y}) + 1 = 2.$$

Note also that all the contracted $\mathbb{P}^1$'s are on the same extremal ray of the Mori cone $\overline{NE}(Y)$, (see 9.5.4), i.e. ϕ cannot be factored in other contractions. ϕ is called a primitive contraction of type I (see 9.1.2). However, $\overline{Y} \subset \mathbb{P}^4$ can be deformed to the generic quintic threefold $\widehat{Y} \subset \mathbb{P}^4$ which is again a Calabi–Yau. The process going from Y to $\widehat{Y}$ is a (primitive) extremal transition of type I. We will see in section 9.1.1 that the topology of these singularities is that of a node: this transition is often called as the conifold transition.

By Clemens' topological analysis one can see that Y and $\widehat{Y}$ do not have the same topology. See section 9.1.1 and theorem 9.1.5 for more details.

9.1.1 The local topology of a conifold transition

Here we analyse the local geometry and topology of a conifold transition Y to $\widehat{Y}$ presented in example 9.1.3.

Definition 9.1.4. *A threefold singularity defined by the equation*

$$x^2 + y^2 + z^2 + v^2 = 0$$

is called a node (nodal singularity) (see appendix 9.5).

By a change of coordinates, the equation of the node can be re-written as

$$z_1 z_3 + z_2 z_4 = 0 \tag{9.3}$$

via the affine transformation

$$\begin{aligned} x &= z_1 + \mathrm{i} z_3 \\ y &= z_3 + \mathrm{i} z_1 \\ z &= z_2 + \mathrm{i} z_4 \\ v &= z_4 + \mathrm{i} z_2. \end{aligned} \tag{9.4}$$

The singularities of example 9.1.3 are nodes:

The conifold, revisited

The original threefold $\overline{Y} \subset \mathbb{P}^4$ is given by the equation:

$$x_3 g(x_0, \ldots, x_4) + x_4 h(x_0, \ldots, x_4) = 0.$$

By a linear projective transformation we may assume the point $P = (1 : 0 : \ldots : 0)$ to be one of the 16 singular points of $\overline{Y}$ and localize our analysis in a neighbourhood $\overline{U}$ of P. By intersecting $\overline{Y}$ with the affine open subset of $\mathbb{P}^4$ defined by $x_0 \neq 0$ we get the local equation of $\overline{U} \subset \mathbb{C}^4$:

$$z_3 \widetilde{g}(z_1, \ldots, z_4) + z_4 \widetilde{h}(z_1, \ldots, z_4) = 0$$

where $z_i := x_i / x_0$ for $i = 1, \ldots, 4$, $\widetilde{g} := g/x_0^4$ and $\widetilde{h} := h/x_0^4$. Since g and h are generic we may assume $\widetilde{g}$ and $\widetilde{h}$ to be smooth maps $\mathbb{C}^4 \longrightarrow \mathbb{C}$ submersive at the origin (i.e. at $P \in \overline{U}$) and by the inverse function theorem we have locally

$$\begin{aligned} \widetilde{g}(z_1, \ldots, z_4) &= z_1 \\ \widetilde{h}(z_1, \ldots, z_4) &= z_2 \end{aligned}$$

up to a suitable analytic change of coordinates (it is the well-known *local submersion theorem*).

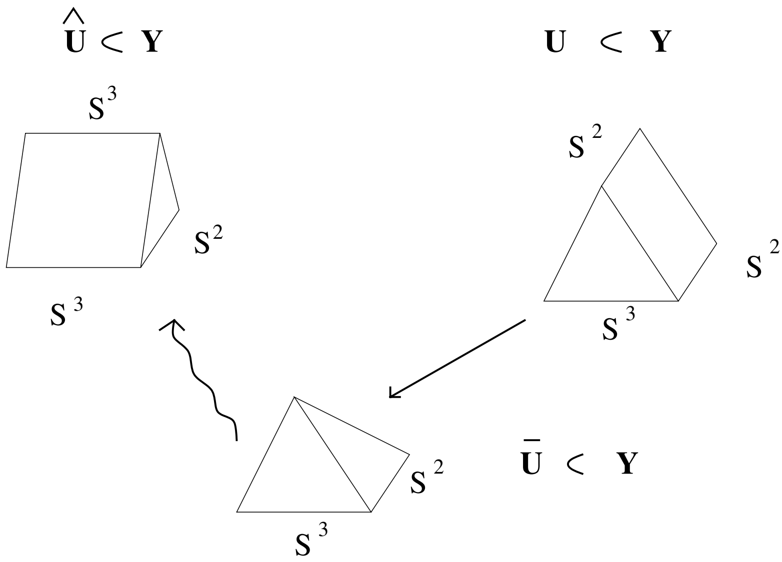

Figure 9.1. The topology of the conifold transition.

Theorem 9.1.5 (Clemens 1983, lemma 1.11).

(i) Let $\overline{U}$ be the neighbourhood of a threefold nodal singularity, then $\overline{U}$ is a real cone over $S^2 \times S^3$.

(ii) Let U be a neighbourhood of the strict transform of a node in Y, then: $U \cong D^4 \times S^2 \subset \mathbb{C}^2 \times S^2$. Furthermore $\mathcal{N}_{U|\mathbb{P}^1} \cong \mathcal{O}_{\mathbb{P}^1}(-1) \oplus \mathcal{O}_{\mathbb{P}^1}(-1)$.

(iii) Let $\widehat{U}$ be the deformed neighbourhood of a node, then $\widehat{U} \cong D^3 \times S^3 \subset T^*S^3 \cong \mathbb{R}^3 \times S^3$. In particular the non-strivial S^3 is the vanishing cycle of $\widehat{U}$ and it is locally embedded as a Lagrangian submanifold in T^*S^3.

(iv) The conifold transition is a local surgery which replaces a tubular neighbourhood $D^4 \times S^2$ of the exceptional fibre $\mathbb{P}^1_{\mathbb{C}} \cong S^2$ in U by $S^3 \times D^3$ to obtain a smoothing $\widehat{U}$ of $\overline{U}$. This is the classical surgery between two manifolds with the same boundary. In particular, U and $\widehat{U}$ are topologically different.

(v) More generally, there are relations between the Betti numbers of the Calabi–Yau manifolds Y and $\widehat{Y}$ as in example 9.1.3.

The invariants discussed in the rest of the article are determined by the local geometry around the singular locus, so we identify (sometimes perhaps too freely) the Calabi–Yau manifolds $\widehat{Y}$ and Y with the affine varieties $\mathbb{R}^3 \times S^3$ and $\mathbb{R}^4 \times S^2$ containing the local neighbourhoods $\widehat{U}$ and U.

The following proof of the theorem is a review of what's explained in the first section of Clemens (1983) and also Candelas and de la Ossa (1990).

Proof

(i)

As we have seen in (9.3), the local equation of a threefold $\overline{U}$ with a nodal singularity at the origin is:

$$z_1 z_3 + z_2 z_4 = 0. \tag{9.5}$$

Consider now the affine transformation

$$\begin{aligned} w_1 &= (z_1 + z_3)/2 \\ w_2 &= \mathrm{i}(-z_1 + z_3)/2 \\ w_3 &= (z_2 + z_4)/2 \\ w_4 &= \mathrm{i}(-z_2 + z_4)/2 \end{aligned} \tag{9.6}$$

and set $w_j = u_j + \mathrm{i}v_j$; we can now identify $\overline{U}$ with the subset $\overline{V} \subset \mathbb{R}^8$ defined by the equation:

$$\begin{aligned} &\sum_{j=1}^{4} u_j^2 - \sum_{j=1}^{4} v_j^2 = 0 \\ &\sum_{j=1}^{4} u_j v_j = 0. \end{aligned} \tag{9.7}$$

Note now that there is a diffeomorphism

$$\overline{V} \setminus \{(0, \ldots, 0)\} \cong (\mathbb{R}^4 \setminus \{(0, \ldots, 0)\}) \times S^2$$

where S^2 is the unitary sphere in $\mathbb{R}^3$. In fact, for every positive real number ρ we can consider the radius ρ hypersphere $S^7_\rho \subset \mathbb{R}^8$ and the section $V_\rho := S_\rho \cap (\overline{V} \setminus \{(0, \ldots, 0)\})$. Clearly we get

$$\overline{V} \setminus \{(0, \ldots, 0)\} = \coprod_{\rho \in \mathbb{R}_{>0}} V_\rho.$$

However, V_ρ has equations

$$\begin{aligned} &\sum_{j=1}^{4} u_j^2 = \sum_{j=1}^{4} v_j^2 = \tfrac{1}{2}\rho^2 \\ &\sum_{j=1}^{4} u_j v_j = 0. \end{aligned}$$

Hence $V_\rho \cong S^3 \times S^2$ since the fibre over a fixed point $(u_1^o, \ldots, u_4^o) \in S^3_{\rho/\sqrt{2}}$ is given by the subset of $\mathbb{R}^4\,(v_1, \ldots, v_4)$ defined by

$$\sum_{j=1}^{4} v_j^2 = \tfrac{1}{2}\rho^2$$
$$\sum_{j=1}^{4} u_j^o v_j = 0$$

which is clearly a S^2. Therefore

$$\coprod_{\rho\in\mathbb{R}_{>0}} V_\rho \cong (\mathbb{R}_{>0} \times S^3) \times S^2 \cong (\mathbb{R}^4 \setminus \{(0, \ldots, 0)\}) \times S^2 \tag{9.8}$$

and $\overline{U} \cong \overline{V}$ identifies with the real cone over $S^3 \times S^2$.

(ii) The blown-up conifold, the small resolution of a nodal singularity

Motivated by formula (9.2), we consider the standard projection $\phi : \mathbb{C}^4 \times \mathbb{P}^1 \to \mathbb{C}^4$ and its restriction to open smooth threefold $U \subset \mathbb{C}^4 \times \mathbb{P}^1$ defined by

$$\begin{aligned} y_0 z_4 - y_1 z_3 &= 0 \\ y_0 z_1 + y_1 z_2 &= 0 \end{aligned} \tag{9.9}$$

with $[y_0, y_1] \in \mathbb{P}^1$. $\phi_{|U} = \varphi : U \longrightarrow \overline{U}$. Recall that $\overline{U}$ is defined by the equation $z_1 z_3 + z_2 z_4 = 0$ and has a nodal threefold singularity at the origin. φ induces an isomorphism between the open sets $U \setminus \phi^{-1}(P) \cong \overline{U} \setminus \{(0, \ldots, 0)\} \cong \overline{V} \setminus \{(0, \ldots, 0)\}$. As in the previous, compact example, $U \to \overline{U}$ is a birational resolution of $\overline{U}$ (see appendix 9.5).

This 'small resolution' of $\overline{U}$, was obtained by 'blowing up' the plane $z_3 = z_4 = 0$; by blowing up the plane $z_3 = z_2 = 0$ we would have another small resolution U_+ isomorphic to U outside the locus of the exceptional curves. U_+ is called the *flop* of U and the birational transformation

$$U \leftarrow \cdots \rightarrow U_+ \tag{9.10}$$

the 'flop'. By analogy the transformation in section 9.4 will also be called a flop.

In particular, we then have a diffeomorphism

$$U \setminus \phi^{-1}(P) \cong (\mathbb{R}^4 \setminus \{(0, \ldots, 0)\}) \times S^2 \tag{9.11}$$

and we want to extend it to the exceptional fibre $\phi^{-1}(P) \cong \mathbb{P}^1 \cong S^2$ to give a diffeomorphism

$$U \cong \mathbb{R}^4 \times S^2. \tag{9.12}$$

In order to construct it, observe that under the affine transformation (9.6) and the previous identification $\mathbb{C}^4(w_1,\dots,w_4) \cong \mathbb{R}^8(u_1,\dots,u_4,v_1,\dots,v_4)$ the neighbourhood U is sent diffeomorphically onto the subset of $\mathbb{R}^8 \times \mathbb{P}^1_{\mathbb{C}}$ defined by

$$\begin{aligned} &y_0u_3 + y_0v_4 - y_1u_1 - y_1v_2 + i(y_0v_3 - y_0u_4 - y_1v_1 + y_1u_2) = 0 \\ &y_0u_1 - y_0v_2 + y_1u_3 - y_1v_4 + i(y_0v_1 + y_0u_2 + y_1v_3 + y_1u_4) = 0. \end{aligned} \tag{9.13}$$

Hence the fibre over a fixed point $(y_0 : y_1)^o \in \mathbb{P}^1_{\mathbb{C}}$ is a $\mathbb{R}^4 \subset \mathbb{R}^8$ ensuring the existence of the diffeomorphism (9.12) up to eventually shrink U. Moreover, by splitting y_0 and y_1 into real and imaginary parts, equations (9.13) reduce to the following matricial form:

$$\mathbf{v} = A\mathbf{u}$$

where $\mathbf{u}$ and $\mathbf{v}$ are vectors whose entries are given by u_j and v_j, respectively, and A is an antisymmetric matrix uniquely determined by the fixed projective point $(y_0 : y_1)^o$. Since outside of the origin the coordinates u_j and v_j have to satisfy the equations (9.7) this suffices to show that the restriction of the diffeomorphism (9.12) to $U \setminus \phi^{-1}(P)$ gives the diffeomorphism (9.11) precisely. Note that U can be identified with the total space of the normal bundle $\mathcal{N}_{U|\mathbb{P}^1}$ which is a holomorphic vector bundle of rank 2 over $\mathbb{P}^1$. By the Grothendieck theorem (see, for instance, Okonek *et al* 1980) we have the splitting

$$\mathcal{N}_{U|\mathbb{P}^1} \cong \mathcal{O}_{\mathbb{P}^1}(d_1) \oplus \mathcal{O}_{\mathbb{P}^1}(d_2)$$

for some $d_1, d_2 \in \mathbb{Z}$. The local equations (9.9) allows us to determine these integers. In fact, we can choose two local charts on $S^2 \cong \mathbb{P}^1(y_0 : y_1)$ around the north and south poles respectively. Say $\tau := y_0/y_1$ and $\sigma := y_1/y_0$ the two local coordinates on $\mathbb{P}^1$. Lifting these charts to $\mathcal{N}_{U|\mathbb{P}^1}$ we can choose the two local parameterizations

$$(\tau, z_1) \oplus (\tau, z_4) \qquad (\sigma, -z_2) \oplus (\sigma, z_3).$$

Look at the fibre over a fixed point $(y_0 : y_1) = (\tau : 1) = (1 : \sigma)$ in the gluing of the charts. Since here $\sigma = \tau^{-1}$ by the local equations (9.12) we get

$$\begin{aligned} -z_2 &= \sigma^{-1} z_1 = \tau z_1 \\ z_3 &= \sigma^{-1} z_4 = \tau z_4 \end{aligned}$$

which means that the transition functions $\tau^{-d_1}, \tau^{-d_2} \in \mathbb{C}^* = GL(1,\mathbb{C})$ are given by τ, i.e. $d_1 = d_2 = -1$.

(iii)

Consider the (real) one-parameter family of local smoothing $\widehat{U}_t$ of $\overline{U}$ assigned by

$$\left.\begin{aligned}&\sum_{j=1}^{4} u_j^2 - \sum_{j=1}^{4} v_j^2 = t\\&\sum_{j=1}^{4} u_j v_j = 0\end{aligned}\right.\qquad t \in \mathbb{R}_{>0}. \tag{9.14}$$

Note that the generic quintic hypersurface $\widehat{Y} \subset \mathbb{P}^4$ smoothing $\overline{Y}$ in example 9.1.3 can be chosen to admit local equations as in (9.14) for some real $t_0 > 0$ since the real one-dimensional arc parametrized by t can be chosen transversely with respect to the Zariski closed subset of singular quintic hypersurfaces and connecting $\overline{Y}$ to $\widehat{Y}$. Consider now the map

$$\mathbb{R}^8(u_1, \ldots, u_4, v_1, \ldots, v_4) \longrightarrow \mathbb{R}^8(q_1, \ldots, q_4, p_1, \ldots, p_4)$$

assigned by setting

$$\begin{aligned}q_j &= \frac{u_j}{\sqrt{t + \sum_i v_i^2}}\\p_j &= v_j.\end{aligned} \tag{9.15}$$

For every $t > 0$ it maps $\widehat{U}_t$ diffeomorphically onto the *cotangent bundle* $T^*S^3 \cong S^3 \times \mathbb{R}^3$ *to the unitary sphere* $S^3 \subset \mathbb{R}^4(q_1, \ldots, q_4)$ embedded in $\mathbb{R}^8$ as follows.

$$\begin{aligned}&\sum_{j=1}^{4} q_j^2 = 1\\&\sum_{j=1}^{4} q_j p_j = 0.\end{aligned} \tag{9.16}$$

Note that the 3-cycle $S_t \subset \widehat{U}_t$ described in $\mathbb{R}^8$ by

$$\begin{aligned}&\sum_{j=1}^{4} u_j^2 = t\\&v_1 = \cdots = v_4 = 0\end{aligned}$$

which vanishes when $t = 0$, is diffeomorphically sent onto the unitary sphere $S^3 \subset T^*S^3$.

The deformed conifold as a symplectic manifold

The canonical symplectic form given by

$$\omega := \mathrm{d}\vartheta$$

where $\vartheta := \sum_{j=1}^4 p_j\, dq_j$ is the *Liouville form of* $\mathbb{R}^8$, induces a vanishing symplectic form on S^3 since this sphere is described in T^*S^3 by $p_1 = \cdots = p_4 = 0$ (locally only three of these equations are needed). This shows that S^3 is a *Lagrangian subvariety of* T^*S^3:

Definition 9.1.6. *$Y \subset X$ a subvariety is called Lagrangian if* $\dim Y = 1/2 \dim(X)$ *and the symplectic form ω of X annihilates on every tangent vector to Y, i.e.*

$$\forall p \in Y,\ \forall u, v \in T_p Y \quad \omega(u, v) = 0.$$

The same is then true for the vanishing cycle $S_t \subset \widehat{U}_t$.

(iv) The local description of the conifold transition

Consider the diffeomorphism:

$$\alpha : (\mathbb{R}^4(\mathbf{u}) \setminus \mathbf{0}) \times \mathbb{R}^4(v) \longrightarrow (\mathbb{R}^4(\mathbf{q}) \setminus \mathbf{0}) \times \mathbb{R}^4(\mathbf{p}) \tag{9.17}$$

given by

$$q_j = \frac{u_j}{\sqrt{\sum_i u_i^2}}$$
$$p_j = v_j \sqrt{\sum_i u_i^2}.$$

Note that, by (9.7) and (9.16), α restricts to a diffeomorphism

$$U \setminus \phi^{-1}(P) \cong (\mathbb{R}^4 \setminus \{\mathbf{0}\}) \times S^2 \overset{\alpha}{\cong} S^3 \times (\mathbb{R}^3 \setminus \{\mathbf{0}\}). \tag{9.18}$$

In particular the fibre over a fixed point $\mathbf{u}^o \in \mathbb{R}^4 \setminus \{\mathbf{0}\}$ such that $\sum_i (u_i^o)^2 = \rho^2$, which is the 2-sphere $S^2_\rho \subset \mathbb{R}^4(\mathbf{v})$ given by $\sum_j v_j^2 - \rho^2 = \sum_{j=1}^4 u_j^o v_j = 0$, is diffeomorphically sent onto the fibre over the fixed point $\mathbf{q}^o = \alpha(\mathbf{u}^o)$, which is the 2-sphere $S^2_{\rho^2} \subset \mathbb{R}^4(\mathbf{p})$ given by $\sum_j p_j^2 - \rho^4 = \sum_{j=1}^4 q_j^o p_j = 0$. Calling D^n the closed unitary ball in $\mathbb{R}^n$, this means that α *restricts to give a diffeomorphism*

$$(D^4 \setminus \{\mathbf{0}\}) \times S^2 \overset{\alpha}{\cong} S^3 \times (D^3 \setminus \{\mathbf{0}\}) \tag{9.19}$$

which reduces to the identity on their boundaries $S^3 \times S^2$. Hence recalling (9.12) we can cut the interior of a $D^4 \times S^2$ around the exceptional fibre $\phi^{-1}(P)$ in U and paste by α the interior of a $S^3 \times D^3$ to get $\widehat{U}_t$ for some $t > 0$.

(v) The Betti numbers

If $\overline{Y}$ has N nodes (and no other singular points) and δ is the number of linearly independent vanishing cycles in the smoothing $\widehat{Y}$, we get the following relationship between the Betti and the Euler numbers of Y and $\widehat{Y}$:

$$\begin{aligned} &b^3(Y) = b^3(\widehat{Y}) - 2\delta \\ &b^2(Y) + b^4(Y) = b^2(\widehat{Y}) + b^4(\widehat{Y}) + 2(N - \delta) \\ &\chi(Y) = \chi(\widehat{Y}) + 2N \end{aligned} \tag{9.20}$$

(see Clemens (1983) and Werner and van Geemen (1990) for detailed proofs). Note that by the Calabi–Yau condition the first equation gives the following relationship between the Hodge numbers of Y and $\widehat{Y}$:

$$h^{2,1}(Y) = h^{1,2}(Y) = h^{2,1}(\widehat{Y}) - \delta = h^{1,2}(\widehat{Y}) - \delta. \qquad \diamond$$

The invariants discussed in the rest of the article are determined by the local geometry around the conifold locus, so we identify the local Calabi–Yau Y, $\widehat{Y}$ and $\overline{Y}$ with the local neighbourhoods U, $\widehat{U}$ and $\overline{U}$.

9.1.2 Transitions of Calabi–Yau threefolds

Let Y and $\overline{Y}$ be projective Calabi–Yau manifolds and ϕ is a birational contraction. See appendix 9.5 for definitions of the different types of singularities used in this section.

Definition 9.1.7. *$\phi : Y \to \overline{Y}$ is a primitive contraction if it cannot be further factored into birational morphisms of normal varieties.*

Non-primitive Calabi–Yau contraction may be factored into a composite of primitive contractions (see Wilson 1989) so, without loss of generality we can consider ϕ to be primitive. In this case the pullback $\phi^* H$ of an ample divisor H on $\overline{Y}$ will cut the Mori cone (see section 9.5.4) $\overline{NE}(Y)$ along an *extremal face*. Such contractions are also called *extremal* and the associated transitions *primitive extremal transitions*.

Definition 9.1.8 (Wilson 1992). *A primitive contraction is:*

- *of type I if the exceptional locus E of the associated primitive contraction ϕ is composed by finitely many curves,*
- *of type II if ϕ contracts a divisor down to a point and*
- *of type III if ϕ contracts a divisor down to a curve.*

In the first case $\phi(E)$ is composed by a finite number of isolated singularities, each with a small resolution. Since Y is smooth these singularities are necessarily terminal of index 1 and therefore cDV points. In the second case, E must be irreducible and, more precisely, it is a del Pezzo surface (see Reid 1980); $\phi(E)$ is a canonical singular point of index 1. In the third case E is again an irreducible surface contracted down to a curve $\phi(E)$ of canonical singularities for $\overline{Y}$. In particular, if ϕ is crepant then E is a conic bundle over the curve $\phi(E)$ which is a smooth curve of (generically cA_1 or cA_2) cDV points (see Reid 1980, Wilson 1992, theorem 2.2).

The simplest example of a non-trivial transition of type I is the conifold transition of example 9.1.3, i.e. a transition allowing only isolated simple double points (nodes) for $\overline{Y}$. In fact these singularities can (at least locally) be smoothed. The following results also hold:

Theorem 9.1.9 (Friedman 1986). *If ϕ is of type I and the singularity is an ordinary double point, then $\overline{Y}$ is smoothable unless ϕ is the contraction of a single $\mathbb{P}^1$ to an ordinary double point.*

Theorem 9.1.10 (Altmann 1994, Gross 1997a, b, Schlessinger 1971).

- *If ϕ is of type II and $\overline{Y}$ is $\mathbb{Q}$-factorial, then $\overline{Y}$ is smoothable unless $E \cong \mathbb{P}^2$ or $E \cong \mathbb{F}_1$.*
- *If ϕ is of type III and $\overline{Y}$ is $\mathbb{Q}$-factorial, then $\overline{Y}$ is smoothable unless $\phi(E) \cong \mathbb{P}^1$ and $E \cong E^7, E^8$, the del Pezzo surfaces of degree 7 and 8 respectively. (Equivalently, E^j is the surface obtained by blowing-up $\mathbb{P}^2$ at j points in general position.)*

After Clemens' work (see definition 9.1.1), Reid (1987) suggested that the birational classes of Calabi–Yau threefolds would *fit together into one irreducible family.* In fact, he speculated that transitions may connect a general Calabi–Yau threefold to a non-Kähler analytic threefold with trivial canonical class, Betti number $b_2 = 0$ and diffeomorphic to a connected sum of N copies of $S^3 \times S^3$, where N is arbitrarily large. This conjecture is usually known as *Reid's fantasy*. There exists various evidence for this conjecture (the Calabi–Yau web: see, e.g., Avram *et al* 1996, Chiang *et al* 1996).

9.1.3 Transitions and mirror symmetry

Assume that there exists a transition from Y_1 to $\widehat{Y}_1$, factorizing through a birational contraction $\phi : Y_1 \longrightarrow \overline{Y}_1$; assume also that the *mirror partners* (see, for example, Morrison 1999) Y_2 of Y_1 and $\widehat{Y}_2$ of $\widehat{Y}_1$ exist (see, for example, Morrison 1999). It is believed that the mirror partners $\widehat{Y}_2$ and Y_2 are also connected by a transition, which factorizes through a birational contraction $\phi^\circ : \widehat{Y}_2 \longrightarrow \overline{Y}_2$; the transition between $\widehat{Y}_2$ and Y_2 is often called the 'reverse transition'. It is not known whether this conjecture holds; see, for example, Batyrev *et al* (1998), for the case of the conifold transition.

The mirror symmetry exchanges the Hodge numbers $h^{1,2}$ (representing the dimension of the complex moduli space) with $h^{1,1}$ (the Kähler moduli space) of the Calabi–Yau mirror partners; this exchange is consistent with a partner mirror transition as we will see in the section 9.1.1. Greene and Plesser (1990) outlined

$$\begin{tikzcd}
 & Y_1 \arrow[d] & & \widehat{Y}_2 \arrow[d] \\
\widehat{Y}_1 & Y_1 \arrow[l, squiggly] & Y_2 & Y_2 \arrow[l, squiggly]
\end{tikzcd}$$

Figure 9.2. The mirror transition.

an heuristic approach to 'continuously' extend mirror symmetry to all the Calabi–Yau threefolds belonging to the same connected component of the web generated by conifold transitions. Actually *if transitions would connect each other to all Calabi–Yau threefolds*, which is a rough version of the Reid's fantasy, then it could give an approach to establishing mirror symmetry for all of them.

In the examples studied by Candelas *et al* (1994) and Morrison (1999) Y_1, $\widehat{Y_1}$ and their mirrors are related by a primitive contraction of type III (see appendix 9.6).

9.1.4 Transitions, black holes etc

The transitions among Calabi–Yau manifolds are also crucial in the context of string theory, as they connect two topologically distinct compactifications of a ten-dimensional type II string theory (to four-dimensional string vacua). Since in spite of the small number of consistent ten-dimensional string theories, their Calabi–Yau compactifications give rise to a multitude of four-dimensional topologically distinct string vacua, the transitions may prove to be the suitable mathematical tool which is able to restore a concept of *uniqueness* in compactified string theory when mirror symmetry and a version of Reid's fantasy (the Calabi–Yau web) is assumed. The physical interpretation would then be that *two four-dimensional topologically distinct string vacua may be connected to each other by means of a black hole condensation.* This is the work of Greene *et al* (1995) and Strominger (1995).

Strominger gave a physical explanation of how to resolve the conifold singularities of the moduli space of classical string vacua by means of massless Ramond–Ramond (R–R) black holes (see appendix 9.7).

In Greene *et al* (1995) the transformation of a massive black hole into a massless one at the conifold model is called *condensation.* Not only conifold transitions have a physical counterpart. For example a similar interpretation involves type II transitions in the context of the string–string duality (see Katz *et al* 1996, Berglund *et al* 1995, 1997).

Transitions of Calabi–Yau manifolds also have a role in five-dimensional supersymmetric theories (see, for example, Morrison and Seiberg (1997) and Douglas *et al* (1997)).

9.2 Chern–Simons theory

One side of the conjecture involves Chern–Simons theory on S^3: this section is an overview of this theory. We start with a quick review of the mathematical background, principal bundles and connections: appendix 9.8 contains more details. Next we discuss some basics of classical Chern–Simons theory (following Freed 1995), and of its quantum version (following Witten 1989). The first evidence for the conjecture comes from comparing an expansion of the Chern–Simons partition function, so the last section is dedicated to the computational aspects and link invariants.

Let $\pi : P \to M$ be a principal G-bundle with G acting on the right (see definition 9.8.1). In particular, for any $m \in M$, $\pi^{-1}(m) \cong G$. The differential of this map gives an isomorphism

$$\mathrm{d}\pi : T_p\pi^{-1}(m) \xrightarrow{\cong} T_{\mathrm{id}}G \cong \mathfrak{g}.$$

Definition 9.2.1. *The vertical bundle on P is the vector sub-bundle $\mathcal{V}P$ of TP given by* $\ker(\mathrm{d}\pi)$ *that is for every $p \in P$*

$$\mathcal{V}_pP := \ker[d_p\pi : T_pP \longrightarrow T_{\pi(p)}M].$$

Then the vertical bundle $\mathcal{V}P$ associated with the principal G-bundle (P, π) is a vector bundle whose standard fibre is the Lie algebra $\mathfrak{g}$ associated with G (see remark 9.8.4).

A connection is an infinitesimal version of a G-equivariant family of sections of $\pi : P \to M$.

Definition 9.2.2. *A connection on a principal G-bundle (P, π) is a vector sub-bundle $\mathcal{H}P$ of TP such that*

$$TP = \mathcal{H}P \oplus \mathcal{V}P \tag{9.21}$$

and for every $p \in P$ and $\sigma \in G$

$$d_pR(\sigma)(\mathcal{H}_pP) = \mathcal{H}_{p\sigma}P \tag{9.22}$$

where R is the right action of G on P (see definition 9.8.1).

Definition 9.2.3. *(i) The connection form of a connection $\mathcal{H}P$ is the $\mathfrak{g}$-valued 1-form $A \in \Omega^1(P, \mathfrak{g})$ such that, for every $p \in P$ and $u \in T_pP$,*

$$A_pu := (d_{\mathrm{id}}\lambda_p)^{-1}(\mathcal{V}_pu) \in T_{\mathrm{id}}G \cong \mathfrak{g} \tag{9.23}$$

where $\lambda_p : G \xrightarrow{\cong} \pi^{-1}(\pi(p)) \subset P$ is the diffeomorphism given by $\lambda_p(\sigma) := p\sigma$. It is a characteristic form of the connection $\mathcal{H}P$ since $\mathcal{H}P = \ker A$ (see proposition 9.8.6).

(ii) The curvature form of a connection $\mathcal{H}P$ is the $\mathfrak{g}$-valued 2-form $\Omega \in \Omega^2(P, \mathfrak{g})$ defined by

$$\Omega_p(u, v) := -A_p[U, V]_p \qquad \forall p \in P, u, v \in T_pP \tag{9.24}$$

where U, V are any horizontal vector fields on P extending the horizontal parts $\mathcal{H}_p u$ and $\mathcal{H}_p v$ of u and v respectively, (recall the splitting (9.103)).

Definition 9.2.4. *A gauge transformation of P is an automorphism φ of P which induces the identity map on the base manifold M.*

Gauge transformations on P form a group $\mathcal{G}_P$, and (9.114) defines an action of $\mathcal{G}_P$ on the affine space of connections $\mathcal{A}_P$ (see proposition 9.8.6).

Definition 9.2.5. *Let $\gamma : I := [0, 1] \longrightarrow M$ be a loop with base point $m \in M$ and let $\widetilde{\gamma}_p : I \longrightarrow P$ be the unique horizontal lift of γ with initial point $p \in P$, i.e. such that*

$$d\widetilde{\gamma}_p(TI) \subset \mathcal{H}P \qquad \text{and} \qquad \widetilde{\gamma}_p(0) = p.$$

Define a diffeomorphism of the fibre $\pi^{-1}(m)$ by

$$\begin{array}{rccc} h_\gamma : & \pi^{-1}(m) & \longrightarrow & \pi^{-1}(m) \\ & p & \longmapsto & \widetilde{\gamma}_p(1). \end{array} \tag{9.25}$$

Then;

$$\mathrm{Hol}_{\mathcal{H}P}(m) := \{h_\gamma : \gamma \textit{ is a loop based at } m\} \tag{9.26}$$

is a group (with the composition of morphisms), called the holonomy group of the connection $\mathcal{H}P$ at $m \in M$.

If the base manifold M is connected all these groups are isomorphic by (9.116). Then $Hol_{\mathcal{H}P}$ is called the holonomy group of the connection $\mathcal{H}P$.

Note that for every $p \in P$ it is possible to identify $\mathrm{Hol}_{\mathcal{H}P}(\pi(p))$ with the subgroup of G

$$G_{\mathcal{H}P}(p) := \{\sigma_\gamma(p) \in G : h_\gamma(p) = p\sigma_\gamma(p) \text{ and } h_\gamma \in \mathrm{Hol}_{\mathcal{H}P}(\pi(p))\}. \tag{9.27}$$

If $p, q \in \pi^{-1}(m)$ then $G_{\mathcal{H}P}(p)$ and $G_{\mathcal{H}P}(q)$ are conjugate subgroups and they coincide if p and q can be joined by an horizontal curve in P.

Definition 9.2.6. *The restricted holonomy group of the connection $\mathcal{H}P$ at $m \in M$*

$$H^{(o)}_{\mathcal{H}P}(m) \subset \mathrm{Hol}_{\mathcal{H}P}(m) \tag{9.28}$$

is defined by considering homotopically trivial loops based at m.

As before, if M is connected we can define the restricted holonomy group $H^{(o)}_{\mathcal{H}P} \subset \mathrm{Hol}_{\mathcal{H}P}$. Moreover, for every $p \in P$ we can identify the restricted holonomy subgroup $H^{(o)}_{\mathcal{H}P}(\pi(p))$ with a suitable subgroup $G^{(o)}_{\mathcal{H}P}(p) \subset G_{\mathcal{H}P}(p) \subset G$.

9.2.1 Chern–Simons' form and action

In this section we follow the notation of Freed (1995), Witten (1989) and Labastida (1999); the reader should consult these papers for more details. Let us assume the base manifold $M = \pi(P)$ to be a smooth and compact 3-manifold. Let $\mathcal{A}_P$ be the affine space of all the possible connection on P and choose $A \in \mathcal{A}_P$ with associated connection $\mathcal{H}P = \ker A$. If $\Omega \in \Omega^2(P, \mathfrak{g})$ is the $\mathfrak{g}$-valued curvature 2-form of the chosen connection then

$$\Omega \wedge \Omega \in \Omega^4(P, \mathfrak{g} \otimes \mathfrak{g}).$$

Definition 9.2.7. *The Chern–Weil 4-form associated with the Killing form $\langle\ ,\ \rangle$ (see definition 9.8.7) is $\langle \Omega \wedge \Omega \rangle \in \Omega^4(P)$.*

Definition 9.2.8. *A Chern–Simons form is an antiderivative $\alpha \in \Omega^3(P)$ of $\langle \Omega \wedge \Omega \rangle$.*

Proposition 9.2.9. *Let $\alpha := \langle A \wedge \Omega \rangle - \frac{1}{6}\langle A \wedge [A, A] \rangle$. Then,*

(i) $\mathrm{d}\alpha = \langle \Omega \wedge \Omega \rangle$,
(ii) if φ is a gauge transformation of P,

$$(\delta\varphi)\alpha = \alpha - \frac{1}{6}\langle \phi \wedge [\phi, \phi] \rangle + d\langle (Ad_{\sigma_\varphi^{-1}} \circ A) \wedge \phi \rangle \tag{9.29}$$

where δ is the codifferential, σ_φ is associated with φ like in (9.113), $\phi := (\delta\sigma_\varphi)(\delta\lambda)A$ *and $(\delta\lambda)A$ is the Maurer–Cartan form of the connection $\mathcal{H}P$ as defined in (9.108).*
(iii) If α' is a Chern–Simons form, the 3-form $(\delta\varphi)\alpha' - \alpha' + \frac{1}{6}\langle \phi \wedge [\phi, \phi] \rangle$ is exact.

The proof follows directly by the definition 9.2.8 of α and by the gauge action on connections (9.114). By (9.115) and the Ad-invariance (see (9.105)) of the Killing form *the Chern–Weil form $\langle \Omega \wedge \Omega \rangle$ is gauge invariant*. Moreover:

Proposition 9.2.10. *$\alpha' - (\delta\varphi)\alpha'$ defines a cohomology class*

$$(\delta\sigma_\varphi)\Phi_A \in H^3(P, \mathbb{R})$$

which is independent by the chosen Chern–Simons form α'. We can also assume that

$$\rho\Phi_A \in H^3(G, \mathbb{Z}) \tag{9.30}$$

for a suitable real number ρ.

In fact, the 3-form $\alpha' - (\delta\varphi)\alpha'$ is closed for every gauge transformation φ and any Chern–Simons form α'. Also it is the image by the codifferential $\delta\sigma_\varphi$ of the cohomology class $\Phi_A \in H^3(G, \mathbb{R})$ associated with the closed 3-form

$$\frac{1}{6}\langle (\delta\lambda)A \wedge [(\delta\lambda)A, (\delta\lambda)A] \rangle \in \Omega^3(G).$$

Note that the choice of $\rho \in \mathbb{R}$ depends only on the connection $\mathcal{H}P$.

Definition 9.2.11. *If there exists a global section*

$$s : M \longrightarrow P$$

the Chern–Simons Lagrangian on M is the 3-form

$$\mathcal{L}(A, s) := \rho(\delta s)\alpha \in \Omega^3(M) \tag{9.31}$$

and the associated Chern–Simons action is obtained by integrating it over M

$$S(\mathcal{L}) := \int_M \mathcal{L}(A, s). \tag{9.32}$$

Remark 9.2.12.

(i) The existence of a section means that P is parallelizable which is the case for example when G is simply connected (see Freed (1995), lemma 2.1, for a proof of this fact.)
(ii) By Stokes' theorem the Chern–Simons action S does not depend on the choice of the Chern–Simons form α when M is assumed without boundary.
(iii) For any gauge transformation φ, the 3-form $\mathcal{L}(A, s) - (\delta\varphi)\mathcal{L}(A, s)$ defines the integral cohomology class

$$\rho\delta(\sigma_\varphi \circ s)\Phi_A \in H^3(M, \mathbb{Z})$$

hence

$$S(\mathcal{L}) - S((\delta\varphi)\mathcal{L}) = \rho \int_M \delta(\sigma_\varphi \circ s)\Phi_A \in \mathbb{Z}. \tag{9.33}$$

(iv) For the particular case $G = SU(2)$ the integral bilinear forms on $\mathfrak{g} = \mathfrak{su}_2$ are parameterized by $k \in \mathbb{Z}$ as follows

$$\forall X, Y \in \mathfrak{su}_2 \qquad \langle X, Y\rangle_k = \frac{k}{8\pi^2}\operatorname{Tr}(XY).$$

Then the real coefficient in (9.30) can be given by $\rho := (8\pi^2)^{-1}$ and the Chern–Simons Lagrangian (9.31) becomes

$$\mathcal{L}(A, s) = \frac{1}{8\pi^2}\operatorname{Tr}(A' \wedge \mathrm{d}A' + \tfrac{2}{3}A' \wedge A' \wedge A')$$

where $A' := (\delta s)A$ (see section 6 in Freed, 1995). This is the typical shape of a Chern–Simons Lagrangian usually adopted in physical literature although the gauge group G is more general than $SU(2)$.

Proposition 9.2.13. *The Chern–Simons action*

$$S[A] := \exp(\mathrm{i}k2\pi S(\mathcal{L})) \tag{9.34}$$

is well defined and gauge invariant, where $k \in \mathbb{Z}$ is called the level of the theory. Furthermore, $S[A]$ depends only on the choice of the gauge equivalence class of connections $[A] \in \mathcal{A}_P/\mathcal{G}_P$ where $\mathcal{G}_P$ acts on $\mathcal{A}_P$ as in (9.114).

In fact any two sections of P are related by a gauge transformation and the assumption (9.30) holds.

By the physical point of view, it is relevant to point out the *quantization law* expressed by (9.33) and (9.34). The real factor ρ defined in (9.30) may be considered to be a *normalizing factor of the Killing form of* $\mathfrak{g}$. Then we can write (9.33) as

$$S(\mathcal{L}) - S((\delta\varphi)\mathcal{L}) = \int_M \delta(\sigma_\varphi \circ s)\Phi_A \in \mathbb{Z}.$$

We can also relate any gauge transformation φ with a map $M \to G$ by taking $\sigma_\varphi \circ s$. In this way we get an immersion of the gauge group $\mathcal{G}_P$ into the group of maps from M to G. $\int_M \delta(\sigma_\varphi \circ s)\Phi_A$ is called the *winding number of the gauge transformation* φ. Since this number is homotopically invariant it is revealing to count the homotopy classes of gauge transformations giving two relevant consequences:

(i) the Chern–Simons action (9.32) is invariant under any gauge transformation homotopically equivalent to the identity; and
(ii) as in the Dirac's well-known work on magnetic monopole, the integer k in (9.34) is found to be closely related to the central charge of the theory. Moreover, in the quantum field theory defined by the following partition function (9.35) k^{-1} is proportional, for large k, to the square λ of the *coupling constant* of the theory (see (9.83)).

Definition 9.2.14. *The Chern–Simons partition function is the Feynman integral of the Chern–Simons action (9.34) taken over all the gauge equivalence classes of connections:*

$$Z(M) := \int_{\mathcal{A}_P/\mathcal{G}_P} S[A]D[A]. \tag{9.35}$$

This defines the *Chern–Simons quantum field theory* (see, for example, Deligne *et al* 1999) whose fields are precisely the elements of $\mathcal{A}_P/\mathcal{G}_P$.

Definition 9.2.15. *Let K be a knot in M, i.e. an embedding of the circle S^1 and R, a representation of G.* The Wilson line *W_K^R is the functional*

$$W_K^R : \mathcal{A}_P/\mathcal{G}_P \longrightarrow \mathbb{R} \tag{9.36}$$

where $W_K^R[A] := \mathrm{Tr}_R(h_K)$ and h_k is the holonomy around K.

Note that the real number $\mathrm{Tr}_R(h_K)$ is well defined for any representation R of G. K can be thought of as a closed loop in M; for every point $m \in K$ we obtain an element $h_K \in \mathrm{Hol}_{\mathcal{H}P}(m)$ as in (9.25). If M is connected h_K does not depend on the choice of $m \in K$ since we can proceed as in (9.116) to obtain $h_K \in \mathrm{Hol}_{\mathcal{H}P}$. By (9.27) h_K defines a conjugacy class in G.

The Wilson line are metric independent (i.e. covariant) and gauge-invariant functionals of the fields; they are then *observables* of the theory.

Since $\mathrm{Tr}_R(h_K)$ is gauge invariant, we define:

Definition 9.2.16. *The unnormalized expectation value is then formally assigned by the Feynman integral*

$$Z(M; K, R) := \int_{\mathcal{A}_P/\mathcal{G}_P} S[A] W_K^R D[A] \tag{9.37}$$

and its expectation value is given by

$$\langle W_K^R \rangle := Z(M; K, R)/Z(M). \tag{9.38}$$

If we now consider a link L in M, i.e. the union of $r \geq 1$ oriented and non-intersecting knots $\{K_i\}_{i=1}^r$ in the oriented manifold M and a collection of irreducible representations $\mathcal{R} := \{R_i\}_{i=1}^r$ of G, one for each knot K_i, we have:

Definition 9.2.17. *The correlation function of our quantum field theory*

$$Z(M; L, \mathcal{R}) := \int_{\mathcal{A}_P/\mathcal{G}_P} S[A] \prod_{i=1}^r W_{K_i}^{R_i} D[A]. \tag{9.39}$$

9.2.2 The Hamiltonian formulation of the Chern–Simons QFT (following Witten's canonical quantization)

Although the mathematical definitions of path integrals in (9.35), (9.37) and (9.39) are quite delicate, the explicit integrals are calculated in Witten (1989). Witten first uses the *stationary-phase approximation* in the 'classical limit' $k \to \infty$ and then *canonical quantization*. Here we present the basic ideas of this second method set-up. A very useful and pleasant reference on the argument is Atiyah (1990a), to which we refer the reader for a deeper understanding. We will not discuss the stationary-phase approximation since it lies outside the aim of the present work, although its relevance is fundamental in giving the confirmation that the partition functions introduced by the Feynman approach in the previous section are the same as the ones we will evaluate in the next section by the Hamiltonian approach: see the first part of section 2 in Witten (1989) and section 7.2 in Atiyah (1990a).

The main purpose in QFT of a Feynman path integral is to provide a relativistically invariant approach since this is a fundamental property of the *Lagrangian density* which, in our case, is expressed by the Chern–Simons action (9.32) multiplied by $2\pi k$. If we want to enucleate a time-evolution in the theory we have to break the relativistic symmetry by constructing a *time-evolution operator* $\exp(\mathrm{i}tH)$ in a certain 'Hilbert' space $\mathcal{H}$ representing the space of physical states. The generator H is the *Hamiltonian operator* of the theory. In general there are formal rules which allow us to produce the space $\mathcal{H}$ and the Hamiltonian H of a QFT whose partition function is known.

In the case of Chern–Simons QFT the spacetime is represented by the 3-manifold M. We can separate out space and time by 'cutting' M along a surface Σ. Near the cut M looks like $\Sigma \times \mathbb{R}$ giving us the desired separation of space and time. Let us then limit ourselves to considering the particular case $M = \Sigma \times \mathbb{R}$ which can be treated by means of *canonical quantization* to construct the physical space $\mathcal{H} = \mathcal{H}(\Sigma)$ of the Chern–Simons theory *quantized on* Σ. More precisely, this means to 'quantize' the space of classical solutions which are the critical fields of the Chern–Simons action (9.32).

Proposition 9.2.18. *The space of classical solutions of Chern–Simons theory is the subspace of gauge equivalence classes of flat connections in $\mathcal{A}_P/\mathcal{G}_P$ which can be naturally identified with the following*

$$\mathcal{M}_M := \hom(\pi_1(M), G)/G$$

where G acts by conjugation (see Freed (1995, proposition 3.5) for more details).

The statement follows by (9.31) and the fact that α is, by definition, an antiderivative of $\langle \Omega \wedge \Omega \rangle$. In fact

$$\mathrm{d}S(\mathcal{L}(A, s)) = 0 \quad \Longleftrightarrow \quad \Omega = 0 \tag{9.40}$$

i.e. the latter is the *Euler–Lagrange equation* of the classical Chern–Simons theory whose solutions are given by *flat connections*. See Freed (1995, proposition 3.1) for details in differentiating. Note that by (9.104) this Euler–Lagrange equation involves only first-order derivatives of the fields. This is a peculiarity of Chern–Simons gauge theory together with the independence from the choice of any metric. Since the restricted holonomy subgroups (9.28) of a flat connections are always trivial it is possible to define a morphism

$$\pi_1(M) \longrightarrow \mathrm{Hol}_{\mathcal{H}P}$$

(see, e.g., Poor 1981, proposition 2.40). By recalling (9.27) we actually get a morphism from $\pi_1(M)$ to G which is well defined up to conjugation. In contrast a similar equivalence class of morphisms allows us to determine a flat connection on P.

Since we are in the particular case $M = \Sigma \times \mathbb{R}$ our space of classical solutions reduces to

$$\mathcal{M}_\Sigma := \hom(\pi_1(\Sigma), G)/G. \tag{9.41}$$

This space is not dependent on the time variable described by $\mathbb{R}$ implying that we do not actually have time-evolution in our theory, i.e. *we have no dynamics and it is all purely topological*: *hence the Hamiltonian H must be trivial.*

The following result allows to 'quantize' $\mathcal{M}_\Sigma$.

Theorem 9.2.19 (Narasimhan and Seshadri 1965, Donaldson 1983). *The space of classical solutions $\mathcal{M}_\Sigma$ is homeomorphic to the moduli space M_τ of holomorphic G-bundles over the Riemann surface Σ_τ obtained by the choice of a complex structure τ on Σ. On M_τ we have a natural choice for a holomorphic line bundle L. The finite dimensional complex vector space*

$$\mathcal{H}^k_\tau(\Sigma) := H^0(M_\tau, L^{\otimes k}) \tag{9.42}$$

of global holomorphic sections of $L^{\otimes k}$ gives the Hilbert space of the quantized theory at level k .

When $G = SU(N)$ the moduli space M_τ turns out to be a projective algebraic variety. Hence we have the natural choice $L := \mathcal{O}_{M_\tau}(1)$, i.e. the line-bundle associated with the hyperplane section. Otherwise when G is more general, the choice of the complex structure τ on Σ gives a natural complex structure on the infinite dimensional affine space $\mathcal{A}_P$. The moduli space M_τ can then be identified with the *symplectic quotient* $\mathcal{A}_P//\mathcal{G}_P$ (see Atiyah (1990a, chapter 4) for a definition) under the action (9.114) of the gauge group $\mathcal{G}_P$ (see Atiyah and Bott (1982) for the details). On $\mathcal{A}_P$ the *Quillen line-bundle* $\mathcal{L}$ (see Quillen 1986), whose curvature is $-2\pi \mathrm{i}$ times the Kähler form of $\mathcal{A}_P$, descends to give a well-defined line-bundle L on M_τ.

The crucial point now is that apparently the vector space $\mathcal{H}^k_\tau(\Sigma)$ *depends on the choice of the complex structure τ on* Σ against the desired general covariance of our theory. Actually $\mathcal{H}^k_\tau(\Sigma)$ varies holomorphically with τ giving rise to a holomorphic vector bundle over the moduli space of compact Riemann surfaces of fixed genus which turns out to admit a canonical *projectively flat connection which permits us to identify the fibres up to a scalar factor.* This fact can be proved in several ways as described in chapter 6 of Atiyah (1990a). See also Hitchin (1990) and Axelrod *et al* (1991) for more details.

The choice (9.42) then give rise to a *modular functor*

$$\Sigma \longrightarrow \mathcal{H}^k(\Sigma) \tag{9.43}$$

in the spirit of a *rational conformal field theory* as defined in Segal (1988): such a functor is well defined *up to a scalar factor.* It is a particular case of a *topological quantum field theory* (TQFT). Let us now briefly recall what it is that is axiomatized in Atiyah (1989). The interested reader may also consider chapter 2 in Atiyah (1990a) and appendix B.6 in Cox and Katz (1999) for some short reviews on the subject and Quinn (1995) for a broader treatment.

Definition 9.2.20 (Axiomatic TQFT). *A $(d+1)$-dimensional topological quantum field theory is a functor Z which associates*

- *with each compact oriented d-dimensional manifold Σ a finite-dimensional complex vector space Z_Σ,*

- *with each compact oriented* $(d+1)$*-dimensional manifold* M *whose boundary is* $\partial M = \Sigma$ *a vector* $Z(M) \in Z_\Sigma$,

and which satisfies the following axioms:

(i) (involutory) if $\overline{\Sigma}$ *denotes* Σ *with the opposite orientation and* Z^*_Σ *denotes the dual vector space of* Z_Σ *then*

$$Z_{\overline{\Sigma}} = Z^*_\Sigma$$

(ii) (multiplicativity) if $\amalg$ *denotes the disjoint union of* d*-manifolds then*

$$Z_{\Sigma_1 \amalg \Sigma_2} = Z_{\Sigma_1} \otimes Z_{\Sigma_2}$$

(iii) (associativity) if $\partial M_1 = \overline{\Sigma}_1 \amalg \Sigma_2$, $\partial M_2 = \overline{\Sigma}_2 \amalg \Sigma_3$ *and* $M = M_1 \cup_{\Sigma_2} M_2$ *is the gluing of* M_1 *and* M_2 *along* Σ_2 *then*

$$Z(M) = Z(M_2) \circ Z(M_1)$$

where, by the previous axioms,

$$\begin{aligned} Z(M_1) &\in Z^*_{\Sigma_1} \otimes Z_{\Sigma_2} = \hom_{\mathbb{C}}(\Sigma_1, \Sigma_2) \\ Z(M_2) &\in Z^*_{\Sigma_2} \otimes Z_{\Sigma_3} = \hom_{\mathbb{C}}(\Sigma_2, \Sigma_3) \\ Z(M) &\in Z^*_{\Sigma_1} \otimes Z_{\Sigma_3} = \hom_{\mathbb{C}}(\Sigma_1, \Sigma_3) \end{aligned}$$

(iv) (unity) if the empty set is considered as a compact d*-dimensional oriented manifold then*

$$Z_\emptyset = \mathbb{C}$$

(v) (identity) if I *denotes the oriented interval* $[0, 1]$ *let us consider the product* $(d+1)$*-manifold* $\Sigma \times I$ *whose boundary is* $\partial(\Sigma \times I) = \overline{\Sigma} \amalg \Sigma$*; then*

$$Z(\Sigma \times I) = \mathbb{I} \in \hom_{\mathbb{C}}(\Sigma, \Sigma)$$

where $\mathbb{I}$ *is the identity endomorphism of* Σ.

Let us now come back to the Hamiltonian formulation of Chern–Simons quantum field theory. In (9.43) we defined a correspondence

$$Z : \Sigma \longmapsto Z_\Sigma := \mathcal{H}^k(\Sigma)$$

between a compact surface $\Sigma \subset M$ and the finite dimensional complex vector space of 'physical states' of the k-level theory quantized along Σ by 'canonical quantization' . This turns out to give a TQFT giving the Hamiltonian interpretation of the partition function unrigorously expressed by path integral in (9.35). Precisely by writing

$$\begin{aligned} M &= M_1 \cup_\Sigma M_2 \\ \partial M_1 &= \overline{\emptyset} \amalg \Sigma \\ \partial M_2 &= \overline{\Sigma} \amalg \emptyset \end{aligned} \tag{9.44}$$

axioms 1, 2, 3 and 4 give

$$Z(M) = Z(M_2) \circ Z(M_1) \in \hom_{\mathbb{C}}(\mathbb{C}, \mathbb{C}) = \mathbb{C}. \tag{9.45}$$

It is the mathematically well defined evaluation of the homonymous partition function. It is completely topological and the scalar indetermination in defining Z_Σ does not even influence its value: actually $Z(M)$ does not even depend on the choice of Σ since $\partial M = \emptyset$ and $Z(M) \in Z_\emptyset$.

In order to perform an analogous Hamiltonian interpretation of the correlation function $Z(M; L, \mathcal{R})$ 'defined' by the path integral in (9.39) we have to relativize the definition of the TQFT Z to the tern $(M, L, \mathcal{R})$ given by a 3-manifold M and a link $L \subset M$ marked by a collection of irreducible representations $\mathcal{R}$ of G. Let us assume L to be transverse to $\partial M = \Sigma$ so that it gives a collection ∂L of signed points in Σ. Moreover, we can mark ∂L by a collection $\partial\mathcal{R}$ of irreducible representations of G induced by representations in $\mathcal{R}$. Let us write

$$\partial(M, L, \mathcal{R}) = (\Sigma, \partial L, \partial\mathcal{R}) \tag{9.46}$$

and then relativize Z by defining it as a functor which associates

- with each d-dimensional tern $(\Sigma, \partial L, \partial\mathcal{R})$ a finite-dimensional complex vector space $Z_{(\Sigma,\partial L,\partial\mathcal{R})}$,
- with each $(d+1)$-dimensional tern $(M, L, \mathcal{R})$, whose boundary is as in (9.46), a vector $Z(M; L, \mathcal{R}) \in Z_{(\Sigma,\partial L,\partial\mathcal{R})}$,

and which satisfies the axioms 1–5 of definition 9.2.20. The crucial point now is to relativize (9.43) to give an analogous definition of $Z_{(\Sigma,\partial L,\partial\mathcal{R})}$. Recall that by (9.27) the choice of a point $p \in \partial L \subset \Sigma = \partial M$ determines *a conjugacy class* in G. Since p is marked by an irreducible representation in $\partial\mathcal{R}$ the order of such a conjugacy class turns out to be the level k. Hence the collection ∂L of marked points in Σ gives rise to a set $C_{\partial L} := \{C_p\}_{p\in\partial L}$ of conjugacy class of order k in G. Let us denote by

$$\hom_{\partial L}(\pi_1(\Sigma \setminus \partial L), G)$$

the set of morphisms $\pi_1(\Sigma \setminus \partial L) \longrightarrow G$ sending a homotopy class of loops around $p \in \partial L$ into the conjugacy class C_p. Factoring out by conjugacy leads to the space

$$\mathcal{M}_{(\Sigma,\partial L,\partial\mathcal{R})} := \hom_{\partial L}(\pi_1(\Sigma \setminus \partial L), G)/G \tag{9.47}$$

which is the analogue of $\mathcal{M}_\Sigma$ as defined in (9.41). The quantization of $\mathcal{M}_{(\Sigma,\partial L,\partial\mathcal{R})}$ proceed now in the same way since the results of Narasimhan and Seshadri (1965) and Donaldson (1983) can be applied in this case too.

Theorem 9.2.21. *The space $\mathcal{M}_{(\Sigma,\partial L,\partial\mathcal{R})}$ is homeomorphic to a moduli space $M_\tau^{(k)}$ of holomorphic G-bundles over the Riemann surface Σ_τ obtained by the choice of a complex structure τ on Σ. On this space we have a natural choice for a line-bundle L_k whose holomorphic sections give the quantization at level k, i.e.*

$$\mathcal{H}_\tau^k(\Sigma, \partial L, \partial\mathcal{R}) := H^0(M_\tau^{(k)}, L_k). \tag{9.48}$$

Note that the introduction of Wilson lines also makes the moduli spaces $M_\tau^{(k)}$ dependent on the level k. As before, the finite dimensional complex vector space $\mathcal{H}_\tau^k(\Sigma, \partial L, \partial\mathcal{R})$ varies holomorphically with τ and give rise to a projectively flat holomorphic vector bundle over the moduli space of compact Riemann surfaces of fixed genus. Up to a scalar factor we have got the desired relativized modular functor

$$Z : (\Sigma, \partial L, \partial\mathcal{R}) \longmapsto Z_{(\Sigma,\partial L,\partial\mathcal{R})} := \mathcal{H}^k(\Sigma, \partial L, \partial\mathcal{R}).$$

Note that an evaluation of the expectation value $\langle W_L^{\mathcal{R}}\rangle$ defined by applying (9.38) and (9.39) needs to fix once for all the undefined scalar factor. It can be realized by the choice of a *framing* (see definition 9.2.24) for every knot composing the link L: here we shall not enter in details about by referring to Witten (1989) and Atiyah (1990b) for more details. In the next section we will consider the problem for the particular case in which L is the unknotted knot.

9.2.3 Computability and link invariants

Let us consider M to be as in (9.44). By (9.45) and axiom 1 in definition 9.2.20 we get

$$Z(M) = (\chi_1, \chi_2) \tag{9.49}$$

where $\chi_1, \chi_2 \in Z_\Sigma$. Similarly if we consider a Wilson observable $W_L^{\mathcal{R}}$ on M we get

$$Z(M; L, \mathcal{R}) = (\psi_1, \psi_2) \tag{9.50}$$

where $\psi_1, \psi_2 \in Z_{(\Sigma,\partial L,\partial\mathcal{R})}$.

These are the fundamental relations allowing the effective computation of $Z(M), Z(M; L, \mathcal{R})$ and $\langle W_L^{\mathcal{R}}\rangle$ essentially by connecting them with the link invariants of L in M.

In the present section, following Witten (1989), we compute some of these quantities when $M = S^3$ and $G = SU(N)$.

Proposition 9.2.22. *Assume $M = S^3$ and $G = SU(N)$. Then the expectation value $\langle W_L^{\mathcal{R}}\rangle$ of any Wilson observable can be inductively evaluated like a Jones polynomial $V_L(q)$ in the variable*

$$q := \exp\left(\frac{2\pi\mathrm{i}}{N+k}\right) \tag{9.51}$$

by applying the skein relation (9.73) and the mirror property (9.72), when L is considered in the standard framing and $\mathcal{R}$ is assigned by choosing the defining N-dimensional representation R of $SU(N)$ for every knot composing L. In particular if L is the unknotted knot K:

$$\langle W_K^R\rangle = \frac{q^{\frac{N}{2}} - q^{-\frac{N}{2}}}{q^{\frac{1}{2}} - q^{-\frac{1}{2}}} = \frac{\sin\left(\frac{N\pi}{N+k}\right)}{\sin\left(\frac{\pi}{N+k}\right)}. \tag{9.52}$$

Moreover

$$Z(S^3) = (k+N)^{-N/2}\sqrt{\frac{k+N}{N}}\prod_{j=1}^{N}\left\{2\sin\left(\frac{j\pi}{k+N}\right)\right\}^{N-j} \tag{9.53}$$

and

$$\begin{aligned} Z(S^3; K, R) = {}& \frac{2}{(k+N)^{N/2}}\sqrt{\frac{k+N}{N}}\sin^{N-2}\left(\frac{\pi}{k+N}\right) \\ & \times \sin\left(\frac{N\pi}{k+N}\right)\prod_{j=2}^{N-1}\left\{2\sin\left(\frac{j\pi}{k+N}\right)\right\}^{N-j}. \end{aligned} \tag{9.54}$$

Jones polynomials were first defined in Jones (1985) and then generalized in Jones (1987) as a particular case of a two-variable polynomial associated with a link by means of the Ocneanu trace of a Hecke algebra representation of its braid group. See also sections 1.3 and 1.4 in Atiyah (1990a) and section 2 in Labastida (1999) for quick, but aimed at our purpose, surveys on the argument.

Definition 9.2.23. *Denote by L_n a link whose planar projections admits n normal crossings and by L_{n+} and L_{n-} those links admitting $n+1$ normal crossings composed by the previous n and by a further crossing which is an over-crossing or an under-crossing, respectively. Given a link $L \subset S^3$ the Jones polynomial $V_L(q)$ is a Laurent polynomial in the variable $q^{\frac{1}{2}}$ inductively defined by the skein relation*

$$(q^{\frac{1}{2}} - q^{-\frac{1}{2}})V_{L_n}(q) - q^{\frac{N}{2}}V_{L_{n+}}(q) + q^{-\frac{N}{2}}V_{L_{n-}}(q) = 0 \tag{9.55}$$

and the mirror property

$$V_L(q) = V_{L'}(q^{-1}) \tag{9.56}$$

where L' is the mirror image of the link L.

To fix ideas start by considering the case in which L is given by two unlinked and unknotted circles K_1, K_2 and Σ is a 2-sphere S^2 which separates the two components of L without cutting any of them. Hence we get

$$\begin{aligned} & Z_{(\Sigma,\partial L,\partial\mathcal{R})} = Z_\Sigma = Z_{S^2} \\ & \psi_1 = Z(M_1; K_1, R_1) \\ & (\quad, \psi_2) = Z(M_2; K_2, R_2). \end{aligned}$$

Since $\dim_{\mathbb{C}} Z_{S^2} = 1$, all the vectors $\chi_1, \chi_2, \psi_1, \psi_2$ are multiples of the same vector. By (9.49) and (9.50) this gives

$$\begin{aligned} Z(M; L, \mathcal{R}) \cdot Z(M) &= (\psi_1, \psi_2)(\chi_1, \chi_2) \\ &= (\psi_1, \chi_2)(\chi_1, \psi_2) = Z(M; K_1, R_1) \cdot Z(M; K_2, R_2) \end{aligned}$$

whose quotient by $Z(M)^2$ is

$$\langle W_L^{\mathcal{R}} \rangle = \langle W_{K_1}^{R_1} \rangle \langle W_{K_2}^{R_2} \rangle. \tag{9.57}$$

By iterating such a relation for an arbitrary collection of unlinked and unknotted Wilson lines $L = \{K_i\}_{i=1}^r$ we obtain that

$$\langle W_L^{\mathcal{R}} \rangle = \prod_{i=1}^{r} \langle W_{K_i}^{R_i} \rangle. \tag{9.58}$$

A first consequence of such a multiplicativity on expectation values of unlinked and unknotted Wilson lines is that $\langle W_K^R \rangle \neq 0$ for an unknotted Wilson line, otherwise we would have a Chern–Simons theory which does not distinguish a knot from a link!

Let us now consider four marked points $\{p_j\}_{j=1}^4$ on $\Sigma = S^2$. They may be obtained either as the transversal section of the unlinked and unknotted link $L_0 = \{K_1, K_2\}$ (two-point section for a circle) or as a section of the two links L_+, L_- given by the two oriented knots whose planar normal crossings projection gives a figure eight (again two-point section for a circle): L_+ has an over-crossing while L_- an under-crossing. If we assume that the same representation R of G is associated with every knot composing these links we may arrange the four points to give

$$(\Sigma, \partial L_0, \partial \mathcal{R}_0) = (\Sigma, \partial L_+, \partial \mathcal{R}_+) = (\Sigma, \partial L_-, \partial \mathcal{R}_-) \tag{9.59}$$

$$= (S^2, \{p_j\}_{j=1}^4, \{R, R, \overline{R}, \overline{R}\}) =: \mathcal{H}. \tag{9.60}$$

If we have the decomposition

$$R \otimes R = \bigoplus_{h=1}^{s} E_h$$

where E_h is an irreducible representation of G, it turns out that

$$d := \dim_{\mathbb{C}} \mathcal{H} \leq s \tag{9.61}$$

and we get $d = s$ for large k (see Witten 1989, section 3). In particular if $G = SU(N)$ and R is the defining N-dimensional representation, then $s = 2$ and

$$d = \begin{cases} 1 & \text{if } k = 1 \\ 2 & \text{otherwise.} \end{cases} \tag{9.62}$$

For $i = 1, 2$ let us call M_i^0, M_i^+, M_i^- the two pieces cutted by S^2 in the three different cases. Note that the exterior pieces may be assumed:

$$M_1^0 = M_1^+ = M_1^- =: M_1 \tag{9.63}$$

while the interior pieces M_2^0, M_2^+, M_2^- may be thought to be related by a diffeomorphism on the boundary exchanging two of the four marked points. As in (9.50) the four pieces M_1, M_2^0, M_2^+, M_2^- determine four vectors:

$$\psi_1, \psi_2^0, \psi_2^+, \psi_2^- \in \mathcal{H}$$

whose products evaluate the associated partition functions. In fact these vectors are not known but the dimensional bound (9.61) may give rise to relations among them and their products which are similar to the defining relations of some link invariants. In particular when $G = SU(N)$ and all the knots are associated with the defining N-dimensional representation, the dimensional bound (9.62) allows us to conclude that $\psi_2^0, \psi_2^+, \psi_2^-$ are linearly dependent and so there must exist $\alpha, \beta, \gamma \in \mathbb{C}$ such that

$$\alpha(\psi_1, \psi_2^0) + \beta(\psi_1, \psi_2^+) + \gamma(\psi_1, \psi_2^-) = 0. \tag{9.64}$$

Hence the same relation can be established on the associated correlation functions as follows:

$$\alpha Z(M; L_0, \mathcal{R}_0) + \beta Z(M; L_+, \mathcal{R}_+) + \gamma Z(M; L_-, \mathcal{R}_-) = 0. \tag{9.65}$$

It actually gives a recursive relation among links L_n, L_{n+} and L_{n-}. In fact we can always cut these links by an S^2 leaving outside all the first n crossings: its interior then gives M_2^0, M_2^+, M_2^- again, respectively. Since α, β, γ depend only on the three vectors $\psi_2^0, \psi_2^+, \psi_2^-$, (9.64) does not depend on ψ_1 and we again get

$$\alpha Z(M; L_n, \mathcal{R}_n) + \beta Z(M; L_{n+}, \mathcal{R}_{n+}) + \gamma Z(M; L_{n-}, \mathcal{R}_{n-}) = 0. \tag{9.66}$$

We can then assume $\alpha \neq 0$ otherwise (9.66) would imply that up to a scalar factor we can exchange an over-crossing by an under-crossing i.e. every knot could be untied and our Chern–Simons theory would not distinguish topologically non-equivalent observables!

Since $M = S^3$ it is possible to continuously deform L_+ and L_- to an oriented circle K by applying a *Reidemeister moving*, i.e. a transformation induced on the planar image with normal crossings of a knot in S^3 by a homeomorphism applied to the original spacial knot (see Reidemeister 1933). By (9.65) we can then write

$$\alpha Z(M; \{K_1, K_2\}, \{R, R\}) + (\beta + \gamma) Z(M; K, R) = 0.$$

Divide by $Z(M)$ and recall (9.57) to get

$$\alpha \langle W_K^R \rangle \langle W_K^R \rangle + (\beta + \gamma) \langle W_K^R \rangle = 0.$$

Since $\langle W_K^R \rangle \neq 0$ we obtain

$$\langle W_K^R \rangle = -\frac{\beta + \gamma}{\alpha}. \tag{9.67}$$

Then by knowledge of α, β, γ (9.66) allows us to inductively determine $\langle W_L^{\mathcal{R}} \rangle$ for every L once we know a relation linking $\langle W_L^{\mathcal{R}} \rangle$ and $\langle W_{L'}^{\mathcal{R}'} \rangle$.

To determine α, β, γ let us concentrate on the boundary diffeomorphisms relating M_2^0, M_2^+, M_2^-. We can pass from L_+ to L_0 by exchanging two of the four marked points on the boundary S^2. Let us denote by

$$f : M_2^+ \longrightarrow M_2^0$$

this 'half monodromy' diffeomorphism. Note that

$$f \circ f : M_2^+ \longrightarrow M_2^-$$

since exchanging the same two points again we pass from L_0 to L_-. By the functoriality of TQFT we get an induced isomorphism $Z(f) \in \mathrm{Aut}\,(\mathcal{H})$ such that

$$\psi_2^- = Z(f)\psi_2^0 = Z(f)^2\psi_2^+. \tag{9.68}$$

Since $Z(f)$ must satisfy its characteristic equation we get the relation

$$\psi_2^- - (\mathrm{tr}\, Z(f))\psi_2^0 + (\det Z(f))\psi_2^+ = 0 \tag{9.69}$$

which allows us to completely determine α, β, γ by the knowledge of the eigenvalues of $Z(f)$. The latter are calculated when $M = S^3$ in Moore and Seiberg (1988). By comparing (9.64) and (9.69) and setting q as in (9.51) we can rewrite (9.66) for $M = S^3$ as follows

$$\begin{aligned}(q^{\frac{1}{2}} - q^{-\frac{1}{2}})Z(M; L_n, \mathcal{R}_n) - q^{\frac{1}{2N}} Z(M; L_{n+}, \mathcal{R}_{n+})&\\ + q^{-\frac{1}{2N}} Z(M; L_{n-}, \mathcal{R}_{n-}) = 0.&\end{aligned} \tag{9.70}$$

Hence by (9.67) the expectation value for the unknotted Wilson line is given by

$$\langle W_K^R \rangle = \frac{q^{\frac{1}{2N}} - q^{-\frac{1}{2N}}}{q^{\frac{1}{2}} - q^{-\frac{1}{2}}}. \tag{9.71}$$

This value does not coincide with (9.52) since the relation (9.70) is similar but not equal to the skein relation (9.55). The reason for such a discrepancy must be recovered in the implicit *framing* choice we made to write (9.64) which is not the same as the *standard framing* used in knot theory.

Definition 9.2.24. *A framing of a knot K is a closed curve K_f obtained as a small deformation of K along a normal vector field direction. The pair (K, K_f) is called a framed knot.*

At the end of the section 9.2.2 we noted that the evaluation of a Wilson observable expectation value $\langle W_L^{\mathcal{R}} \rangle$ needs to fix once for all the undefined scalar factors which occur in the projective definition of the Hamiltonian quantities via

TQFT. Actually by making assumptions (9.60) and (9.63) we made a particular choice for these scalar factors which do not coincide with the canonical choice usually adopted for knots in S^3 by requiring that the *Gauss self-linking number* be trivial for every knot (see Witten (1989 section 2.1) for the definition; see also Mariño and Vafa (2001 section 3) for a recent discussion of the problem in connection with the concept of a framed knot): this is what is usually meant by the *standard framing* of a knot.

Note that the coefficient associated with the unknotted unlinked L_0 is $q^{1/2} - q^{-1/2}$ both in (9.70) and in (9.55). Since by (9.68) we pass from ψ_2^0 to ψ_2^- by applying $Z(f)$ while its inverse $Z(f)^{-1}$ allows us to pass to ψ_2^+ we can argue that

$$q^{-\frac{N}{2}} q^{\frac{1}{2N}} = (q^{\frac{N}{2}} q^{-\frac{1}{2N}})^{-1} = \exp\left(\pi \mathrm{i} \frac{(1-N^2)}{N(N+k)}\right)$$

is the factor expressing the framing change through the half-monodromy f. It follows that, by adopting the standard framing, the expectation value (9.71) of the unknotted Wilson line must be rewritten as in (9.52). Although the skein relations (9.70) and (9.55) are not the same, the 'polynomials' defined by the former also satisfy the mirror property

$$\langle W_L^{\mathcal{R}} \rangle(q) = \langle W_{L'}^{\mathcal{R}'} \rangle(q^{-1}). \tag{9.72}$$

We can then conclude that the skein relation

$$(q^{\frac{1}{2}} - q^{-\frac{1}{2}}) \langle W_{L_n}^{\mathcal{R}_n} \rangle - q^{\frac{N}{2}} \langle W_{L_{n+}}^{\mathcal{R}_{n+}} \rangle + q^{-\frac{N}{2}} \langle W_{L_{n-}}^{\mathcal{R}_{n-}} \rangle = 0 \tag{9.73}$$

and the mirror property (9.72) allow us to inductively express in the standard framing the expectation value $\langle W_L^{\mathcal{R}} \rangle$ of any Wilson observable in S^3, when $G = SU(N)$ and all the representations associated with knots are the defining N-dimensional ones.

Note that when we fix $N = 2$ the unique variable is the level k of the theory while when N is general $\langle W_L^{\mathcal{R}} \rangle$ can be interpreted also as a *HOMFLY polynomial* (see Freyed *et al* (1985) for the definition of this two-variable polynomial invariant of links).

The skein relation (9.73) cannot evaluate the partition function $Z(S^3)$ and consequently the correlation function of any Wilson observable. Their evaluation follows by generalizing the previous procedure to every 3-manifold M.

Definition 9.2.25. *Let $K \subset S^3$ be an unknotted circle and T a tubular neighbourhood of K, i.e. a solid torus centred in K. Then*

$$S^3 = (S^3 \setminus T) \cup_\Sigma T$$

where $\Sigma := \partial T$ is a two-dimensional torus. If before the gluing we apply a diffeomorphism on the boundary ∂T then the gluing will give us a new three-manifold M which is said to be obtained by S^3 after a surgery on the knot K.

Proposition 9.2.26. *Any 3-manifold M can be obtained by S^3 up to a finite number of surgeries on knots. Hence the partition functions and expectation values on a general M can then be evaluated by those on S^3 once it is known how the repeated surgeries act on these quantities and on knots framing.*

An important application of this proposition is given by the manifold

$$M := S^2 \times S^1.$$

If we assume S^3 to be the compactification by a point of $\mathbb{R}^3$ and K to be the unit circle in the plain $z = 0$, consider the following surgery on K. Let Σ be a two-dimensional torus around K invariant under an inversion of $\mathbb{R}^3$: the tubular neighbourhood of K is the interior T_1 of Σ. Note that the exterior $T_2 = S^3 \setminus T_1$ is a solid torus too and we get

$$S^3 = T_1 \cup_\Sigma T_2. \tag{9.74}$$

However, if T_1, T_2 are considered as two solid tori which can be identified by a translation of $\mathbb{R}^3$ we get

$$S^2 \times S^1 = T_1 \cup_\Sigma T_2 \tag{9.75}$$

since $T_i = D_i \times S^1$, $\Sigma = S^1 \times S^1$ and $S^2 = D_1 \cup_{S^1} D_2$. (9.74) and (9.75) differ simply by the diffeomorphism applied on the boundary Σ to glue the solid tori T_i: in the former it is given by an inversion while in the latter by a translation.

This example is important because $Z(S^2 \times S^1; L, \mathcal{R})$ can be obtained by the TQFT axioms easier than $Z(S^3; L, \mathcal{R})$. Then we get a method to evaluate our partition functions on S^3 which is the main ingredient of the Witten's proof of a conjecture by Verlinde (1988) already proved in Moore and Seiberg (1988). In Verlinde (1988) it is shown how to get a basis $\{v_0, \ldots, v_{t-1}\}$ of Z_Σ canonically after the choice of an homology basis $\{\gamma_1, \gamma_2\}$ for $H_1(\Sigma, \mathbb{Z})$: T the interior of Σ the first basis vector v_0 is chosen to give $Z(T) \in Z_\Sigma$. The two solid tori T_1, T_2 giving $S^2 \times S^1$ in (9.75) are two identical copies of T identified by a translation. This gives

$$v_0 = Z(T_2) \qquad (v_0, \quad) = Z(T_1) \qquad (v_0, v_0) = Z(S^2 \times S^1). \tag{9.76}$$

However, if we consider Σ to be as in (9.74) the inversion of $\mathbb{R}^3$ acts on $H_1(\Sigma, \mathbb{Z})$ by sending

$$\begin{aligned} \gamma_1 &\longmapsto -\gamma_1 \\ \gamma_2 &\longmapsto \gamma_2. \end{aligned} \tag{9.77}$$

Let $\tau = a + ib$ be the complex number in the Siegel upper half-plane

$$\mathbb{H} := \{\tau \in \mathbb{C} : \mathrm{Im}(\tau) > 0\}$$

representing the isomorphism class of the complex torus Σ. The transformation induced on $\mathbb{H}$ by the inversion acts as follows

$$\tau = a + ib \longmapsto \frac{1}{|\tau|^2}(-a + ib) = -\tau^{-1}.$$

This is the modular transformation represented by the element

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

in the modular group

$$\Gamma := SL(2, \mathbb{Z})/\{\pm I\}$$

where $I = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. Γ acts on $\mathbb{H}$ by setting

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \tau = (a\tau + b)(c\tau + d)^{-1}.$$

Since the isomorphism classes of complex tori are parametrized by the *modular curve* $\Gamma \setminus \mathbb{H}$ it turns out that the inversion realizes a diffeomorphism of Σ which preserves the complex structure (see the first chapter in Silverman [1994] for further details about and a careful construction of the quotient $\Gamma\setminus\mathbb{H}$). It induces an isomorphism on Z_Σ which can be represented on the Verlinde basis by a complex $t \times t$ matrix S_i^j such that

$$v_i = \sum_j S_i^j v_j.$$

Therefore by (9.74) and (9.76) we get

$$Z(S^3) = \left(v_0, \sum_j S_0^j v_j \right) = \sum_j S_0^j (v_0, v_j).$$

This formula gives an effective evaluation of $Z(S^3)$ since the numbers $g_{ij} := (v_i, v_j)$ and the matrix S_i^j are given by knowledge of the Verlinde basis of Z_Σ. Hence by setting $S_{i,j} := \sum_k S_i^k g_{jk}$ we get

$$Z(S^3) = S_{0,0}.$$

When $G = SU(N)$ we obtain the following result:

$$S_{0,0} = (k + N)^{-N/2} \sqrt{\frac{k + N}{N}} \prod_{j=1}^{N} \left\{ 2 \sin \left(\frac{j\pi}{k + N} \right) \right\}^{N-j} \tag{9.78}$$

allowing to conclude (9.53). By recalling (9.52) we are able to write $Z(S^3; K, R)$ as in (9.54) for the unknotted knot K in the defining N-dimensional representation R of $SU(N)$.

9.3 The Gopakumar–Vafa conjecture

This section discusses the conjecture itself, its origin and its relation to geometric transitions. We also presents supporting evidence, which leads to the uncharted territory of 'open Gromow–Witten invariants'.

We start with the original observation by Gopakumar and Vafa (by comparing the partition functions) and show, in this first part, how Witten's interpretation of the Chern–Simons theory as an open-string theory Witten (1992) provides the tools for the geometric interpretation of the duality.

Conjecture 9.3.1 (Gopakumar and Vafa 1998a, b). *(notation as in 9.1.1) The $SU(N)$–Chern–Simons theory on $S^3 \subset \widehat{Y} := T^*S^3$ of level k is equivalent, for large, N to a type IIA closed-string theory (with fluxes) on the local Calabi–Yau $Y := \mathcal{O}_{\mathbb{P}^1}(-1) \oplus \mathcal{O}_{\mathbb{P}^1}(-1)$.*

(The language used here reflects the reformulation of the conjecture given in Ooguri and Vafa (2000) rather then the original one.)

Theorem 9.3.2 (Witten 1992). *Let $\widehat{Y} = T^*L$ be a local Calabi–Yau threefold. Then there exist topological string theories with $\widehat{Y}$ as target space, such that their open sectors are exactly equivalent to a QFT on L.*

Conjecture 9.3.3 (Gopakumar and Vafa after Witten). *A topological open-string theory of type IIA on $\widehat{Y} := T^*S^3$ with N D6-branes wrapped around the base S^3 is equivalent, for large N, to a type IIA closed-string theory on the local Calabi–Yau $Y := \mathcal{O}_{\mathbb{P}^1}(-1) \oplus \mathcal{O}_{\mathbb{P}^1}(-1)$ with N units of 2-form Ramond–Ramond flux through the exceptional S^2.*

The transition from Y to $\widehat{Y}$ realizes the geometrical model of a physical closed/open duality among string theories of type IIA. That is, the transition from Y to $\widehat{Y}$ realizes the geometrical model of a physical duality relating a particular type IIA closed string theory on Y and the $SU(N)$–Chern–Simons QFT on the Lagrangian submanifold S^3 of $\widehat{Y}$ for large N.

This formulation of conjecture 9.3.1 has already been given in Gopakumar and Vafa (1999); see also Ooguri and Vafa (2000). See Vafa (2001), for the correspondence among D6-branes and units of R–R flux.

Witten's work is more general: he proposes a string theory interpretation of the Chern–Simons $U(N)$-gauge theory on a real three-dimensional Lagrangian submanifold L of a complex Calabi–Yau threefold $\widehat{Y}$ and also extends it beyond the hypothesis

$$\widehat{Y} = T^*L. \tag{9.79}$$

We refer to appendix 9.9 for more details.

Sketch of the proof: How theorem 9.3.2 implies 9.3.1 $\leftrightarrow$ 9.3.3.

Witten constructs an 'A-twisted sigma model' on $\widehat{Y}$. In particular he consider maps ϕ from a Riemann surface Σ with boundary $\partial\Sigma$, to the target space Y (i.e. ϕ

is a bosonic field of the open sector of this A-model) satisfying some conditions. The most important assumption is that

$$\phi(\partial\Sigma) \subset L. \tag{9.80}$$

There are also boundary conditions, involving derivatives of ϕ along the components of $\partial\Sigma$ and the fermionic fields. These conditions are needed to preserve fermionic symmetry but they do not directly enter the geometric picture (see section 3.1 in Witten [1992] for more details). If $Y = T^*L$, the weak coupling limit of the abstract string Lagrangian reduces exactly to the Lagrangian, of a QFT on L *'there are neither perturbative correction nor instanton, i.e. corrections'* (see definition 9.3.8). *In the A-twisted case such a limit turns out to be exactly a Chern–Simons $U(N)$-gauge theory.*

Gopakumar and Vafa observed these boundary conditions may be expressed in terms of D-branes (see A Lerda's article in the same volume) by saying that the Witten's open string theory is an *A-model topological open string theory with N topological D6-branes wrapped on L.* ◇

9.3.1 Matching the free energies

In the next two subsections, we review the evidence for the conjectures 9.3.1 and 9.3.3. The first evidence is given by the *matching of the 'free energies'* (or equivalently *partition functions*) for the theories involved by the conjecture. The second one is given by *comparisons of the expectation values of observables* in the two theories.

Theorem 9.3.4. *The genus g contribution to the free energy (9.82) of the Chern–Simons theory on S^3 coincides with the genus g contribution to the free energy of the closed string theory on $\mathcal{O}_{\mathbb{P}^1}(-1) \oplus \mathcal{O}_{\mathbb{P}^1}(-1)$.*

The Chern–Simons side

Definition 9.3.5. *Let $Z(S^3)$ be the partition function given by (9.53). Set*

$$F(S^3) = -\log Z(S^3). \tag{9.81}$$

Proposition 9.3.6 ('t Hooft 1974, Periwal 1993). *For large N, the free energy (9.81) of a $SU(N)$-gauge Chern–Simons QFT on S^3 can be expanded as follows,*

$$F(S^3) = \sum_{g\geq 0} \mathcal{F}_g(\tau) N^{2g-2}. \tag{9.82}$$

Here

$$\lambda := \frac{2\pi}{k+N} \tag{9.83}$$

is the Chern–Simons coupling constant, $\tau := \lambda N$ the 't Hooft coupling constant. The weak-coupling limit $\lambda \to 0$, $N \to +\infty$ leave constant the 't Hooft coupling constant.

Sketch of the proof. The statement follows by observing that, in the 'double line notation', Feynman diagrams contributing to the free energy F may be thought of as a sort of 'triangulation' of a compact, connected topological surface given by an admissible subdivision of the topological surface in polygons and discs. The latter occur as the internal planar regions of loops in Feynman diagrams: they should be understood as polygons admitting two edges and two vertices. 't Hooft observed that the contribution due to a Feynman diagram is proportional to $\lambda^{e-v}N^{h-l}$ where l is the number of diagram loops (*quark loops* in 't Hooft notation) and e, v, h are the number of edges, vertices and faces respectively, in the induced 'triangulation'. Since a diagram loop increases h by 1 and e, v by 2, the contribution due to a Feynman diagram without loops and admitting $h' = h - l$ faces is proportional to $\lambda^{e-v}N^{h-l}$ as well. The Euler characteristic formula

$$2 - 2g = h - e + v$$

allows us to conclude that the Feynman diagrams' contributions to the free energy F can be labelled by the genus g of the topological surface and the number of faces h of the induced 'triangulation'. The associated contribution is then proportional to $\lambda^{2g-2+h}N^h$ giving

$$F = \sum_g \left(\sum_h C_{g,h} \lambda^{2g-2+h} N^h \right)$$

where $C_{g,h}$ are suitable coefficients computed by Periwal. If we now consider the weak-coupling limit $\lambda \to 0$, $N \to +\infty$ leaving $\tau = \lambda N$ constant, then the free-energy expansion can be reorganized as follows.

$$F = \sum_g \left(\sum_h C_{g,h} \tau^{2g-2+h} \right) N^{2-2g} = \sum_g \mathcal{F}_g(\tau) N^{\chi(g)}. \qquad \diamond$$

Lemma 9.3.7. *Let*

$$Z(S^3) = (k+N)^{-N/2} \sqrt{\frac{k+N}{N}} \prod_{j=1}^{N} \left\{ 2 \sin\left(\frac{j\pi}{k+N} \right) \right\}^{N-j}$$

be the Chern–Simons partition function, as in (9.53).

Set $F(S^3) = -\log Z(S^3)$ and

$$t = \frac{2\pi \mathrm{i} N}{k+N} \qquad \lambda = \frac{2\pi}{k+N}$$

as in (9.83). The 't Hooft topological expansion for large N (9.82) becomes, for small λ,

$$F(\lambda, t) = \sum_{g=0}^{+\infty} F_g(t)\lambda^{-\chi(g)} \tag{9.84}$$

where $F_g(t) = \tau^{\chi(g)}\mathcal{F}_g(\tau) = (-1)^{g+1}t^{\chi(g)}\mathcal{F}_g(-it)$. In particular,

$$\begin{aligned}
F_0(t) &= \frac{i\pi^2}{6}t - i\left(m + \frac{1}{4}\right)\pi t^2 + \frac{i}{12}t^3 - \sum_{d=1}^{+\infty} d^{-3}(1 - e^{-dt}) \\
F_1(t) &= \tfrac{1}{24}t + \tfrac{1}{12}\log(1 - e^{-t}) \\
F_g(t) &= \frac{(-1)^g B_{2g}}{2g(2g-2)!}\left(\frac{B_{2g-2}}{(2g-2)} + \sum_{d=1}^{+\infty} d^{2g-3}e^{-dt}\right) \qquad \forall g \geq 2
\end{aligned} \tag{9.85}$$

where m is an arbitrary integer coming from the polydromic behaviour of the complex logarithm and B_h is the hth Bernoulli number defined by

$$\frac{x}{e^x - 1} = \sum_{h=0}^{+\infty} B_h \frac{x^h}{h!}.$$

In the physics literature the $2g$th Bernoulli number is often denote by B_g instead of B_{2g}.

The explicit computation of the expansion coefficients can be performed either starting from $\mathcal{F}_g(\tau)$ as in Periwal (1993) (expansion for large N) or from $F_g(t)$ by following Gopakumar and Vafa (1999, 1998a) (expansion for small λ). The key ingredient in expanding $F(S^3)$ is to employ the Mittag–Leffler expansion for the logarithmic derivative of the complex function $\sin(z)/z$. When $z = j\lambda/2$ we get the following relation:

$$\sin\left(\frac{j}{2}\lambda\right) = \frac{j}{2}\lambda \prod_{d=1}^{+\infty}\left(1 - \frac{j^2\lambda^2}{4\pi^2 d^2}\right)$$

which introduced in (9.78) gives (9.84). The interested reader should consult the references cited earlier.

The IIA theory side

Definition 9.3.8. *Given a topological string theory whose target space is a complex manifold Y, a world-sheet instanton (or simply instanton) of genus g is a holomorphic map*

$$\phi : \Sigma \longrightarrow Y$$

from a Riemann surface of genus g. If the boundary $\partial\Sigma$ is not empty ϕ is said to be open, since a similar instanton is typical of an open string. In the A-twisted context ϕ represent a bosonic elementary field.

In our case the only non-trivial homology class in Y is the exceptional $\mathbb{P}^1$: the 'string amplitude' 'counts' instantons with the exceptional $\mathbb{P}^1$ as image:

Lemma 9.3.9. *Let*

$$F^{(s)}(\lambda, t) = \sum_{g=0}^{+\infty} F_g^{(s)}(t)\lambda^{-\chi(g)}$$

be the perturbative expansion of the free energy (or better: the 'string amplitude') of the type IIA (closed) string theory (with string constant λ) on the local Calabi–Yau $Y = \mathcal{O}_{\mathbb{P}^1}(-1) \oplus \mathcal{O}_{\mathbb{P}^1}(-1)$. The coefficients $F_g^{(s)}(t)$ determine the 'local' Gromow–Witten invariants of Y, associated with maps of Riemann surfaces with the homology class of the exceptional locus $\mathbb{P}^1 \cong S^2 \subset Y$ as image. With the identification of λ as in (9.83) and

$$t := \frac{2\pi \mathrm{i} N}{k+N} \tag{9.86}$$

we have

$$F_g^{(s)}(t) = F_g(t) \qquad \forall g.$$

(t is interpreted as the Kähler modulus of the exceptional locus $S^2 \cong \mathbb{P}^1$ in Y.)

The contribution $F_g^{(s)}(t)$ to the string amplitude $F^{(s)}(\lambda, t)$ given by all the genus g instantons is called *the genus g instanton correction.*

Definition 9.3.10. *Let (X, g) be a Kähler manifold; fix a closed 2-form B on X and denote by $J \in H^2(X, \mathbb{R})$ the Kähler class of the Hermitian metric g. The cohomology class of the form $\omega = B + \mathrm{i}J$ is called the complexified Kähler class associated with g. The Kähler modulus of a given real 2-cycle $Z \subset X$ is defined by the period*

$$\int_Z \omega \in \mathbb{C}$$

of the complexified Kähler class on it. By Stoke's theorem and Kähler condition on J it is well defined for the entire homology class of Z.

Sketch of the proof of lemma 9.3.9: $F_0^{(s)}(t)$ is recovered in Candelas *et al* (1991): with our parameters we get

$$F_0^{(s)}(t) = \frac{\mathrm{i}\pi^2}{6}t - \mathrm{i}a\pi t^2 + \frac{\mathrm{i}}{12}t^3 - \sum_{d=1}^{+\infty} d^{-3}(1 - \mathrm{e}^{-dt}).$$

The coefficient a does not have a direct topological interpretation on Y. Hosono *et al* (1995) argue that $a = 1/4$ giving the match with $F_0(t)$ when $m = 0$.

The computation of $F_1^{(s)}(t)$ and $F_2^{(s)}(t)$ can be found in Bershadsky *et al* (1993) and (1994) respectively: in our situation they match $F_1(t)$ and $F_2(t)$ exactly.

Faber and Pandharipande (2000) compute $F_g^{(s)}(t)$ for every genus $g \geq 2$, i.e. for all the values of g for which the compactified moduli space $\overline{M}_g$, given by projective, connected, nodal, Deligne–Mumford stable curves of arithmetic genus g, is an irreducible variety of dimension $3g-3$ with orbifold singularities if regarded as an ordinary coarse moduli space (it is smooth if regarded as Deligne–Mumford stacks: see Fulton and Pandharipande (1995) and chapter 7 in Cox and Katz (1999) for a general reference). Faber and Pandharipande (2000) show that

> *The coefficients $F_g(t)$ determine the 'local' Gromow–Witten invariants of Y, associated with maps of Riemann surfaces with the homology class of the exceptional locus $\mathbb{P}^1 \cong S^2 \subset Y$ as image.*

In particular for $g \geq 2$ we can write

$$F_g^{(s)}(t) = -\langle 1 \rangle_{g,0}^Y - \sum_{d=1}^{+\infty} C(g,d) \mathrm{e}^{-dt} \tag{9.87}$$

where $\langle 1 \rangle_{g,0}^Y$ is the genus g, degree 0 Gromov–Witten invariant of our Calabi–Yau Y giving the instanton correction due to constant maps. On the other hand the series on the right gives, for every d, the instanton correction due to maps realizing a d-covering with genus g of the exceptional $\mathbb{P}^1$. Theorem 3 in Faber and Pandharipande (2000) gives

$$C(g,d) = |\chi(\overline{M}_g)| \frac{d^{2g-3}}{(2g-3)!}$$

where $\chi(\overline{M}_g)$ is the orbifold Euler characteristic of the coarse moduli space $\overline{M}_g$. It can be expressed in terms of Bernoulli numbers by means of the following Harer–Zagier formula

$$\chi(\overline{M}_g) = \frac{B_{2g}}{2g(2g-2)}.$$

Therefore we get

$$C(g,d) = \frac{|B_{2g}| d^{2g-3}}{2g(2g-2)!}. \tag{9.88}$$

Note that when $g = 0, 1$ the instanton correction due to non-constant maps admits a similar series presentation whose coefficients are known. In particular, the genus 0 case is settled by the Aspinwall–Morrison formula

$$C(0,d) = d^{-3}$$

(see Aspinwall and Morrison (1993), Manin (1995), Voisin (1996)) and it is easy to recover its contribution in the series comparing in $F_0(t)$. For the genus 1 case see Graber and Pandharipande (1999): in our particular situation it turns out that the non-constant instanton correction is due only to 1-coverings and is given by $1/12 \log(1-\mathrm{e}^{-t})$.

We still need to compute $\langle 1\rangle^{Y}_{g,0}$ in (9.87). This is theorem 4 from Faber and Pandharipande (2000). Consider the rank g vector bundle $\mathbb{E} \to \overline{M}_g$ whose fibre over the Deligne–Mumford stable curve C is given by $H^0(C, \omega_C)$ (here ω_C is the dualizing sheaf of C, *the Hodge bundle of* $\overline{M}_g$). If $c_j(\mathbb{E})$ is the jth Chern class of $\mathbb{E}$ then

$$c^3_{g-1}(\mathbb{E}) := c_{g-1}(\mathbb{E}) \wedge c_{g-1}(\mathbb{E}) \wedge c_{g-1}(\mathbb{E})$$

is a top form over $\overline{M}_g$. A result in Getzler and Pandharipande (1998) applied to our Calabi–Yau Y gives

$$\langle 1\rangle^{Y}_{g,0} = (-1)^g \int_{\overline{M}_g} c^3_{g-1}(\mathbb{E}). \tag{9.89}$$

Faber and Pandharipande then show that

$$\int_{\overline{M}_g} c^3_{g-1}(\mathbb{E}) = \frac{|B_{2g}|}{2g} \frac{|B_{2g-2}|}{2g-2} \frac{1}{(2g-2)!}.$$

Since $|B_{2g}| = (-1)^{g+1} B_{2g}$, the relations (9.87), (9.88) and (9.89) imply

$$\begin{aligned} F^{(s)}_g(t) &= (-1)^{g+1} \int_{\overline{M}_g} c^3_{g-1}(\mathbb{E}) - \sum_{d=1}^{+\infty} \frac{|B_{2g}| d^{2g-3}}{2g(2g-2)!} \mathrm{e}^{-dt} \\ &= \frac{B_{2g}}{2g(2g-2)!} \left(\frac{|B_{2g-2}|}{(2g-2)} + (-1)^g \sum_{d=1}^{+\infty} d^{2g-3} \mathrm{e}^{-dt} \right) = F_g(t). \qquad \diamond \end{aligned}$$

9.3.2 The matching of expectation values

Here we discuss the *matching of the expectation values of observables* in the two theories of the conjecture 9.3.1. The conjecture would be proved if the expectation values for any observable would coincide. Unfortunately it is not known how to produce a similar 'universal comparison theorem' but a general set-up to compare some kind of observable has been performed and the matching of expectation values has been proved in some particular case. In this section we present this strategy and its striking mathematical consequences.

The basic idea has already been suggested in Gopakumar and Vafa (1999) and then developed in Ooguri and Vafa (2000), Labastida and Mariño (2001), Katz and Liu (2001) and Li and Song (2001). In Chern–Simons theory observables are assigned by Wilson lines or products of them whose correlation functions are given by (9.37) and (9.39) respectively. It is not clear *a priori* what these functions correspond to on the topological closed-string theory side but there are some leads.

First, Witten's open-string interpretation of Chern–Simons theory also gives a translation of the correlation functions of Wilson observables in terms of instantons contributions:

Proposition 9.3.11. *An observable in $SU(N)$ Chern–Simons gauge theory represented by a link $\mathcal{L}$ corresponds in the Witten open-string theory interpretation to the Lagrangian submanifold $\mathcal{C}_{\mathcal{L}}$ given by the conormal bundle in $T^*S^3|_{\mathcal{L}}$.*

The non-constant instanton contributions of a type IIA open-string theory with non-compact D-branes wrapped on $\mathcal{C}_{\mathcal{L}}$ give a string theory interpretation of the correlation function of $\mathcal{L}$.

Definition 9.3.12. *Let $\mathcal{K}$ be a knot in S^3, parametrized by $\mathbf{q} = \mathbf{q}(s)$ for $s \in [0, 2\pi)$. For any s consider the plane $\pi_s \subset \mathbb{R}^4(\mathbf{p})$ of equations*

$$\sum_{j=1}^{4} q_j(s) p_j = 0$$
$$\sum_{j=1}^{4} \dot{q}_j(s) p_j = 0.$$

The three-dimensional submanifold $\mathcal{C}_{\mathcal{K}} := \coprod_s \pi_s$ is called the conormal bundle of $\mathcal{K}$.

Lemma 9.3.13. *$\mathcal{C}_{\mathcal{K}}$ is a Lagrangian submanifold with respect to the symplectic structure induced on T^*S^3 by the differential of the Liouville form $\vartheta := \sum_{j=1}^{4} p_j \,\mathrm{d}q_j$ of $\mathbb{R}^8$.*

Proof. Consider T^*S^3 as embedded in $\mathbb{R}^8 = \mathbb{R}^4(\mathbf{q}) \times \mathbb{R}^4(\mathbf{p})$ by the equations (9.16). For any s consider the plane $\pi_s \subset \mathbb{R}^4(\mathbf{p})$ of equations

$$\sum_{j=1}^{4} q_j(s) p_j = 0$$
$$\sum_{j=1}^{4} \dot{q}_j(s) p_j = 0.$$

Then

$$\vartheta|_{\mathcal{C}_{\mathcal{K}}} = \sum_{j=1}^{4} \dot{q}_j(s) p_j \,\mathrm{d}s = 0. \tag{9.90}$$

□

Then we can try to understand how the conifold transition acts on those instantons:

Theorem 9.3.14. *For a suitable link $\mathcal{L}$, the correlation function of the related observable in $SU(N)$ Chern–Simons gauge theory corresponds, on the IIA string theory on $\mathcal{O}_{\mathbb{P}^1}(-1)\oplus\mathcal{O}_{\mathbb{P}^1}(-1)$, to 'open Gromov–Witten invariants' of maps from Riemann surfaces with boundary on $\mathbb{P}^1$ determined by $\mathcal{L}$.*

The class of 'suitable' links $\mathcal{L}$ in the statement includes torus knots.

Sketch of the proof of proposition 9.3.11 Witten (1992) shows that one can reproduce the correlation function of a Chern–Simons observable by introducing further D-branes wrapping around a suitable Lagrangian submanifold of $\widehat{Y} = T^*S^3$ which is not the base S^3 and considering the partition function of the limit QFT.

In Gopakumar and Vafa (1999) and Ooguri and Vafa (2000) Wilson line observable represented by a knot $\mathcal{K} \subset S^3$ is associated with the 'conormal bundle' $\mathcal{C}_{\mathcal{K}}$ whose total space turns out be a Lagrangian submanifold. Precisely, let us consider T^*S^3 as embedded in $\mathbb{R}^8 = \mathbb{R}^4(\mathbf{q})\times\mathbb{R}^4(\mathbf{p})$ by the equations (9.16) end let $\mathcal{K} \subset S^3 \subset \mathbb{R}^4(\mathbf{q})$ be assigned by the parametrization $\mathbf{q} = \mathbf{q}(s)$ for $s \in [0, 2\pi)$. For any s consider the plane $\pi_s \subset \mathbb{R}^4(\mathbf{p})$ of equations

$$\sum_{j=1}^{4} q_j(s)p_j = 0$$

$$\sum_{j=1}^{4} \dot{q}_j(s)p_j = 0.$$

The three-dimensional submanifold $\mathcal{C}_{\mathcal{K}} := \coprod_s \pi_s$ (endowed with the induced differential structure) is a Lagrangian submanifold with respect to the symplectic structure induced on T^*S^3 by the differential of the Liouville form $\vartheta := \sum_{j=1}^{4} p_j \,\mathrm{d}q_j$ of $\mathbb{R}^8$ since

$$\vartheta|_{\mathcal{C}_{\mathcal{K}}} = \sum_{j=1}^{4} \dot{q}_j(s)p_j \,\mathrm{d}s = 0. \tag{9.91}$$

Then the open-string theory having T^*S^3 as target space and boundary conditions represented by M topological D6-branes wrapped on $\mathcal{C}_{\mathcal{K}}$ is *exactly equivalent to a $SU(M)$ Chern–Simons gauge theory*, since the boundary condition $\partial\phi \subset \mathcal{C}_{\mathcal{K}}$, which is the analogue of (9.80), is satisfied for 'every bosonic field' ϕ. But globally we now have an 'A-twisted σ-model' whose open sector also contains open strings having one end on S^3 and the other on $\mathcal{C}_{\mathcal{K}}$: the non-constant instantons associated with their world-sheet give a non-trivial contribution to the string amplitude. This means that the low-energy limit QFT is a $SU(N)\otimes SU(M)$ gauge theory which is no more a Chern–Simons theory but a deformation of it.

Because $\mathcal{C}_{\mathcal{K}} \cong \mathcal{K} \times \mathbb{R}^2$ and S^3 is simply connected, Witten's argument shows that *this partition function is strictly related with the correlation function of the original observable associated with $\mathcal{K}$ in the $SU(N)$ Chern–Simons theory on S^3*. Precisely, if $S(\mathcal{L}_{\mathcal{C}_{\mathcal{K}}})$ is the Chern–Simons action of the $SU(M)$ gauge theory on $\mathcal{C}_{\mathcal{K}}$ defined as in (9.32) then the partition function of the limit QFT is defined by a Feynman integration of the following Chern–Simons deformed action

$$S(\mathcal{L}_{\mathcal{C}_{\mathcal{K}}}) - \frac{\mathrm{i}}{2\pi k}\sum_d \eta_d \log(\mathrm{Tr}_R(h_{\mathcal{K}}^d)) \tag{9.92}$$

where $h_{\mathcal{K}}$ is the holonomy operator on $\mathcal{K}$ with respect to a connection $\widetilde{A}$ of the $SU(M)$ principal bundle over $\mathcal{C}_{\mathcal{K}}$ and $\eta_d = \pm 1$ for any d (see corollary 9.9.2 in appendix 9.9).

The statement of proposition 9.3.11 follows by repeating this construction for every knot in $\mathcal{L}$. ◇

Sketch of the proof of theorem 9.3.14 for $\mathcal{L} = \mathcal{K}$, the un-knot: We now fix a knot $\mathcal{K}$, consider the conormal Lagrangian submanifold $\mathcal{C}_{\mathcal{K}}$ and study its image, through the conifold transition, on $Y = \mathcal{O}_{\mathbb{P}^1}(-1) \oplus \mathcal{O}_{\mathbb{P}^1}(-1)$. Such a procedure can easily be realized when $\mathcal{K}$ is the unknotted knot. Consider, in fact, the involution of $\mathbb{C}^4(x, y, z, t)$ given by

$$(x, y, z, t) \longmapsto (\overline{x}, \overline{y}, -\overline{z}, -\overline{t}).$$

Recalling now the chain of transformations given by (9.4), (9.6) and (9.15) we see that such an involution act on $\mathbb{R}^8(\mathbf{q}, \mathbf{p})$ as follows.

$$(q_1, q_2, q_3, q_4, p_1, p_2, p_3, p_4) \longmapsto (q_1, -q_2, -q_3, q_4, -p_1, p_2, p_3, -p_4). \tag{9.93}$$

We then have the following three properties:

(i) T^*S^3 turns out to be fixed by the involution (9.93) as follows from its embedding equations (9.16) in $\mathbb{R}^8$;
(ii) the symplectic form

$$\omega = \mathrm{d}\vartheta = \sum_{j=1}^{4} \mathrm{d}p_j \wedge \mathrm{d}q_j$$

changes its sign under (9.93); and
(iii) the set of fixed points of (9.93) is given by

$$\mathcal{F} := \{(\mathbf{q}, \mathbf{p}) : q_2 = q_3 = p_1 = p_4 = 0\}.$$

These properties imply that $\mathcal{C} := \mathcal{F} \cap T^*S^3$ is a Lagrangian submanifold with respect to the symplectic structure induced by ω on T^*S^3 whose equation in $\mathbb{R}^8(\mathbf{q}, \mathbf{p})$ turns out to be

$$\begin{gathered} q_1^2 + q_2^2 - 1 = q_2 = q_3 = 0 \\ p_1 = p_4 = 0 \end{gathered} \tag{9.94}$$

Hence topologically $\mathcal{C} \cong S^1 \times \mathbb{R}^2$ and $\mathcal{K} := \mathcal{C} \cap S^3$ is an equator of S^3, i.e. it is the unknotted knot on S^3 and $\mathcal{C} = \mathcal{C}_{\mathcal{K}}$. Recall now that, by Clemens theorem 9.1.5, the conifold transition can be locally realized like a surgery by means of the diffeomorphism on boundaries α represented in (9.17) whose equations are

$$q_j = \frac{u_j}{\sqrt{\sum_i u_i^2}}$$
$$p_j = v_j \sqrt{\sum_i u_i^2}.$$

Hence the image of $\mathcal{C}$ in the blow-up

$$Y = \mathcal{O}_{\mathbb{P}^1}(-1) \oplus \mathcal{O}_{\mathbb{P}^1}(-1) \longrightarrow \overline{Y}$$

is the strict transform $\widetilde{\mathcal{C}}$ of the subvariety described in $\overline{Y}$ by conditions (9.94). Recall that $\overline{Y}$ has local equations (9.7) in $\mathbb{R}^8(\mathbf{u}, \mathbf{v})$. Then $\widetilde{\mathcal{C}}$ is the strict transform of the three-dimensional degenerate hyperquadric of rank 4:

$$u_1^2 + u_4^2 - v_2^2 - v_3^2 = u_2 = u_3 = v_1 = v_4 = 0.$$

Restrict the diffeomorphism (9.8) to this hyperquadric: outside of the exceptional fibre it is then topologically equivalent to $(\mathbb{R}_{>0} \times S^1) \times S^1$. By extending (9.8) over the exceptional locus as in (9.12) we get the following topological interpretation of the strict transform

$$\widetilde{\mathcal{C}} \cong \mathbb{R}^2 \times S^1$$

where the second factor S^1 is an equator of the exceptional locus S^2. Note that $\widetilde{\mathcal{C}} \cap S^2 = S^1$, the equator in the exceptional locus S^2.

By this general picture in Ooguri and Vafa (2000) it is argued that the Chern–Simons deformation (9.92) due to the Wilson line associated with the unknot (or, equivalently, due to non-constant instantons landing on $\mathcal{C}$ by proposition 9.3.11) can be evaluated, in the topological type IIA string theory on Y, by holomorphic non-constant instantons sending a Riemann surface with boundary onto either the upper or lower hemisphere of the exceptional S^2, with respect to the equator $\widetilde{\mathcal{C}} \cap S^2$ where the Riemann surface boundary is sent. In fact, starting from (9.52) Ooguri and Vafa compute the deformation term in (9.92) and obtain that, for large N, it is given by $-\mathrm{i}\Phi(\lambda, t, \mathcal{K})$ where

$$\Phi(\lambda, t, \mathcal{K}) = \sum_d \frac{\mathrm{Tr}_R(h_{\mathcal{K}}^d) + \mathrm{Tr}_R(h_{\mathcal{K}}^{-d})}{2d \sin(d\lambda/2)} \mathrm{e}^{-dt/2}.$$

The terms of the series on the right can be thought of as a sort of *Gromov–Witten invariants of maps from Riemann surfaces with boundary to the disc.* Ooguri and Vafa, using ***M***-theory duality (see Gopakumar and Vafa 1998a, b), showed that the right-hand side of this equation can be thought as *Gromov–Witten invariants of maps from Riemann surfaces with boundary to the disc.*

Katz and Liu (2001) and Li and Song (2001), with some assumptions, verified that this is in fact the case with mathematical methods. In particular

$$\Phi(\lambda, t, y) = \sum_d \frac{y^d}{2d\sin(d\lambda/2)} \mathrm{e}^{-dt/2}$$

is the multiple cover formula of the disc (here $t/2$ is the relative homology class of the (upper) hemisphere with orientation represented by y).

On the Chern–Simons side, the computation requires the framing of the knot to be fixed; on the IIA side, the mathematical computations also require the choice of a torus action (on the boundary of the Riemann surface). It has been shown that the ambiguities match. $\diamond$

By a theoretical point of view the geometric set-up allowing us to understand the conifold transition image of the Lagrangian submanifold $\mathcal{C}_{\mathcal{K}}$, when $\mathcal{K}$ is the unknot, can be generalized to every knot or link. In practice the associated Chern–Simons deformation and the corresponding open instantons correction in closed-string theory become very intricate and difficult to compute. In Labastida and Mariño (2001) such a computation is carried out in the highly non-trivial case of torus knots again *showing the conjectured matching of quantities.* The same result is obtained for further knots and links in Ramadevi and Sarkar (2000) and Labastida *et al* (2000).

9.4 Lifting to M-theory

In the previous section we saw that the conifold transition machinery is a nice geometrical setting for the large N open/closed string duality conjectured in physics. Here we describe a geometrical construction which gives striking evidence for the Gopakumar–Vafa conjecture and reduce to the conifold geometry to a 'dimensional reduction'. The main references for this construction are Atiyah *et al* (2001) and the more extensive Atiyah and Witten (2001).

The geometric construction is suggested by the physical 'lift' of IIA theories with branes (resp. fluxes), to $\boldsymbol{M}$-theory. In our situation, $\boldsymbol{M}$-theory is then compactified on seven-dimensional, singular, spaces M_{-r}, M_r with special (G_2) holonomy:

$$\begin{array}{ccc} M_{-r} & \longleftarrow\!\!-\,-\,-\!\!\longrightarrow & M_r \\ \downarrow & & \downarrow \\ \mathbb{R}^4 \times S^2 & < -\mathit{conifold}- > & S^3 \times \mathbb{R}^3. \end{array}$$

The vertical maps are essentially Hopf fibrations, the singularities on M_{-r} and M_{+r} are related to the presence of branes (resp. fluxes) and the special holonomy is needed to preserve the $N = 1$ supersymmetry condition. The conifold transition is lifted to a map between seven-dimensional manifold (the '$\boldsymbol{M}$-theory flop') . The physics statement in Atiyah *et al* (2001) and Atiyah and Witten (2001) is that the theory does not go through a singularity under the $\boldsymbol{M}$-theory flop: this implies the Gopakumar–Vafa conjecture for the conifold transition.

In the following section we discuss Riemannian holonomy groups; next we introduce the geometrical construction of the lift for $N = 1$ branes. We will check later its physical consistence with the $\boldsymbol{M}$-theory lift of IIA with branes. Some basics properties of such lifts are stated in section 9.4.3.

9.4.1 Riemannian Holonomy, G_2 manifolds and Calabi–Yau, revisited

The purpose of this section is to fix some notation and basic properties; details and proofs can be found, for example, in Joyce (2000).

Let ∇ be the Levi-Civita connection on the tangent bundle TM of a Riemannian manifold (M, g) and let $p \in M$:

Definition 9.4.1. *The group* $\mathrm{Hol}_p(g)$

$$\mathrm{Hol}_p(g) := \mathrm{Hol}_\nabla(p) \tag{9.95}$$

is the Riemannian holonomy group of g at $p \in M$; $Hol_\nabla(p)$ was defined in (9.26).

It can be seen that when M is connected the holonomy group $\mathrm{Hol}(g)$ is a subgroup of $O(\dim M)$, fixed up to conjugation. If M is orientable then $\mathrm{Hol}(g) \subset SO(\dim M)$. If (M, g, J) is a Kähler manifold of dimension $2m$, then $\mathrm{Hol}(g) \subset U(m)$.

Theorem 9.4.2. *A compact Kähler manifold (M, g, J) of complex dimension $m \geq 3$ is a Calabi–Yau variety if and only if* $\mathrm{Hol}(g) = SU(m)$ *(for a proof see Joyce (2000)).*

In particular such a (M, g, J) is always projective algebraic. The following definition, often used in the physics literature, is then equivalent for $m \geq 3$ to the one given in section 9.1.1:

Definition 9.4.3 (Calabi–Yau, revisited). *A compact Calabi–Yau manifold is a compact Kähler manifold of dimension $2m$, $m \geq 2$, and* $\mathrm{Hol}(g) = SU(m)$.

From the point of view of physics it is the condition $\mathrm{Hol}(g) \subseteqq SU(m)$ which is relevant, as it preserves the required supersymmetry. On a seven-dimensional manifold, the necessary condition becomes $\mathrm{Hol}(g) = G_2$, where G_2 is as follows. below:

Definition 9.4.4. *Let $(x_1, \ldots, x_7)$ be coordinates on $\mathbb{R}^7$ and set*

$$\mathrm{d}\mathbf{x}_{i_1 \ldots i_r} = \mathrm{d}x_{i_1} \wedge \ldots \wedge \mathrm{d}x_{i_r}.$$

G_2 is the Lie subgroup of $GL(7, \mathbb{R})$ preserving the 3-form

$$\varphi_0 := \mathrm{d}\mathbf{x}_{123} + \mathrm{d}\mathbf{x}_{145} + \mathrm{d}\mathbf{x}_{167} + \mathrm{d}\mathbf{x}_{246} - \mathrm{d}\mathbf{x}_{257} - \mathrm{d}\mathbf{x}_{347} - \mathrm{d}\mathbf{x}_{356}.$$

Proposition 9.4.5. *The following hold:*

(i) G_2 *fixes the 4-form* $*\varphi_0$ *($*$ is the Hodge star), the Euclidean metric* $g_0 := \sum_{i=1}^{7} \mathrm{d}x_i^2$ *and the orientation on* $\mathbb{R}^7$. *In particular,* $G_2 \subset SO(7)$.
(ii) G_2 *is compact, connected, simply connected, semisimple.*
(iii) $\dim G_2 = 14$.

Definition 9.4.6. *Let* M *be an oriented manifold with* $\dim M = 7$. *A 3-form* $\varphi_p \in \Lambda^3 T_p^* M$ *is positive at* p *if there exists an oriented isomorphism* $T_p^* M \cong \mathbb{R}^7$ *identifying* φ_p *with* φ_0. *Set*

$$\Lambda_+^3 T_p^* M := \{\varphi_p \in \Lambda^3 T_p^* M, \textit{ such that } \varphi_p \textit{ is positive }\}.$$

A 3-form φ *on* M *is positive if* $\varphi|_p$ *is positive for every point* $p \in M$*; set*

$$\Omega_+^3(M) := \{\varphi \textit{ such that } \varphi_p \in \Lambda_+^3 T_p^* M, \qquad \forall p \in M\}.$$

Note that, by definition $\Lambda_+^3 T_p^* M \cong GL_+(7, \mathbb{R})/G_2$. A dimensional computation implies immediately that it is a non-empty open subset of $\Lambda^3 T_p^* M$. Then a positive 3-form on M is a global section of the open subbundle $\Omega_+^3 M$. Fix a positive 3-form φ on a Riemannian 7-manifold (M, g). We will write

$$\mathrm{Hol}(g) \subseteq_\varphi G_2$$

when, for any $p \in M$, we get

$$\Phi_p \circ (\mathrm{Hol}_p(g)) \circ \Phi_p^{-1} \subseteq G_2$$

where Φ_p is an oriented isomorphism $T_p^* M \cong \mathbb{R}^7$ representative of the class in $GL_+(7, \mathbb{R})/G_2$ associated with $\varphi|_p$ via the isomorphism (9.4.1). Since G_2 is invariant under conjugation, for any two positive form φ, ψ

$$\mathrm{Hol}(g) \subseteq_\varphi G_2 \Longleftrightarrow \mathrm{Hol}(g) \subseteq_\psi G_2.$$

Without loss of generality we then write $\mathrm{Hol}(g) \subseteq G_2$.

Definition 9.4.7. (M, g) *has a* G_2 *holonomy metric if* $\mathrm{Hol}(g) = G_2$.

The following properties assure that supersymmetry is preserved:

Proposition 9.4.8. *Let* (M, g) *be a Riemannian 7-manifold with* G_2 *holonomy metric. Then*

(i) g *is Ricci flat,*
(ii) M *is an orientable spin manifold and*
(iii) (M, g) *has a non-zero covariant spinor.*

(See, for example, Joyce [2000] for a proof of these statements.)

The existence of manifolds with G_2 holonomy metric was firstly studied in Bryant (1987) and then solved in Bryant and Salamon (1989) and in Gibbons *et al* (1990) for non-compact manifolds. Compact manifolds with G_2 holonomy metric were then constructed in Joyce (1996). See also chapter 11 in Joyce (2000).

9.4.2 The geometry

Lemma 9.4.9. *Fix r in $\mathbb{R}_{>0}$, $\mathbb{C}^4$ with coordinates (z_1, z_2, z_3, z_4) and set*

$$\begin{aligned} M_r : &= \{\mathbf{z} \in \mathbb{C}^4 : |z_1|^2 + |z_2|^2 - |z_3|^2 - |z_4|^2 = r\} \\ M_{-r} : &= \{\mathbf{z} \in \mathbb{C}^4 : |z_1|^2 + |z_2|^2 - |z_3|^2 - |z_4|^2 = -r\}. \end{aligned}$$

Then, topologically,

$$\begin{aligned} M_r &\cong S^3 \times \mathbb{C}^2_{(z_3,z_4)} \cong S^3 \times \mathbb{R}^4 \\ M_{-r} &\cong \mathbb{C}^2_{(z_1,z_2)} \times S^3 \cong \mathbb{R}^4 \times S^3. \end{aligned}$$

The proof of this lemma is presented after the proof of the following proposition.

Proposition 9.4.10. *There exists the following geometric lift of the conifold transition:*

$$\begin{array}{ccccccc} M_{-r} & \cong & \mathbb{R}^4 \times S^3 & \leftarrow \cdots \rightarrow & S^3 \times \mathbb{R}^4 & \cong & M_r \\ & & {\scriptstyle h_-} \downarrow & & \downarrow {\scriptstyle h_+} & & \\ & & \mathbb{R}^4 \times S^2 & < -\text{conifold}- > & S^3 \times \mathbb{R}^3 & & \end{array} \tag{9.96}$$

where

(i) h_- is the identity on the first factor and the Hopf fibration on S^3 and
(ii) h_+ is the identity on the first factor and the non-differentiable extension to $\mathbb{R}^3$ of the Hopf fibration on S^3.

Furthermore $\mathbb{R}^4 \times S^3$ admits a G_2 holonomy metric.

Proof of proposition 9.4.10 The key geometric observation of the following argument is that M_{-r} and M_r are resolutions of real cones over $S^3 \times S^3$, while $\mathbb{R}^3 \times S^3$ and $S^2 \times \mathbb{R}^4$ are resolutions of a real cone over $S^2 \times S^3$. Furthermore the Hopf fibration maps $S^3 \to S^2$.

Clemens theorem (9.1.5) describes the conifold transition as a surgery between topological spaces with the same boundary. This surgery is expressed by the morphism α, which is the identity on $S^3 \times S^2$ (see (9.17)):

$$\alpha : (\mathbb{R}^4 \setminus \{\mathbf{0}\}) \times S^2 \cong S^3 \times (\mathbb{R}^3 \setminus \{\mathbf{0}\}).$$

Since,

$$\begin{aligned} (\mathbb{R}^4 \setminus \{\mathbf{0}\}) \times S^2 &\cong \mathbb{R}_{>0} \times S^3 \times S^2 \\ S^3 \times (\mathbb{R}^3 \setminus \{\mathbf{0}\}) &\cong S^3 \times S^2 \times \mathbb{R}_{>0} \end{aligned}$$

we can re-write α as

$$\begin{aligned} \alpha : \mathbb{R}_{>0} \times S^3 \times S^2 &\longrightarrow S^3 \times S^2 \times \mathbb{R}_{>0} \\ (\rho, \mathbf{u}, \mathbf{v}) &\longmapsto (\mathbf{u}, \mathbf{v}, \rho). \end{aligned} \tag{9.97}$$

As in the previous lemma, we embed $S^3 \subset \mathbb{C}^2_{(z_i, z_{i+1})}$ and consider the compatible *Hopf fibration:*

$$\begin{aligned} h : S^3 &\longrightarrow \mathbb{P}^1_{\mathbb{C}} \cong S^2 \\ (z_i, z_{i+1}) &\longmapsto [z_i, z_{i+1}] = [\lambda z_i, \lambda z_{i+1}] \qquad \lambda \in \mathbb{C}^*. \end{aligned} \tag{9.98}$$

Then the following diagram:

$$\begin{array}{ccc} \mathbb{R}_{>0} \times S^3 \times S^3 & \xrightarrow{\widetilde{\alpha}} & S^3 \times S^3 \times \mathbb{R}_{>0} \\ {\scriptstyle h_3} \downarrow & & \downarrow {\scriptstyle h_2} \\ \mathbb{R}_{>0} \times S^3 \times S^2 & \xrightarrow{\alpha} & S^3 \times S^2 \times \mathbb{R}_{>0} \end{array} \tag{9.99}$$

commutes, where

$$\begin{aligned} h_3 &:= \mathrm{Id}_{\mathbb{R}_{>0}} \times \mathrm{Id}_{S^3} \times h \\ h_2 &:= \mathrm{Id}_{S^3} \times h \times \mathrm{Id}_{\mathbb{R}_{>0}} \\ \widetilde{\alpha}(\rho, \mathbf{u}, \mathbf{u}') &:= (\mathbf{u}, \mathbf{u}', \rho). \end{aligned}$$

Note that, while h_3 can be smoothly extended to a fibration,

$$h_- := \mathrm{Id}_{\mathbb{R}^4} \times h : \mathbb{R}^4 \times S^3 \longrightarrow \mathbb{R}^4 \times S^2$$

this is not true for h_2. There is, however, a topological extension of h_+ h_2. The extensions h_- and h_+ then give the diagram (9.96) in the statement.

Bryant and Salamon (1989) and Gibbons *et al* (1990) explicitly describe a G_2 holonomy metric on $M := S^3 \times \mathbb{R}^4$.

The metric in Gibbons *et al* (1990) is a smooth extension of the metric on the cone over $S^3 \times S^3$. Bryant and Salamon (1989) consider $SU(2) \cong S^3$ and the quaternions $\mathbb{H} \cong \mathbb{R}^4$ as a cone over $SU(2)$. Then $S^3 \times \mathbb{R}^4 \cong (SU(2) \times SU(2) \times \mathbb{H})/SU(2)$, with $SU(2)$ acting on the right, is a rank-four vector bundle on $SU(2)$. With this latter representation is evident that there are two other resolutions of the cone over $S^3 \times S^3$:

$$(\mathbb{H} \times SU(2) \times SU(2))/SU(2) \cong \mathbb{R}^4 \times S^3, \; (SU(2) \times \mathbb{H} \times SU(2))/SU(2).$$

The third manifold fibres, via the Hopf fibration, to the 'flopped' local Calabi–Yau Y_+ of the resolved conifold Y (see (9.10)); we have then the third branch of figure 9.3 (see also Manzoni 1842). $\diamond$

Proof of lemma 9.4.9. Let (z_1, z_2, z_3, z_4) be coordinates in $\mathbb{C}^4$; for every r, positive real number set

$$M_r := \{\mathbf{z} \in \mathbb{C}^4 : |z_1|^2 + |z_2|^2 - |z_3|^2 - |z_4|^2 = r\}.$$

Then,

$$\begin{array}{cccc} \phi_+ : & M_r & \longrightarrow & S^3 \times \mathbb{C}^2 \\ & (z_1, z_2, z_3, z_4) & \longmapsto & \left(\dfrac{z_1}{\rho_+}, \dfrac{z_2}{\rho_+}, z_3 \cdot \rho_+, z_4 \cdot \rho_+ \right) \end{array}$$

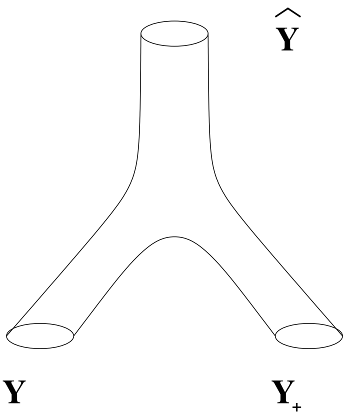

Figure 9.3. The three branches of the moduli. (*'I rami del lago di Como...'*.)

is a isomorphism, where $\rho_+ := \sqrt{|z_1|^2 + |z_2|^2} = \sqrt{r + |z_3|^2 + |z_4|^2}$.

Similarly for M_{-r}. ◇

9.4.3 Branes and *M*-theory lifts

It remains to understand how this geometrical construction reflects a consistent correspondence among physical theories. On the two sides of the conifold transition we have the theories IIA with branes, and IIA with fluxes, compactified on $\widehat{Y}$ and Y respectively. The equivalence of the theory is claimed by the Gopakumar conjecture 9.3.1.

The crucial point is that IIA string theory may itself be regarded as a dimensional reduction of a $\mathcal{N} = 1$ supersymmetric Lorentz invariant theory in 11 dimensions: M-theory (see Sen [1997, section 7], for a quick review and references cited there for details on the argument). $\boldsymbol{M}$-theory was proposed originally in Townsend (1995) and Witten (1995), who observed that the low-energy limit of a type IIA string theory, i.e. a type IIA supergravity theory, can be obtained by a 'Kaluza–Klein' dimensional reduction of a $\mathcal{N} = 1$ supersymmetric gravity theory in 11 dimensions. The reduction is along an S^1, called the 11th *circle*.

When $\boldsymbol{M}$-theory and IIA are 'compactified' on manifolds M and Y respectively, the 'Kaluza–Klein' dimensional reduction induces an S^1 fibration $h : M \to Y$.

If N G-branes are 'wrapped' on a submanifold $L \subset Y$, M is singular along $h^{-1}(L)$; the type of singularity is determined by G (see appendix 9.5) and h is a singular Hopf fibration. For a survey on this topic see, for example, Johnson (1998, 2000). In order to preserve the $\mathcal{N} = 1$ supersymmetry of the theory, M must be a manifold with G_2 holonomy.

9.4.4 $\boldsymbol{M}$-theory lift and $\boldsymbol{M}$-theory flop

Physics statement 9.4.11 (Acharya (1999), Atiyah *et al* (2001), Atiyah–Witten (2001)). *The following diagram is physically consistent with the $\boldsymbol{M}$-theory lift of N D-branes, (i.e. gauge group $SU(N)$):*

$$\begin{array}{ccc} M_{-r} & & M_r \\ \pi_- \downarrow & & \downarrow \pi_+ \\ X_- & & X_+ \\ h_-^{(N)} \downarrow & & \downarrow h_+^{(N)} \\ \mathbb{R}^4 \times S^2 & \leftarrow \text{conifold} \rightarrow & S^3 \times \mathbb{R}^3 . \end{array} \tag{9.100}$$

In particular,

(i) X_- and X_+ are G_2 holonomy spaces.
(ii) The surjections $h_-^{(N)}, h_+^{(N)}$ give rise to N R–R fluxes and N D-branes, respectively, for the type IIA string theories obtained by dimensional reduction on the two sides of the conifold transition.
(iii) $(S^3, 0) \subset X_+$ is a locus of A_{N-1} singularities.

Finally, $\boldsymbol{M}$-theory compactified on X_- is equivalent to $\boldsymbol{M}$-theory on X_+. Note that the physics description is valid only for large N.

At the time this lecture was given the work by Acharya (1999), and Atiyah *et al* (2001) was in print, while the main results of Atiyah and Witten (2001) had just been recently announced. Atiyah and Witten (2001) 'argue that there is a moduli space of theories of complex dimension one that interpolates smoothly, without phase transition, between the three classical spacetimes' obtained by compactification on the three lifts described at the end of proof of 9.4.10 and in figure 9.3.

Sketch of the proof: The geometric lift (9.96) gives an $\boldsymbol{M}$-theory lift of IIA string theories when $N = 1$. The singularity of the map h_+ denotes the presence of branes.

To get $\boldsymbol{M}$-theory lift with N D-branes wrapped on $S^3 \times \{0\} \subset S^3 \times \mathbb{R}^3$ we need to introduce corresponding singularities on M_r (see section 9.4.3). We do so by defining a suitable action of the group of Nth roots of unity on $\mathbb{C}^4$: the induced action on M_{-r} will give N units of R–R fluxes on $\mathbb{R}^4 \times S^2$.

Let $\Gamma_N := \mathbb{Z}/N\mathbb{Z}$ act on $\mathbb{C}^4$ as

$$\begin{aligned} \Gamma \times \mathbb{C}^4 &\longrightarrow \mathbb{C}^4 \\ (n, \mathbf{z}) &\longmapsto (z_1, z_2, \xi_n z_3, \xi_n z_4) \end{aligned} \tag{9.101}$$

where $\xi_n := \exp(2\pi \mathrm{i} n/N)$. The complex plane $F =: \{z_3 = z_4 = 0\}$ is the fixed locus of Γ. Recall that $M_{-r} \cong \mathbb{C}^2{}_{(z_1,z_2)} \times S^3$ and $M_r \cong S^3 \times \mathbb{C}^2{}_{(z_3,z_4)}$. Then,

$$F \cap M_{-r} = \emptyset \qquad F \cap M_r = S^3 \times \{\mathbf{0}\}.$$

The quotient

$$M_{-r} \cong \mathbb{C}^2{}_{(z_1,z_2)} \times S^3 \longrightarrow M_{-r}/\Gamma \cong \mathbb{R}^4 \times (S^3/\Gamma) := X_-$$

is smooth; (S^3/Γ) is called a *lens space* and denoted by $L(N)$. Furthermore, since the Γ-action is restricted to the fibre of the Hopf fibration, the map h_- in (9.96) can be factorized through the canonical projection π_- as follows

$$\begin{array}{ccc} M_{-r} & \xrightarrow{h_-} & \mathbb{R}^4 \times S^2 \\ \pi_- \searrow & & \nearrow_{h_-^{(N)}} \\ & X_-. & \end{array}$$

However, the quotient

$$M_r \cong S^3 \times \mathbb{C}^2{}_{(z_3,z_4)} \longrightarrow M_r/\Gamma \cong S^3 \times (\mathbb{R}^4/\Gamma) := X_+$$

contains a S^3 *of singular points*. Furthermore, since the Γ-action is restricted to the fibre of the Hopf fibration, the map h_+ in (9.96) can topologically be factorized through the canonical projection as follows

$$\begin{array}{ccc} M_{r} & \xrightarrow{h_+} & S^3 \times \mathbb{R}^3 \\ \pi_+ \searrow & & \nearrow_{h_+^{(N)}} \\ & X_+. & \end{array}$$

$\mathbb{R}^4/\Gamma$ is a A_{N-1} singularity, with gauge group $SU(N)$ (see appendix 9.5). In fact with the change of coordinates $w_3 = z_3$, $w_4 = \sqrt{-1}\cdot \bar{z}_4$, the action becomes $(w_3, w_4) \to (\xi w_3, \xi^{-1} w_4)$ as described in appendix 9.5. This is the geometric incarnation of the M-theory lift with $SU(N)$ branes wrapped on S^3.

Furthermore the non-singular $\mathbb{Z}_N$ quotient (on the left of diagram (9.100)) gives rise to N units of R–R flux. In fact, if $V(-r)$ is the volume of $S^3 \times \{\mathbf{0}\}$, then $\mathrm{vol}(S^2) = \mathrm{vol}(S^3/\Gamma) = V(-r)/N$.

Recall that there exists a G_2 holonomy metric (Bryant and Salamon (1989), Gibbons *et al* (1990)) on $M := S^3 \times \mathbb{R}^4$. There is a precise description of the isometry group on M and the action of Γ is included in this subgroup. Hence the quotients X_- and X_+ are also G_2 holonomy spaces.

Atiyah and Witten (2001), following Atiyah *et al* (2001) show that the M-theories compactified on X_- and X_+ are equivalent. Thus, there is no 'phase' transition, exactly as when IIA is compactified on Calabi–Yau varieties related by a 'flop' (see Witten 1993). Hence the name of M-theory flop.

It is worth pointing out that the equivalence of the theory and the relations between the physical parameters derived in Atiyah *et al* (2001) *is only valid for large N*. The equivalence of the theories also implies the relations between the Kähler modulus of Y and the parameters on the Chern–Simons theory conjectured by Gopakumar and Vafa (Atiyah *et al* 2001).

However, the asympotics of the G_2 metric is not what it would be expected from the IIA situation; based on this observation Atiyah *et al* (2001) conjectured the existence of a deformation of the G_2 metric with such properties. This was later shown in Brandhube *et al* 2001. $\diamond$

9.5 Appendix: Some notation on singularities and their resolutions

Here we adopt the same notation and terminology introduced in Reid (1980, 1983, 1987a).

Definition 9.5.1. *A Weil divisor D on a a complex, normal and quasiprojective variety is $\mathbb{Q}$-Cartier, if for some $r \in \mathbb{Z}$, rD is a Cartier divisor (i.e. $D \in Pic(\overline{Y}) \otimes \mathbb{Q}$).*

If $\overline{Y}$ is smooth then any Weil divisor is Cartier.

Definition 9.5.2. *A $\overline{Y}$ be a complex, normal and quasiprojective variety is $\mathbb{Q}$-factorial if any Weil divisor is $\mathbb{Q}$-Cartier.*

Definition 9.5.3. *Let $\overline{Y}$ be a complex, normal and quasiprojective variety and $K_{\overline{Y}}$ be its canonical divisor which is, in general, a Weil divisor. $\overline{Y}$ has canonical (terminal) singularities if*

(i) $K_{\overline{Y}}$ is $\mathbb{Q}$-Cartier,
(ii) given a smooth resolution $f : Y \longrightarrow \overline{Y}$ then

$$rK_Y \equiv f^* K_{\overline{Y}} + \sum_i a_i E_i$$

where $\equiv$ means 'linearly equivalent', E_i are all the exceptional divisors of f and $a_i \geq 0$ (respectively $a_i > 0$).

The smallest integer r for which such conditions hold is called the *(global) index of $\overline{Y}$* and the smallest r' for which $r' K_{\overline{Y}}$ is Cartier in a neighbourhood of $P \in \overline{Y}$ is called the *index of the singularity P*.

The divisor $\Delta := \sum_i a_i E_i$ is called the *discrepancy of the resolution f*. If $\Delta \equiv 0$ then f is called a *crepant resolution of $\overline{Y}$*.

We are interested in transitions of Calabi–Yau manifolds: in particular, if at a point in the complex moduli $\overline{Y}$ is singular and $K_{\overline{Y}} \equiv 0$, its birational resolution should be crepant to preserve the Calabi–Yau condition on the canonical bundle.

Definition 9.5.4. *(see, for example, Clemens* et al *1988) By $NE(Y) \subset \mathbb{R}^\ell$ we denote the cone generated (over $\mathbb{R}_{\geq 0}$) by the effective cycles of (complex) dimension 1, mod. numerical equivalence. $\overline{NE(Y)}$ is the closure of $NE(Y) \subset \mathbb{R}^\ell$ in the finite dimensional real vector space $\mathbb{R}^\ell$ of all cycles of complex dimension 1, mod. numerical equivalence.*

Note that $\ell = rk(Pic(Y))$, and in the cases of Calabi–Yau manifolds, $\ell = b_2(Y)$, the second Betti number of Y.

Definition 9.5.5. *A birational contraction $f : Y \to \overline{Y}$ is called primitive extremal if the numerical class of a fibre of f is on a ray of the Mori cone $NE(Y)$.*

Examples

The surface case

Let X be a surface. It can be proved that a point $P \in X$ is a terminal singularity if and only if it is non-singular. Moreover, the canonical (non-terminal) singular points are given by the *Du Val singularities* (DV points) which are classified as follows in terms of their local equations.

$$\begin{aligned}
&A_n : x^2 + y^2 + z^{n+1} = 0 && n \geq 1\\
&D_n : x^2 + y^2 z + z^{n-1} = 0 && n \geq 4\\
&E_6 : x^2 + y^3 + z^4 = 0\\
&E_7 : x^2 + y^3 + yz^3 = 0\\
&E_8 : x^2 + y^3 + z^5 = 0.
\end{aligned}$$

In particular each of them admits a crepant resolution whose exceptional locus is composed by a set of (-2) curves (i.e. rational curves admitting self-intersection index -2) whose configuration are dually represented by the homonymous Dynkin diagrams: these are particular examples of Hirzebruch–Jung strings (see Peters *et al* 1984, chapters I and III).

Note that an ordinary double point is represented by A_1 and admits a crepant resolution whose exceptional locus is given by a unique (-2) curve. This equation is generalized to the threefold case in equation (9.1.4).

Each of these singularities can be described as a quotient of $\mathbb{C}^2$ by a discrete subgroup $\Gamma \subset SL(2)$. For A_n, Γ is the cyclic group of order $n+1$ generated by a primitive nth root of unity ξ; the action on $\mathbb{C}^2$ sends $(w_1, w_2) \to (\xi w_1, \xi^{-1} w_2)$ (see Slodowy 1990).

The threefold case

Let X be a threefold and $P \in X$ be a canonical singular point of index r. A first important fact is that there exists a finite Galois covering $Y \longrightarrow X$ with group $\mathbb{Z}/r$ which is étale in codimension 1 and such that Y is locally canonical of index 1 (see Reid 1980, corollary 1.9).

Definition 9.5.6. *$P \in X$ is a compound Du Val singularity (cDV point) if the restriction to a surface section is a Du-Val surface singularity.*

The advantage of these kind of singularities is that they admit a simultaneous small resolution, as studied by several authors (see, e.g., Reid (1983), Pinkham

(1983), Morrison (1985), Friedman (1986)). The idea is that of thinking an analytic neighbourhood of an isolated cDV point like the total space of a one-parameter family of deformations of the section over which we get a DV point. The total space of the induced one-parameter family of deformations of a given resolution of such a DV point is then a small resolution of the starting cDV point. One can now apply the theory of simultaneous resolutions of DV points on surfaces (Brieskorn 1966, 1968, Tyurina 1970). The main theorem in Reid (1983) states that

(i) $P \in X$ is a terminal singularity of index r if and only if the local r-fold cyclic covering $Y \longrightarrow X$ has only isolated *compound Du Val singularities*; and
(ii) if X admits at most canonical singularities then there exists a crepant partial resolution $S \longrightarrow X$ such that S admits at most isolated terminal singularities.

These results allows to reduce the problem of resolving canonical singularities to that of resolving cDV points, up to partial resolutions and finite coverings.

9.6 Appendix: More on the Greene–Plesser construction

Here we will quickly sketch an example supporting the Greene–Plesser construction explained in Candelas *et al* (1994) and in Morrison (1999).

Let $\overline{Y}_1$ be the degree 8 weighted hypersurface of $\mathbb{P}(1, 1, 2, 2, 2)$ and Y_1 be the desingularization induced by blowing up the singular locus of $\mathbb{P}(1, 1, 2, 2, 2)$. Here ϕ is a primitive contraction of type III and the transition can be completed by considering the embedding of $\mathbb{P}(1, 1, 2, 2, 2)$ in $\mathbb{P}^5$ by means of the linear system $\mathcal{O}(2)$. The image of $\mathbb{P}(1, 1, 2, 2, 2)$ is a rank 3 hyperquadric of $\mathbb{P}^5$. Hence the image of $\overline{Y}_1$ is the complete intersection of this hyperquadric with the generic quartic hypersurface of $\mathbb{P}^5$. By smoothing the hyperquadric we get $\widehat{Y}_1$. Following the idea of Greene–Plesser (1990) the mirror partners may be found by taking the quotient with the subgroups of automorphisms preserving the holomorphic 3-form. Since the hypersurfaces cohomology can be completely described by means of Poincaré residues (see Griffiths 1969) these subgroups are respectively given by

$$G := \{(a_0, \ldots, a_4) \in (\mathbb{Z}_8)^2 \times (\mathbb{Z}_4)^3 : \sum a_i \equiv 0(8)\}$$
$$H := \{(b_0, \ldots, b_5) \in (\mathbb{Z}_4)^2 \times (\mathbb{Z}_2)^4 : b_0 + b_1 \equiv b_2 + \cdots + b_5 \equiv 0(4)\}.$$

We denote by a_i, b_j the least non-negative integers representing the homonymous class in $\mathbb{Z}_n$. Hence the mirror partner $\widehat{Y}_2$ of $\widehat{Y}_1$ may be obtained by a H-invariant complete intersection of bidegree $(2, 4)$ in $\mathbb{P}^5$ via the desingularization of the quotient $\mathbb{P}^5/H$ where H acts on $\mathbb{P}^5$ as follows.

$$\begin{array}{ccc} (H/\Delta_H) \times \mathbb{P}^5 & \longrightarrow & \mathbb{P}^5 \\ (\mathbf{b}, \mathbf{x}) & \longmapsto & (\beta_j x_j) \end{array}$$

where

$$\beta := \begin{cases} \exp\left(\frac{b_j \pi \mathrm{i}}{4}\right) & \text{for } j = 0, 1 \\ \pm 1 & \text{otherwise} \end{cases}$$

and Δ_H is the subgroup of H giving a trivial action on $\mathbb{P}^5$, i.e.

$$\Delta_H := \{(0, \ldots, 0), (2, 2, 1, \ldots, 1)\}.$$

However, the mirror partner Y_2 of Y_1 may be obtained by a G-invariant hypersurface of degree 8 in $\mathbb{P}(1, 1, 2, 2, 2)$ via the desingularization of the quotient $\mathbb{P}(1, 1, 2, 2, 2)/G$ where G acts on $\mathbb{P}(1, 1, 2, 2, 2)$ as follows.

$$\begin{array}{ccc} G/\Delta_G \times \mathbb{P}(1, 1, 2, 2, 2) & \longrightarrow & \mathbb{P}(1, 1, 2, 2, 2) \\ (\mathbf{a}, \mathbf{x}) & \longmapsto & (\alpha_j x_j) \end{array}$$

where

$$\alpha_j := \begin{cases} \exp\left(\frac{a_j \pi \mathrm{i}}{8}\right) & \text{for } j = 0, 1 \\ \exp\left(\frac{a_j \pi \mathrm{i}}{4}\right) & \text{otherwise} \end{cases}$$

and Δ_G is the diagonal subgroup of G which is

$$\Delta_G := \{(a, \ldots, a) : 0 \leq a \leq 3\}.$$

It can be checked that there is a birational equivalence between $\widehat{Y_2}$ and Y_2 representing a mirror partner of our transition .

9.7 Appendix: More on transitions in superstring theory

Strominger gave a physical explanation of how resolving the conifold singularities of the moduli space of classical string vacua by means of massless Raymond–Ramond (R–R) black holes . More precisely the possible compactifications of a ten-dimensional IIB string theory to four dimensions on a Calabi–Yau manifold Y may be parametrized by the choice of the complex structure characterizing Y. Such a choice may be described by the periods of a holomorphic 3-form Ω over a suitable symplectic basis of $H_3(Y, \mathbb{Q})$ (see de Wit *et al* [1985] and Strominger [1990] for detailed notation in a $N = 2$, four-dimensional supergravity theory and in special geometry) which can be considered as projective coordinates of the moduli space $\mathcal{M}(Y)$ of complex structures. The complex codimension 1 locus defined in $\mathcal{M}$ by the vanishing of one of those periods is composed by singular complex structures generically geometrically realized by a conifold. In fact the generic singularity is given by an ordinary double point. Note that the associated vanishing cycle is represented by the 3-cycle of the symplectic basis corresponding to the vanishing period.

Such singularities induce a polydromic behaviour for the components of the self-dual 5-form giving the classical field. Following an analogous construction given in Seiberg and Witten (1994) and applied in the completely different context of $N = 2$ supersymmetric Yang–Mills theory, Strominger resolved this problem by means of a low-energy effective Wilsonian field defined by including the light fields associated with extremal black 3-branes which can wrap around the vanishing 3-cycles and are always contained in a ten-dimensional compactified type IIB theory (see Horowitz and Strominger 1991). These 3-branes represent black holes whose mass is proportional to the volume of the vanishing cycles they wrap around. Hence they are massless at the conifold and by integrating out the smooth so defined Wilsonian field we get exactly the polydromic behaviour of the classical field. This is enough to ensure that the theory may smoothly extend to the conifold.

However, in the case of a ten-dimensional compactified type IIA theory we get a similar picture by taking the periods of a complexified Kähler form $\omega \in H^2(Y, \mathbb{C}) = H^{1,1}(Y)$ over a suitable basis of $H_2(Y, \mathbb{Q})$ as projective coordinates of the moduli space $\mathcal{M}^{'}(Y)$ of all possible Kähler structure on Y (which parametrizes all the possible compactifications of a ten-dimensional IIA string theory to four dimensions on the Calabi–Yau manifold Y). We get now black 2-branes (see Horowitz and Strominger 1991) which can wrap around vanishing 2-cycles and represent massless black holes at conifold. Since, in this case, these massless states are a result of large instanton corrections the resolution of singularities can be obtained by passing to the dual IIB compactification on a mirror model Y° of Y and by proceeding like before.

9.8 Appendix: Principal bundles, connections etc

Here we review some terminology, concepts and properties from differential geometry: for more details, see, for example, Helgason (1978), Poor (1981) and Warner (1983).

Definition 9.8.1. *Let G be a Lie group. A left (right) action of G on a manifold M is a homomorphism (anti-homomorphism) to the group of diffeomorphisms of M:*

$$L(resp.\ R) : G \longrightarrow \mathrm{Diff}(M)$$

In particular for every $\sigma, \tau \in G$ we have $L(\sigma) \circ L(\tau) = L(\sigma\tau)$ (resp. $R(\sigma) \circ R(\tau) = R(\tau\sigma)$).

Definition 9.8.2. *An action is free if* id *is the unique element of G mapping to the identity of* $\mathrm{Diff}(M)$. *Note that if the G-action is free then it is an injection of G into* $\mathrm{Diff}(M)$.

Definition 9.8.3. *A principal G-bundle on a manifold M is a manifold P on which G acts freely on the right together with a smooth, surjective map* $\pi : P \to M$ *such that* *(i) for every point $m \in M$ there is a local trivialization of P i.e. an open neighbourhood $\{U_a\}$ and a local diffeomorphism $\varphi_{U_a} : \pi^{-1}(U_a) \overset{\cong}{\to} U_a \times G$ making the following diagram be commutative*

$$\begin{array}{ccc} \pi^{-1}(U) & \xrightarrow{\varphi_U} & U \times G \\ {\scriptstyle\pi}\downarrow & \overset{pr_1}{\swarrow} & \\ U & & \end{array} \tag{9.102}$$

(ii) π is G-invariant i.e. for every $p \in P$ and every $\sigma \in G$

$$\pi(p\sigma) = \pi(p)$$

where $p\sigma := R(\sigma)p$.

Remark 9.8.4. *For a principal bundle (P, π) the map π is a submersion implying that*

$$\mathcal{V}_p P := \ker(d_p\pi) = T_p\pi^{-1}(\pi(p))$$

for every $p \in \pi^{-1}(\pi(p))$. Set $m := \pi(p) \in M$ and let (U, φ_U) be a local trivialization of P near m. The commutative diagram (9.102) allows us to define a diffeomorphism σ_m^U such that

$$(\sigma_m^U)^{-1} := (\varphi_U^{-1}) \mid_{\{m\}\times G}: G \overset{\cong}{\longrightarrow} \pi^{-1}(m).$$

Its differential gives the isomorphism

$$\mathrm{d}_p\sigma_m^U : T_p\pi^{-1}(m) \overset{\cong}{\longrightarrow} T_{\sigma_m^U(p)}G.$$

However, by differentiating the automorphism r_σ of G, given by right multiplication by $\sigma \in G$, we get the isomorphism

$$\mathrm{d}_{\mathrm{id}}r_\sigma : \mathfrak{g} \cong T_{\mathrm{id}}G \overset{\cong}{\longrightarrow} T_\sigma G$$

where $\mathfrak{g}$ is the Lie algebra associated with G whose elements are all the left invariant vector fields on G. Hence, for every $p \in \pi^{-1}(m)$, we get the isomorphism

$$\mathrm{d}_p(r^{-1}_{\sigma_m^U(p)} \circ \sigma_m^U) : \ker(\mathrm{d}_p\pi) \overset{\cong}{\longrightarrow} \mathfrak{g}.$$

This suffices to conclude that the vertical bundle $\mathcal{V}P$ associated with the principal G-bundle (P, π) is a vector bundle whose standard fibre is the Lie algebra $\mathfrak{g}$ associated with G. In particular near a point $p \in P$ we have the local trivialization $(\pi^{-1}(U), \varphi_{\pi^{-1}(U)})$ where

$$\varphi_{\pi^{-1}(U)} : \mathcal{V}P \mid_{\pi^{-1}(U)} \overset{\cong}{\longrightarrow} \pi^{-1}(U) \times \mathfrak{g}$$

is the diffeomorphism defined by setting

$$\varphi_{\pi^{-1}(U)}(u) := (q, \mathrm{d}_q(r^{-1}_{\sigma^U_{\pi(q)}(q)} \circ \sigma^U_{\pi(q)})(u))$$

for every $q \in \pi^{-1}(U)$ *and* $u \in \mathcal{V}_q P$.

Recall the definition 9.2.2 of a connection on a principal G-bundle (P, π). It is not difficult to show that every principal bundle on a paracompact manifold M admits a connection (see, e.g., Poor 1981, theorems 2.35 and 9.3). Given a connection $\mathcal{H}P \subset TP$ we can uniquely split a vector field $X : P \longrightarrow TP$ into a horizontal part $\mathcal{H}X : P \longrightarrow \mathcal{H}P$ and a vertical part $\mathcal{V}X : P \longrightarrow \mathcal{V}P$ such that, for every $p \in P$,

$$X_p = \mathcal{H}_p X + \mathcal{V}_p X. \tag{9.103}$$

Recalling definition 9.2.3 let $A \in \Omega^1(P, \mathfrak{g})$ be the $\mathfrak{g}$-valued 1-form associated with the connection $\mathcal{H}P$ and $\Omega \in \Omega^2(P, \mathfrak{g})$ be its curvature $\mathfrak{g}$-valued 2-form. These forms are each other related by the *structure equation*

$$\Omega(X, Y) = \mathrm{d}A(X, Y) + [AX, AY]$$

for every vector fields X, Y on P. We can rewrite it in the following shorter shape

$$\Omega = \mathrm{d}A + \tfrac{1}{2}[A, A] \tag{9.104}$$

by setting $[A, A](X, Y) := [AX, AY] - [AY, AX]$.

Let l_σ be the automorphism of G given by left multiplication by $\sigma \in G$. The dual vector space $\mathfrak{g}^*$ of the Lie algebra $\mathfrak{g}$ can be canonically identified with the vector space of all *left invariant* 1*-forms on* G since all such forms assume constant values on left invariant vector fields. The composition

$$a_\sigma := l_\sigma \circ r_{\sigma^{-1}} : G \longrightarrow G$$

is an automorphism of G fixing id $\in G$. Therefore its differential

$$Ad_\sigma := \mathrm{d}_{\mathrm{id}} a_\sigma \tag{9.105}$$

may be thought like an automorphism of $\mathfrak{g} \cong T_{\mathrm{id}}G$ and its codifferential $\delta_{\mathrm{id}} a_\sigma$ like an automorphism of $\mathfrak{g}^*$.

Proposition 9.8.5. *Let us consider* $\theta \in \mathfrak{g}^*$ *and* $X, Y \in \mathfrak{g}$. *Then for every* $\sigma \in G$

$$(\delta r_\sigma)\theta X = (\theta \circ Ad_\sigma)X \tag{9.106}$$

and they satisfy the Maurer–Cartan equation[1]

$$\mathrm{d}\theta(X, Y) = -\theta[X, Y]. \tag{9.107}$$

[1] For this reason left invariant 1-forms are also called *Maurer–Cartan forms*.

Proof. To prove (9.106) note that, for every $\tau \in G$, left invariance of θ gives

$$\theta_{\tau\sigma} = (\delta_{\tau\sigma} l_{\sigma^{-1}})\theta_{\sigma^{-1}\tau\sigma}$$

which implies

$$(\delta_\tau r_\sigma)\theta_{\tau\sigma} = (\delta_\tau r_\sigma \circ \delta_{\tau\sigma} l_{\sigma^{-1}})\theta_{\sigma^{-1}\tau\sigma} = (\delta_\tau a_{\sigma^{-1}})\theta_{\sigma^{-1}\tau\sigma} = \theta_{\sigma^{-1}\tau\sigma} \circ \mathrm{d}_\tau a_\sigma.$$

To restrict this relation to a left invariant vector field $X \in \mathfrak{g}$ means to choose $\tau = \mathrm{id}$ and so to obtain just (9.106). For (9.107) let us observe that, almost by definition,

$$\mathrm{d}\theta(X, Y) = X\theta Y - Y\theta X - \theta[X, Y].$$

Since $X, Y \in \mathfrak{g}$ left invariance of θ implies that both θY and θX are constant functions. This suffices to end up the proof. $\square$

Given a point $p \in P$ let us now consider the codifferential

$$\delta\lambda_p : T^*P \longrightarrow T^*G$$

and let A be the connection form of $\mathcal{H}P$. We can then define the $\mathfrak{g}$-valued 1-form $(\delta\lambda)A \in \Omega^1(G, \mathfrak{g})$ by setting

$$((\delta\lambda)A)_\sigma := (\delta_\sigma \lambda_p)A_{p\sigma} \tag{9.108}$$

for every $\sigma \in G$. This definition is not dependent on the choice of $p \in P$ since by (9.23) we have for every $v \in T_\sigma G$

$$(\delta_\sigma \lambda_p)A_{p\sigma}(v) = A_{p\sigma}((\mathrm{d}_\sigma \lambda_p)v) = (\mathrm{d}_{\mathrm{id}}\lambda_{p\sigma})^{-1}(\mathcal{V}_{p\sigma}(\mathrm{d}_\sigma \lambda_p)v).$$

Since λ_p is a diffeomorphism of G onto the fibre $\pi^{-1}(\pi(p))$ it follows that $(\mathrm{d}_\sigma \lambda_p)v \in \mathcal{V}_{p\sigma} P$ and

$$(\delta_\sigma \lambda_p)A_{p\sigma}(v) = (\mathrm{d}_{\mathrm{id}}\lambda_{p\sigma})^{-1}((\mathrm{d}_\sigma \lambda_p)v) = \mathrm{d}_\sigma(\lambda_{p\sigma}^{-1} \circ \lambda_p)v = (\mathrm{d}_{\mathrm{id}} l_\sigma)^{-1} v \tag{9.109}$$

where the last equality follows by differentiating the commutative diagram

$$\begin{tikzcd} G \arrow[rr, "\lambda_p"] \arrow[dr, "l_{\sigma^{-1}}"'] & & P \arrow[dl, "\lambda_{p\sigma}^{-1}"] \\ & G. & \end{tikzcd}$$

The $\mathfrak{g}$-valued 1-form $(\delta\lambda)A$ is actually left invariant since

$$\delta_\sigma l_\tau((\delta\lambda)A)_{\tau\sigma} = (\delta_\sigma l_\tau \circ \delta_{\tau\sigma}\lambda_p)A_{p\tau\sigma} = A_{p\tau\sigma} \circ d_\sigma(\lambda_p \circ l_\tau)$$

and, given $v \in T_\sigma G$, we get

$$\begin{aligned} \delta_\sigma l_\tau((\delta\lambda)A)_{\tau\sigma} v &= (\mathrm{d}_{\mathrm{id}}\lambda_{p\tau\sigma})^{-1}(\mathcal{V}_{p\tau\sigma}\mathrm{d}_\sigma(\lambda_p \circ l_\tau)v) \\ &= \mathrm{d}_\sigma(\lambda_{p\tau\sigma}^{-1} \circ \lambda_p \circ l_\tau)v = (\mathrm{d}_{\mathrm{id}} l_\sigma)^{-1} v = ((\delta\lambda)A)_\sigma v. \end{aligned}$$

Therefore $(\delta\lambda)A \in \mathfrak{g}^* \otimes \mathfrak{g} \cong Hom(\mathfrak{g}, \mathfrak{g})$: call it *the Maurer–Cartan form associated with the connection $\mathcal{H}P$*. By (9.109) it is the $\mathfrak{g}$-valued 1-form which assigns to each tangent vector to G its left invariant extension: hence its representative in $Hom(\mathfrak{g}, \mathfrak{g})$ is the identity $\mathrm{id}_{\mathfrak{g}}$ and the Maurer–Cartan equation (9.107) gives

$$\mathrm{d}(\delta\lambda)A(X, Y) = -(\delta\lambda)A[X, Y] = -[X, Y] = -[(\delta\lambda)AX, (\delta\lambda)AY].$$

Then we get

$$\mathrm{d}(\delta\lambda)A + \tfrac{1}{2}[(\delta\lambda)A, (\delta\lambda)A] = 0.$$

By defining $(\delta\lambda)\Omega$ just like we did for $(\delta\lambda)A$ in (9.108) the structure equation (9.104) and the last one allows us to conclude that

$$(\delta\lambda)\Omega = 0. \tag{9.110}$$

Since $\delta_{\mathrm{id}}\lambda_p$ realizes the isomorphism $\mathcal{V}_p^*P \cong \mathfrak{g}^*$ this actually means that *the curvature 2-form Ω vanishes on the tangent space to the fibre of P*. Hence the structure equation (9.104) can be rewritten as follows

$$\mathrm{d}A = \Omega - \tfrac{1}{2}[A, A]$$

to give a decomposition of $\mathrm{d}A$ into horizontal and vertical parts.

Let us now come back to consider the connection form A of $\mathcal{H}P$. It can be defined as in (9.23) since the connection $\mathcal{H}P$ determines a splitting in the tangent bundle TP. But the converse is also true and the connection $\mathcal{H}P$ may be obtained by the $\mathfrak{g}$-valued 1-form A just like the vector sub-bundle $\ker A$.

Proposition 9.8.6. *If A is the connection form of a connection $\mathcal{H}P$ then*

$$\begin{aligned} \forall p \in P, \forall u \in \mathcal{V}_p P \qquad (\mathrm{d}_{\mathrm{id}}\lambda_p)A_p u &= u \\ \forall \sigma \in G \qquad \delta R(\sigma)A &= \mathrm{Ad}_{\sigma^{-1}} \circ A. \end{aligned} \tag{9.111}$$

Conversely given a $\mathfrak{g}$-valued 1-form A on P satisfying these conditions the vector sub-bundle $\ker A \subset TP$ *gives a connection on P whose connection form is A. Hence the set $\mathcal{A}_P$ of all the connection on P can be identified with the affine subspace of $\Omega^1(P, \mathfrak{g})$ defined by conditions (9.111). Furthermore the curvature form $\Omega \in \Omega^2(P, \mathfrak{g})$ of $\mathcal{H}P$ is a $\mathfrak{g}$-valued 2-form such that*

$$\begin{aligned} \forall p \in P, \forall u, v \in \mathcal{V}_p P \qquad \Omega_p(u, v) &= 0 \\ \forall \sigma \in G \qquad \delta R(\sigma)\Omega &= \mathrm{Ad}_{\sigma^{-1}} \circ \Omega. \end{aligned} \tag{9.112}$$

Proof. The first equality in (9.111) follows immediately by the definition of the connection form A. For the second one note that

$$\delta_p R(\sigma)A_{p\sigma}(u) = A_{p\sigma}(\mathrm{d}_p R(\sigma)u) = (\mathrm{d}_{\mathrm{id}}\lambda_{p\sigma})^{-1}\mathcal{V}_{p\sigma}(\mathrm{d}_p R(\sigma)u).$$

The condition (9.22) for the connection $\mathcal{H}P$ implies that $\mathcal{V}_{p\sigma}(\mathrm{d}_p R(\sigma)u) = \mathrm{d}_p R(\sigma)(\mathcal{V}_p u)$. However, $\mathcal{V}_p u = \mathrm{d}_{\mathrm{id}}\lambda_p(A_p u)$ and we can write

$$\delta_p R(\sigma) A_{p\sigma}(u) = (\mathrm{d}_{\mathrm{id}}\lambda_{p\sigma})^{-1} \circ \mathrm{d}_p R(\sigma) \circ \mathrm{d}_{\mathrm{id}}\lambda_p(A_p u) = Ad_{\sigma^{-1}} \circ A(u)$$

where the last equality follows by the commutative diagram

$$\begin{array}{ccc} \pi^{-1}(\pi(p)) & \stackrel{R(\sigma)}{\longrightarrow} & \pi^{-1}(\pi(p\sigma)) \\ \lambda_p \uparrow & & \downarrow \lambda_{p\sigma}^{-1} \\ G & \stackrel{a_{\sigma^{-1}}}{\longrightarrow} & G. \end{array}$$

For the converse it suffices to observe that the first equality in (9.111) gives the splitting condition (9.21) and the second one ensures the G-invariance (9.22) for ker A. Hence it is a connection on P whose connection form is clearly A.

Finally the first equality in (9.112) follows by (9.110) and the second one by applying the second equality in (9.111) to the definition (9.24) of Ω. □

Let us recall that a *gauge transformation of* P is an automorphism φ of P which induces the identity map on the base manifold M. Then it leaves every fibre fixed and it takes sense to define the associated map

$$\sigma_\varphi : P \longrightarrow G \tag{9.113}$$

such that $\varphi(p) = p\sigma_\varphi(p)$. By applying the *Liebnitz rule* to the connection form A we get that

$$(\delta_p\varphi)A_{\varphi(p)} = \delta_p R(\sigma_\varphi(p))A_{p\sigma_\varphi(p)} + (\delta_p\sigma_\varphi)(\delta\lambda)A_{\sigma_\varphi(p)}$$

where $(\delta\lambda)A$ is the Maurer–Cartan form of the given connection. The second equation in (9.111) allows us to conclude that *under a gauge transformation* φ *the connection form A behaves as follows*

$$(\delta\varphi)A = Ad_{\sigma_\varphi^{-1}} \circ A + (\delta\sigma_\varphi)(\delta\lambda)A. \tag{9.114}$$

If Ω *is the associated curvature then by* (9.110) *and* (9.112) *it transforms under* φ *as follows*

$$(\delta\varphi)\Omega = Ad_{\sigma_\varphi^{-1}} \circ \Omega. \tag{9.115}$$

Since gauge transformations on P form a group $\mathcal{G}_P$ with respect to the composition, (9.114) defines an action of $\mathcal{G}_P$ on the affine space of connections $\mathcal{A}_P$.

Let us now consider the *exponential map* $\exp : \mathfrak{g} \longrightarrow G$ which assigns to a left invariant vector field $X \in \mathfrak{g}$ the point $\exp_X(1) \in G$ where $\exp_X(t)$ is the unique one-parameter group whose tangent vector at $0 \in \mathbb{R}$ is $X_{\mathrm{id}} \in T_{\mathrm{id}}G$. Since

$Ad_\sigma \in \mathrm{Aut}(\mathfrak{g})$, for every $\sigma \in G$, and the Lie algebra of $\mathrm{Aut}(\mathfrak{g})$ is $\mathrm{End}(\mathfrak{g})$ we get the following commutative diagram

$$\begin{array}{ccc} G & \xrightarrow{Ad} & \mathrm{Aut}(\mathfrak{g}) \\ \exp \uparrow & & \uparrow \exp \\ \mathfrak{g} & \xrightarrow{ad} & \mathrm{End}(\mathfrak{g}) \end{array}$$

where $ad := d(Ad)$.

Definition 9.8.7. *For every $X, Y \in \mathfrak{g}$ the symmetric bilinear form*

$$\langle X, Y \rangle := \mathrm{tr}(ad_X \circ ad_Y)$$

is called the Killing form of the lie algebra $\mathfrak{g}$.

Given a point $m \in M$ recall the definition (9.26) of the *holonomy group* $\mathrm{Hol}_{\mathcal{H}P}(m)$ *of a connection $\mathcal{H}\ P$ at $m \in M$*. If the base manifold M is connected all these groups are isomorphic when m varies in M since we can send

$$h_\gamma \in \mathrm{Hol}_{\mathcal{H}P}(m_1) \longmapsto h_{\alpha * \gamma * \overline{\alpha}} \in \mathrm{Hol}_{\mathcal{H}P}(m_2) \tag{9.116}$$

where α is a path from m_1 to m_2 and $\overline{\alpha}$ its reversed path. Then it make sense to define *the holonomy group* $\mathrm{Hol}_{\mathcal{H}P}$ *of the connection* $\mathcal{H}P$.

9.9 Appendix: More on Witten's open-string theory interpretation of QFT

Sketch of proof of theorem 9.3.2: We have to show that under the assumptions (9.79) and (9.80) the weak-coupling limit of the abstract string Lagrangian reduces exactly to the Lagrangian of a QFT on L.

The low-energy (or weak-coupling) limit of a string theory is only approximated by a QFT since the limit Lagrangian admits perturbative corrections depending on the coupling constant and non-constant instanton corrections (see definition 9.3.8). The string theory analysed in Witten (1992) is a topological theory given by an A-twisted σ-model. At first Witten observes that this model does not depend on the coupling constant of the theory implying that there cannot be any perturbative correction in the limit Lagrangian.

It remains then to show that *all the non-constant instanton contributions vanish*. Let σ be the canonical symplectic form on $\widehat{Y} = T^*L$. It is the differential of the Liouville form, i.e. in local canonical coordinates $\sigma = \mathrm{d}\vartheta$ where $\vartheta := \sum_{j=1}^{3} p_j\, \mathrm{d}q_j$. The Liouville form vanishes on L given by $p_1 = p_2 = p_3 = 0$. Note that the bosonic sigma model action reduces for instantons to be

$$I = \int_\Sigma \phi^*(\sigma).$$

Stokes' theorem and condition (9.80) suffice to conclude that

$$I(\phi) = 0 \tag{9.117}$$

for *all* instantons ϕ. However, by its definition the bosonic σ-model action I vanishes *only* for constant instantons. Hence we can admit only constant instanton corrections and the abstract string Lagrangian reduces exactly to the Lagrangian of the QFT on L realizing the low-energy limit. In the A-twisted case such a limit turns out to be exactly a Chern–Simons $U(N)$ gauge theory.

Dropping assumption (9.79). The main result of Witten (1992) is more general than theorem 9.3.2. In fact he analyses (section 4.4) the low-energy limit of an A-twisted topological open-string theory whose target space is given by a Calabi–Yau threefold $\widehat{Y}$ admitting L as a Lagrangian submanifold.

Theorem 9.9.1. *Let $\widehat{Y}$ be a local Calabi–Yau threefold and $L \subset \widehat{Y}$ a Lagrangian submanifold. Then there exist topological string theories with $\widehat{Y}$ as target space, such that their open sectors are equivalent to a QFT on L up to the convergence of non-constant instanton contributions. In the A-twisted case the Lagrangian action of the limit QFT is (if convergent) a deformation of a Chern–Simons action.*

This result follows by assuming the same boundary conditions as before. But now (9.80) is no more sufficient to conclude the vanishing (9.117) for non-constant instantons: given ϕ its *instanton number* is

$$q(\phi) := \int_{\Sigma} \phi^*(\omega)$$

where ω is the symplectic form of $\widehat{Y}$. Instanton numbers turns out to be non-negative. For any knot $\mathcal{K} \subset \phi(\partial\Sigma) \subset L$ consider the Wilson line $W^R_{\mathcal{K}}$ constructed by holonomy on L. For a given connection A on a $U(N)$ principal bundle Witten shows that the instanton contribution of ϕ is given by

$$-\frac{\mathrm{i}\eta(\phi)\mathrm{e}^{-\theta q(\phi)}}{2\pi k} \sum_{\mathcal{K}\subset\phi(\partial\Sigma)} \log(\mathrm{tr}_R(h_{\mathcal{K}}))$$

where θ is a positive real parameter, $\mathrm{e}^{-\theta q(\phi)}$ a suitable weighting factor and $\eta(\phi) = \pm 1$. If $S(\mathcal{L}(A))$ is the Chern–Simons action on L the limit action turns out to be

$$S' = S(\mathcal{L}(A)) - \frac{\mathrm{i}}{2\pi k} \sum_{\phi} \left[\eta(\phi)\mathrm{e}^{-\theta q(\phi)} \sum_{\mathcal{K}\subset\phi(\partial\Sigma)} \log(\mathrm{tr}_R(h_{\mathcal{K}})) \right]. \tag{9.118}$$

Under suitable assumptions on the 'moduli space' of instantons ϕ the sum can be perturbatively evaluated for $\theta \gg 0$.

Corollary 9.9.2. *Assume that* $\widehat{Y} = T^*S^3$ *and* $L = \mathcal{C}$ *is the Lagrangian submanifold given by the conormal bundle of the unknot knot in* S^3 *as in proposition 9.3.11. Then the low-energy limit QFT on* $\mathcal{C}$ *of the open sector of a type IIA string theory with* M *D-branes wrapped around* $\mathcal{C}$ *is a* $SU(M)$ *Chern–Simons gauge theory on* $\mathcal{C}$*. Moreover, the global open-string theory with* N *D-branes wrapped around* S^3 *and* M *D-branes wrapped around* $\mathcal{C}$ *admits a low-energy limit QFT whose action is the following deformation of the* $SU(M)$ *Chern–Simons action on* $\mathcal{C}$*;*

$$S' = S(\mathcal{L}) - \frac{\mathrm{i}}{2\pi k} \sum_d \eta_d \log(\mathrm{tr}_R(h_{\mathcal{K}}^d)).$$

The first part of the statement can be proven as theorem 9.3.2 since the Liouville form of $\mathbb{R}^8 \supset T^*S^3$ vanishes when restricted to $\mathcal{C}$, as in (9.91). That is enough to guarantee the vanishing (9.117).

To prove the second part, note that the only non-trivial non-constant contributions comes from instantons ϕ such that $\phi(\partial\Sigma)$ is a d-covering of the unknot in S^3. For these instantons $q(\phi) = 0$ by Stokes' theorem and the thesis follows by (9.118).

References

Acharya B S 1999 M theory Joyce orbifold and super Yang–Mills *Adv. Theor. Math. Phys.* **3** 227–48

Altmann K 1997 The versal deformation of an isolated toric singularity *Inv. Math.* **128**(3) 443–79

Aspinwall P S and Morrison D R 1993 Topological field theory and rational curves *Commun. Math. Phys.* **151** 245–62

Aspinwall P S, Greene B and Morrison D R 1994 Calabi–Yau moduli space, mirror manifolds and spacetime topology change in string theory *Nucl. Phys.* B **416** 414–80

Atiyah M F 1989 Topological quantum field theories *Publ. Math. I.H.E.S.* **68** 175–86

——1990a *The Geometry and Physics of Knots* (Cambridge: Cambridge University Press)

——1990b On framing of 3-manifolds *Topology* **29** 1–8

Atiyah M F and Bott R 1982 The Yang–Mills equations over Riemann surfaces *Phil. Trans. R. Soc.* A **308** 523–615

Atiyah M F, Maldacena J and Vafa C 2001 An M-theory flop as a large N duality *J. Math. Phys.* **42** 3209-20

Atiyah M F and Witten E 2001 M-theory dynamics on a manifold of G_2 holonomy *Preprint* hep-th/0107177

Avram A, Candelas P, Jančić D and Mandelberg M 1996 On the connectedness of the moduli space of Calabi–Yau manifolds *Nucl. Phys.* B **465** 458–72

Axelrod S, Della Pietra S and Witten E 1991 Geometric quantization of Chern–Simons gauge theory *J. Diff. Geom.* **33** 787–902

Barth W, Peters C and van de Ven A 1984 *Compact Complex Surfaces* (Berlin: Springer)

Batyrev V, Ciocan-Fontanine I, Kim B and van Straten D 1998 Conifold transitions and mirror symmetry for Calabi–Yau complete intersections in Grassmannians *Nucl. Phys.* B **514** 640–6

Berglund P, Katz S and Klemm A 1995 Mirror symmetry and the moduli space for the generic hypersurfaces in toric varieties *Nucl. Phys.* B **456** 153–204

Berglund P, Katz S, Klemm A and Mayr P 1997 New Higgs transition between dual $N = 2$ string models *Nucl. Phys.* B **483** 209–28

Bershadsky M, Cecotti S, Ooguri H and Vafa C 1993 Holomorphic anomalies in topological field theories (appendix by Katz S) *Nucl. Phys.* B **405** 279–304 and in *Mirror Symmetry (Stud. Adv. Math. 1)* vol II (Providence, RI: American Mathematical Society and International Press) pp 655–82

——1994 Kodaira–Spencer theory of gravity and exact results for quantum string amplitudes *Commun. Math. Phys.* **165** 311–428

Brandhuber A, Gomis J, Gubser S S and Gukov S 2001 Gauge theory at large N and new G_2 holonomy metrics *Nucl. Phys.* B **611** 179–204

Brieskorn E 1966 Über die Auflösung gewisser Singularitäten von holomorphen Abbildungen *Math. Ann.* **166** 76–102

——1968 Die Auflösung der rationalitaten Singularitäten holomorpher Abbildungen *Math. Ann.* **178** 255–70

Bryant R L 1987 Metrics with exceptional holonomy *Ann. Math.* **126** 525–76

Bryant R L and Salamon S M 1989 On the construction of some complete metrics with exceptional holonomy *Duke Math. J.* **58** 829–50

Candelas P and de la Ossa X C 1990 Comments on conifolds *Nucl. Phys.* B **342**(1) 246–68

Candelas P, de la Ossa X C, Green P S and Parkes L 1991 A pair of Calabi–Yau manifolds as an exactly soluble superconformal theory *Nucl. Phys.* B **359** 21–74 and *Essays on Mirror Manifolds* (International Press) pp 31–95

Candelas P, de la Ossa X C, Font A, Katz S and Morrison D R 1994 Mirror symmetry for two parameter models I *Nucl. Phys.* B **416** 481–538

Chern S and Simons J 1974 Characteristic forms and geometric invariants *Ann. Math.* **99** 48–69

Chiang T–M, Greene B, Gross M and Kanter Y 1996 Black hole condensation and the web of Calabi–Yau manifold *Nucl. Phys. Proc. Suppl.* **46** 82–95

Clemens C H 1983 Double solids *Adv. Math.* **47** 107–230

Clemens H, Kollár J and Mori S 1988 Higher dimensional geometry *Asterisque* **166**

Cox A D and Katz S 1999 *Mirror Symmetry and Algebraic Geometry (Math. Surveys and Monographs 68)* (Providence, RI: American Mathematical Society)

Deligne P, Etingof P, Freed D S, Jeffrey L C, Kazhdan D, Morgan J W, Morrison D and Witten E (ed) 1999 *Quantum Fields and Strings: A Course for Mathematicians* vols I, II (Providence, RI: American Mathematical Society)

de Wit B, Lauwers P and van Proeyen A 1985 Lagrangians of $N = 2$ supergravity–matter systems *Nucl. Phys.* B **255** 569–608

Donaldson S K 1983 A new proof of a theorem of Narasimhan and Seshadri *J. Diff. Geom.* **18** 269–77

Douglas M, Katz S and Vafa C 1997 Small instantons, del Pezzo surfaces and type I-theory *Nucl. Phys.* B **497** 155–72

Faber C and Pandharipande R 2000 Hodge integrals and Gromov–Witten theory *Inv. Math.* **139** 173–99

Freed D S 1995 Classical Chern–Simons theory, Part 1 *Adv. Math.* **113** 237–303

Freyd P, Yetter D, Hoste J, Lickorish W B R, Millet K and Ocneanu A 1985 A new polynomial invariant of knots and links *Bull. Am. Math. Soc.* **12** 183–312

Friedman R 1986 Simultaneous resolution of threefold double points *Math. Ann.* **274** 671–89

Fulton W and Pandharipande R 1995 Notes on stable maps and quantum cohomology *Algebraic Geometry–Santa Cruz 1995 (Proc. Symp. Pure Math. 62)* Part 2 (Providence, RI: American Mathematical Society) pp 45–96

Getzler E and Pandharipande R 1998 Virasoro constraints and the Chern classes of the Hodge bundle *Preprint* math.AG/9805114

Gibbons G W, Page D N and Pope C N 1990 Einstein metrics on S^3, $\mathbb{R}^3$ and $\mathbb{R}^4$ bundles *Commun. Math. Phys.* **127** 529–53

Gopakumar R and Vafa C 1998a M-theory and topological strings—I *Preprint* hep-th/9809187

——1998b *M*-theory and topological strings–II *Preprint* hep-th/9812127

——1999 On the gauge theory/geometry correspondence *Adv. Theor. Math. Phys.* **3** 1415–43

Greene B, Morrison D R and Strominger A 1995 Black hole condensation and the unification of string vacua *Nucl. Phys.* B **451** 109–20

Greene B and Plesser M R 1990 Duality in Calabi–Yau moduli space *Nucl. Phys.* B **338** 15–37

Griffiths P 1969 On the periods of certain rational integrals I, II *Ann. Math.* **90** 460–95, 498–541

Gross M 1997a Deforming Calabi–Yau threefolds *Math. Ann. 308* **2** 187–220

——1997b Primitive Calabi–Yau threefolds *J. Diff. Geom.* **45** 288–318

Helgason S 1978 *Differential Geometry, Lie Groups and Symmetric Spaces* (New York: Academic)

Hitchin N J 1990 Flat connection and geometric quantization *Commun. Math. Phys.* **131** 347–80

Horowitz G and Strominger A 1991 Black Strings and *p*-branes *Nucl. Phys.* B **360** 197–209

Hosono S, Klemm A, Theisen S and Yau S T 1995 Mirror symmetry, mirror map and applications to complete intersection Calabi–Yau spaces *Nucl. Phys.* B **433** 501–54 and *Mirror Symmetry (Stud. Adv. Math. 1)* vol II (Providence, RI: American Mathematical Society and International Press) pp 545–606

Jones V 1985 A polynomial invariant for knots via von Neumann algebras *Bull. Am. Math. Soc.* **12** 103–11

——1987 Hecke algebras representations of braid groups and link polynomials *Ann. Math.* **126** 335–88

Johnson C V 1999 Etudes on D-Branes *Non-perturbative Aspects of Strings Branes and Supersymmetry* (Singapore: World Scientific)

——2000 D-Brane primer *Preprint* hep-th/0007170

Joyce D 1996 Compact Riemann 7-manifolds with holonomy G_2 I, II *J. Diff. Geom.* **43** 291–328, 329–75

Joyce D 2000 *Compact Manifolds with Special Holonomy* (Oxford: Oxford Science Publications)

Katz S and Liu C C M 2001 Enumerative geometry of stable maps with Lagrangian boundary conditions and multiple covers of the disc *Adv. Theor. Math. Phys.* **5** 1

Katz S, Morrison D R and Plesser M 1996 Enhanced gauge symmetry in type II string theory *Nucl. Phys.* B **477** 105–40

Kodaira K 1954 On Kähler varieties of restricted type (an intrinsic characterization of algebraic varieties) *Ann. Math.* **60** 28–48

Kollár J 1989 Flops *Nagoya Math. J.* **113** 14–36

Labastida J M F 1999 Chern–Simons gauge theory: ten years after *Eprint* hep-th/9905057

Labastida J M F, Mariño M and Vafa C 2000 Knots, links and branes at large N *J. High Energy Phys.* 0011:007

Labastida J M F and Mariño M 2001 Polynomial invariants for torus knots and topological strings *Commun. Math. Phys.* **217** 423–49

Li J and Song Y S 2001 Open string instantons and relative stable morphisms *Adv. Theor. Math. Phys.* **5** 67–91

Manin Y 1995 Generating functions in algebraic geometry and sums over trees 2 *The Moduli Space of Curves* (Boston, MA: Birkäuser) pp 401–17

Manzoni A 1842 I Promessi Sposi *The Betrothed*

Mariño M and Vafa C 2001 Framed knots at large *N Preprint* hep-th/0108064

Moore G and Sieberg N 1988 Polynomial equations for rational conformal field theories *Phys. Lett.* B **212** 451–66

Morrison D R 1985 The birational geometry of surfaces with rational double points *Math. Ann.* **271** 415–38

——1999 Through the looking glass *Mirror Symmetry* vol III (Providence, RI: American Mathematical Society and International Press) pp 263–77

Morrison D R and Seiberg N 1997 Extremal transitions and five-dimensional supersymmetric field theories *Nucl. Phys.* B **483** 229–47

Narasimhan M S and Seshadri C S 1965 Stable and unitary vector bundles on a compact Riemann surface *Ann. Math.* **82** 540–67

Okonek C, Schneider M and Spindler H 1980 *Vector Bundles on Complex Projective Spaces (Progr. Math. 3)* (Boston, MA: Birkäuser)

Ooguri H and Vafa C 2000 Knot invariants and topological strings *Nucl. Phys.* B **577** 419–38

Periwal V 1993 Topological closed–string interpretation of Chern–Simons theory *Phys. Rev. Lett.* **71** 1295–8

Pinkham H 1983 Factorization of birational maps in dimension 3 *Singularities (Proc. Sym. Pure Math. 40)* (Providence, RI: American Mathematical Society) pp 343–72

Poor W 1981 *Differential Geometric Structures* (New York: McGraw-Hill)

Quillen D G 1986 Determinants of Cauchy–Riemann operators over a Riemann surface *Funct. Anal. Appl.* **19** 31

Quinn F 1995 Lectures on axiomatic topological quantum field theory *Geometry and Quantum Field Theory (IAS/Park City Math. Series 1)* (Providence, RI: American Mathematical Society) pp 325–459

Ramadevi P and Sarkar T 2001 On link invariants and topological string amplitude *Nucl. Phys.* B **600** 487–511

Reid M 1980 Canonical 3-Fods *Journées de Géométrie Algébrique d'Angers* (Sijthoff and Noorddhoff) pp 273–310

——1983 Minimal models of canonical 3-Fods *Algebraic Varieties and Analytic Varieties (Adv. Stud. Pure Math. 1)* (Amsterdam: North-Holland) pp 131–80

——1987a Young person's guide to canonical singularities *Algebraic Geometry 1, Bowdoin 1985 (Proc. Sym. Pure Math. 46)* (Providence, RI: American Mathematical Society) pp 354–414

——1987b The moduli space of 3-Fods with $K = 0$ may neverthless be irreducible *Math. Ann.* **278** 329–34

Reidemeister K 1933 Zür dreidimensionalen topologie *Abh. Math. Sem. Univ. Hamburg* **9** 189–94

Schlessinger M 1971 Rigidity of quotient singularities *Inv. Math.* **14** 17–26

Segal G 1988 Two-dimensional conformal field theories and modular functions *Proc. Int. Congr. Math. Phys., Swansea* pp 22–37

Seiberg N and Witten E 1994 Electromagnetic duality, monopole condensation and confinement in $N = 2$ supersymmetric Yang–Mills theory *Nucl. Phys.* B **426** 19–52

Sen A 1997 An introduction to non-perturbative string theory *Duality and Supersymmetric Theories* (Cambridge: Cambridge University Press)

Silverman J H 1994 *Advanced Topics in the Arithmetics of Elliptic Curves (Graduate Texts in Mathematics 151)* (Berlin: Springer)

Slodowy 1980 *Simple Singularities and Simple Algebraic Groups (Lecture Notes in Mathematics 815)* (Berlin: Springer)

Strominger A 1990 Special geometry *Commun. Math. Phys.* **133** 163–80

——1995 Massless black holes and conifolds in string theory *Nucl. Phys.* B **451** 97–109

't Hooft G 1974 A planar diagram theory for strong interactions *Nucl. Phys.* B **72** 461–73

Townsend P 1995 The eleven-dimensional supermembrane revisited *Nucl. Phys.* B **350** 184–7

Tyurina G 1970 Resolution of the singularities of flat deformations of rational double points *Funkts. Analiz. Prilozh.* **4** 77–83

Verlinde E 1988 Fusion rules and modular transformations in 2*d* conformal field theory *Nucl. Phys.* B **300** 360–83

Vafa C 2001 Superstrings and topological strings at large *N* *J. Math. Phys.* **42** 2798–817

Voisin C 1996 A mathematical proof of a formula of Anspinwall and Morrison *Comput. Math.* **104** 135–51

Warner F 1983 *Foundations of Differentiable Manifolds and Lie Groups (Graduate Texts in Mathematics 94)* (Berlin: Springer)

Werner J and van Geemen B 1990 New examples of threefolds with $c_1 = 0$ *Math. Z.* **203** 211–25

Wilson P M H 1989 Calabi–Yau manifolds with large Picard number *Invent. Math.* **98** 139–55

——1992 The Kähler cone on Calabi–Yau threefolds *Inv. Math.* **107** 561–83

Witten E 1989 Quantum field theory and the Jones polynomial *Commun. Math. Phys.* **121** 351–99

——1992 Chern–Simons gauge theory as a string theory *The Floer Memorial Volume* (Boston, MA: Birkhäuser) p 637

——1993 Phase of $N = 2$ theories in two-dimension *Nucl. Phys.* B **403** 159–222

——1995 String theory dynamics in various dimensions *Nucl. Phys.* B **443** 85–126

Index