СООБЩЕНИЯ
ОБЪЕДИНЕННОГО
ИНСТИТУТА
ЯДЕРНЫХ
ИССЛЕДОВАНИЙ

Дубна

E10-95-268

V.I.Ilyushchenko

LATHOLOGY — THE REVERSIBLE DETERMINISTIC THEORY OF ERRORS

1995

1 Introduction

The main trend in modern high energy physics is the gradual transition from relatively high statistics experiments to relatively low statistics ones [1].

As a semiquantitative boundary between these two regions we use the value 10^4 events/bin, which corresponds to the relative error of 1% for Poisson statistics. This limit corresponds to the applicability threshold of the PTT (Philips-Twomey-Tikhonov) - regularization, which is presently the only mathematical means to solve inverse problems with erraneous input data. For a standard 100 bin histogram the overall low statistics will amount to 10^6 events.

Another semiquantitative estimate follows from probabilistic nature of the generally accepted and the only existing now Gaussian theory of errors rooted in the algebraic theory of quadratic forms. The logically nonmotivated choice of the squared discrete Holder norm with $p = 2$ (Euclidean norm) results in nonrobust (fragile, i.e. error-dependent) local optimization solutions of scalar objective functionals. All estimates of data sample -- mean values, deviations etc. - are valid only in the limits due to the Big Number Theorem (BNT). In other words, the main underlying mathematical hypothesis of the standard theory of errors is that of normal distribution.

Here we describe the first version of the deterministic theory of errors -- lathology (from the Greek word lathos - error - VII) based on the reversibility axiom and free of any hypotheses concerning underlying statistical distributions.

2 The Second General Form of Inverse Problem Solution

The standard mixed matrix-vector form of direct problem (folding or convolution) solution -- mainly known in the discrete form of the System of Linear Algebraic Equations (SLAE) - is

$$At = f + n \tag{1}$$

where A - a square matrix, t - a column vector of true solution, f - a column vector of input data and n - that of input data errors. It is useful to note the three following important features: first, the very elements of the SLAE (1) do not form an algebraic group, second, this standard mixed matrix-vector form allows only an additive error hypothesis to be described algebraically and, third, all computations are usually performed within a real number basis with the inherent rounding-off errors.

The first general form (and up to now the only known one – VII) of inverse problem solution (unfolding or deconvolution) looks like

$$t = A^{-}(f + n) \tag{2}$$

where A^{-} is an inverse matrix of A. In trying to compute (2) we always observe the fundamental computational instability patterns in the form of a bin-by-bin periodic (sinus-like) function within positive and negative envelopes of input data vector f, which we call the Signature-Envelope Function (SEF) and which looks like $SEF(m) = (-1)^m f(m)$, where m - the bin number and $f(m)$ - the input data vector (envelope).

What is the source of this instability ? The only presently proposed answer is based on the Riemann-Lebesgue theorem well-known in the theory of Fourier series coefficients [2], i.e.

$$\int_{\alpha \to \infty} A(x, s) \sin(\alpha * s) ds \to 0 \tag{3}$$

This Limiting Virtual Zero (LVZ) is usually added to the left-hand-side (l.h.s.) of the SLAE (1) in a discrete form to yield a solution with the accompanying periodic instability

$$SEF(m) = A^{-}(f + n) \tag{4}$$

The Riemann-Lebesgue theorem is not, however, the only version of the LVZ – any converging series tending to zero will serve as well, i.e. the number of potential candidates is infinite.

Now let us differentiate the above relation (1):

$$A't + At' = (f + n)' \tag{5}$$

to get the second general form of the inverse problem solution

$$t = (A')^{-}[(f + n)' - At'] \tag{6}$$

where all r.h.s. terms contain derivatives. The behaviour of numerical derivatives of error-stricken functions has been systematically studied ealier [3]. Their patterns closely fit those observed in solving (2).

3 Standard Gaussian Theory of Additive Errors

3.1 General Exposition

C.F.Gauss (1777-1855) developed his theory of additive errors in 1821-1823 for processing mainly astronomical and magnetic measurements [4], [5] specified by typical relative errors of 10^{-4}. It is important to note that the modern-like notions of matrix and SLAE were advanced for the first time by A.Cayley (1821-1895) only in 1851 [6] and were quite naturally not used by C.F.Gauss. The Gaussian theory of additive errors is strongly motivated by the main object of mathematical studies by C.F.Gauss, i.e. quadratic forms.

Part I of the above cited work [4] is dedicated to the systematic and theoretical development of the theory of additive errors based on probability theory. From two principally different types of errors - systematic and statistical (Zufallsfehler, i.e. random errors - VII) the Part I deals only with the latter ones. Gauss defines the function $\varphi(x)$ as a relative error for an observation x. Then $\varphi(x)dx$ is the probability of an error between x and $x + dx$, with $\varphi(x)$ to be normalized via the condition

$$\int_{-\infty}^{+\infty} \varphi(x)dx = 1.0 \tag{7}$$

The subsequent basic minimum requirement leads to the Legendre-Gauss Least SQuares (LSQ) being the main topic of Part II of the cited Gaussian work:

$$\int x^2\varphi(x)dx = MIN! \tag{8}$$

Here Gauss used the implicit hypothesis that the most suitable weight for an error is its square.

With the conceptual basis of the theory of errors formed, one must find out a suitable function $\varphi(x)$. As has been truly indicated in [5], generally the error distribution is unknown beforhand and one can choose any arbitrary function satisfying (7). Once again Gauss introduces some kind of (exponential) quadratic form

$$\varphi(x) = e^{-x^2} \tag{9}$$

which is well known presently as the normal distribution.

The modern noncritical version of the Gaussian theory of errors can be found e.g. in [1]. In particular, there it is shown that "...the Gaussian or normal distribution gets its importance in large part from the Central Limit Theorem..." also known as the Big Number Theorem (BNT). The BNT applicability limits are, however, known only qualitatively - in general, its validity diminishes with the statistics involved. Here again the absence of alternative theories of errors forces these BNT applicability limits to be implicitely extended to low statistics as well.

3.2 Main Defects of the Gaussian Theory of Errors

From our point of view, based on many years of computational experience, the first main defect of the Gaussian theory of errors is due to its arbitrary conceptual basis, with all underlying initial hypotheses taken from the theory of quadratic forms.

The second defect is illustrated by the author of the theory himself in the Part II of [4], where he describes an involved procedure for discarding "bad" observations. In other words, already Gauss observed the nonrobust, i.e. error-dependent, character of statistical estimates – means, deviations etc. – of the statistical samples under study. In the language of normalized spaces Gauss used the squared discrete Holder norm with $p = 2$ (Euclidean norm) as an objective functional to be minimized. All the subsequent attempts to find out some discrete robust Holder norms e.g. with $p = 1$ (Manhattan norm) or $p = \infty$ (minimax or Chebyshev norm) failed.

Moreover, the analysis of the weighed LSQ results shows that the relevant objective functional behaves like a disrete Holder norm with $p < 1$, i.e. it possesses a set of weight-induced parasitic local minima. Since all the presently used minimization codes with discrete objective functionals realize only local minimization algorithms, all the final data processing results are nonrobust by definition. This latter point is amply illustrated by the so-called history plots (see [1], Fig.2 on p.1183) and multipeak structures in the weighed average plots (see [1], p.1343 f.) witnessing the effects of systematic and other non-random errors. As a net result, the widely popular χ^2-criterion tending to 1.0 per each degree-of-freedom becomes obsolete.

To sum up, the only existing now theory of errors is probabilistic in its conceptual basis, with all underlying analytical hypotheses chosen from the Gaussian theory of quadratic forms. This theory deals only with statistical (random) errors, the BNT controlled samples and is nonrobust (error-dependent) by definition.

4 Reversibility Axiom

The reversibility axiom forming the logical basis of reversible mathematics can be formulated as follows: " The left-hand-side and the right-hand-side of any equation must satisfy reversibility criterions, i.e. the requirements of the unique term-by-term correspondence and equivalence relations" [7].

This axiom allows to identify a set of mathematical and/or physical transforms resulting in the all-matrix SLAE form:

$$AT = F + N \tag{10}$$

It is very important that the very elements of the latter SLAE now form a group. The inverse transforms from the old SLAE (1) to the novel one (10) are very simple, thus preserving all the algebraic results so far valid for the standard mixed vector-matrix SLAE form.

5 Fundamentals of Lathology

By means of the reversibility axiom the all-matrix SLAE (10) can be written down as

$$(A_F + A_{+N})T = F + N \tag{11}$$

where A_F and A_{+N} are the square matrices corresponding to the relevant r.h.s. terms F and +N. First, the novel all-matrix SLAE form enables these new matrices to be computed explicitely, while the standard mixed form (1) does not.

Second, the true solution T is assumed to be devoid of any errors.

Third, we can map to the l.h.s. of the basic equation (10) the r.h.s. error terms of any type, i.e. statistical, systematic, individual etc. without any explicit assumptions about the underlying distributions, in other words, in a quite deterministic way.

As an example, let us consider a Poisson data vector f specified by the error interval $\pm n$. In lathology this vector will be specified by the "data" matrix A_f and two lathological ("error") matrices A_{+n} and A_{-n}. For the purpose of convenience these three basic matrices can be described by some relevant matrix Holder norms $\| * \|_p$, i.e. by three scalars,
$d = \| A_f \|_p$,
$+l = \| A_{+n} \|_p$ and
$-l = \| A_{-n} \|_p$,
or another characteristic matrix parameters like determinants etc.

Another example is illustrated by the so-called regular unfolding, when the direct (folding) problem is written down as

$$At = (I^{\%} + R)t = f_{reg} = t + r \tag{12}$$

where $I^{\%}$ - the diagonal scalar unit matrix and R - the residual (error) matrix with the zero main diagonal. This is the case of a spectrometer with the unity acceptance factor matrix and only the resolution errors present in the r.h.s. input data vector f_{reg}. Then the non-resolution (systematic etc.) error vector, s, can be evaluated from the difference

$$s = f - f_{reg} \tag{13}$$

Another version of the systematic error evaluation can be realized in the case of two or more spectrometers measuring the same physical process in the same particle beam, which situation is quite typical for modern big accelerators and colliders.

The problem of the robustness, i.e. the error independence of final computational results (vector t) can be also efficiently solved within the reversible mathematical approach by means of generalized multiple constraint mathematical means and will be exposed by the present author in some subsequent paper.

By using the all-matrix SLAE form (10) one can evaluate the systematic errors of different origin, e.g. intrarun ones.

6 Classification of Systematic Errors

The problem of systematic errors seems to be most difficult to be handled even at conceptual and/or phenomenological level. The existing data on the properties of elementary particles [1] demonstrate, however, their presence in all published data sets – see, e.g. the summed–up data with two and three Gaussians and history plots with different systematic trends.

The possible sources of systematic errors can be, nevertheless, classified into two main categories: mathematical and engineering-physical.

6.1 Some Software (Mathematical) Sources of Systematic Errors

1. The first source of these errors is self–evident – this is due to the very folding (convolution) operation like (1) or (10). For a typical effective mass spectrum (S) with two main components – peaks (P) and combinatorial background (B) – the first will be broadened and shifted, while the second will be gradually enhanced. The standard fitting procedures usually approximate B as a smoothed polynomial, which is afterwards subtracted from S to provide "clean" final data on P. This procedure, if done without a previous unfolding, results in an elimination of lower (substrate) parts of P, i.e. in a systematic underestimation of the P content (e.g. area). The second important systematic error term stems from nonresolved multiplets seen and interpreted in the folded spectrum S as singlets.

2. The second source of the systematic errors is due to the run structure of the data samples gathered in different experimental runs (intrarun shifts). The standard linear model (1) deals either with partial run data samples or an integral data sample. The principal feature of the all-matrix linear model (10) is the possibility of individual treatment of run samples as elements of the r.h.s. matrix F and of the integral statistics with a preserved run structure. In addition, we can process randomized matrix samples, thus eliminating artificially the effect of the intrarun shifts. The resulting difference is an estimate of this type of systematic errors.

3. At the majority of modern high energy accelerators and colliders several similar setups are positioned along the same beam line. This provides a possibility of measuring the same physical process simultaneously by means of all acting spectrometer setups to evaluate intrasetup shifts. With one of the setups chosen as a basic one it is possible to analyze the relevant systematic error terms.

6.2 The Folding Shifts Induced along Abscissa Axis

The above phenomenology pertains mainly to the folding shifts induced along ordinate (amplitude, number of events etc.) axis.

The folding shifts induced along abscissa (wavelength, energy etc.) axis are due to the symmetry of the apparatus matrix A and of its factor matrices. The resolution factor matrix is mainly Toeplitzean and hence symmetric, while remaining three factor matrices are asymmetric by definition. The resulting shifts are bipolar so that e.g. the folded peaks can shift to any position relative to the true peak one depending on the specific form of the factor matrices.

6.3 MC Modelling Versus Experimental Data

The physical sources of systematic errors are of the most involved origin anf can be extracted only after very intricate crosschecks. One of this sources is, however, quite evident – in many cases systematic errors are estimated from the difference observed between MC modelled and experimental data sets. Here again we have to remember that the linear model forms the basis of both such data sets, hence the extracted difference is in fact the difference between the MC modelling systematic error and the experimental data systematic error.

7 Some Hardware Sources of Systematic Errors

The main hardware source of systematic errors is due to data transmission channels between detectors and data acquisition systems, between front-end electronics and on-line processing units etc. The produced systematic errors can be corrected here by means of the so-called error-correcting codes effectively used for writing and reading information on compact discs, in transmitting information in space communication lines etc.

Another set of the potential systematic error sources is associated with improperly accounted parameters of experimental setup. For example, when using a target with a large aspect (length/diameter) ratio irradiated with a high energy diverging particle beam and measuring the beam flux by means of a monitor detector positioned at the target front end one will systematically overestimate the real beam flux, thus producing systematic errors in particle yields, production cross-sections and other flux-associated values.

In short, the hardware sources of systematic errors are crucially dependent on the specific structure of the relevant setups and need to be considered on the more down-to-earth basis.

8 Conclusions

To sum up, it is possible to formulate the following general conclusions:

1. In addition to the reversible arithmetic [7] the reversibility axiom allows to develop the deterministic theory of errors - lathology.

2. As opposed to the Gaussian probabilisitic theory of errors, dealing only with additive statistical errors, the lathology can deal with errors of arbitrary origin without any limiting assumptions concerning their distribution and/or underlying statistics.

3. The transition from the standard mixed vector-matrix form of linear model computed within a real number basis to its all-matrix one computed within an integer number basis allows to solve unfolding problems with arbitrary input errors and without any regularization imposed.

4. The preservation of close and simple analytical links between the old and new linear models allows their cross-checks to be easily performed.

5. The true reversible solutions are assumed to be devoid of any errors except of those associated with numerical basis and computer's *macheps*.

6. The problems to be solved deal with extensions of the reversible mathmetics into different application domain, e.g. LSQ.

7. The reversibility axiom results not only in a novel arithmetic but also in the deterministic theory of errors – lathology – and the robust techniques

for solving LSQ, SLAE and many other fundamental problems of data processing.

9 Acknowledgments

I wish to thank those people who contributed to the development of the theory of errors for their deep insight into the problem "To err is human...". The generous support provided by Prof.A.I.Malakhov within the SPHERE collaboration is to be also acknowledged.

References

[1] Review of Particle Properties,
Phys. Rev., 50D, N 3, 1175 f., (1994).

[2] Ilyushchenko V.I.,
Reversible Mathematics - The Second Unfolding Solution and the Source of Unfolding Instability,
JINR Preprint, JINR, Dubna (1995) (to be published).

[3] Ilyushchenko V.I.,
Numerical Differentiation and Integration of Noisy Model Data,
JINR Preprint E10-90-542, JINR, Dubna (1990).

[4] Gauss C.F.,
Theoria Combinationis Observationum Erroribus Minimis Obnoxiae (1823), in vols.I,II,IV of Werke, I-XII, Georg Olms Verlag, Hildesheim (1973).

[5] Buhler W.K.,
Gauss (A Biographical Study),
Springer, Berlin (1981).

[6] Cayley W.,
Collected Mathematical Papers,
University Press, Cambridge (1889-1898).

[7] Ilyushchenko V.I.,
Reversible Arithmetic - Addition, Subtraction, Multiplication and Division,
JINR Preprint E10-94-450, JINR, Dubna (1994).

Received by Publishing Department
on June 26, 1995.