объединенный
институт
ядерных
исследований
дубна

E10-93-403

V.I.Ilyushchenko

THE NOVEL APPROACH TO SOLVING SYSTEMS OF LINEAR ALGEBRAIC EQUATIONS (SLAE)

Submitted to journal «Inverse Problems»

1993

1 Introduction

System of linear algebraic equations (SLAE) is a fundamental mathematical object both in linear algebra and numerical mathematics at large. In linear algebra the SLAE is the basis of the first and second main problems, in numerical mathematics it is the basic computational model, to which can be reduced the whole realm of integral, differential, nonlinear etc. equations, i.e. main analytic tools of modern mathematics and physics.

On the other hand, it is possible to formulate direct and inverse SLAE problems, thus introducing a structural subdivision within the SLAE solution itself. The direct SLAE problem is traditionally considered to be solvable and stable, the inverse SLAE problem is often prone to inherent instabilities due to multiple known and unknown sources of experimental and computational errors, both statistical and systematic.

Here we will present the outline of the novel approach in solving the SLAE with arbitrary initial errors. All the crucial points of the novel approach have been verified via numerous tests with different input samples by means of dedicated computer codes.

2 State of the art

The standard direct problem for the SLAE with noisy input data and an additive noise (error) model can be formulated as

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

where A — the coefficient (apparatus) $(m \times m)$ -matrix, t — the true solution column vector, f — the input data column vector and n — the additive noise (error) column vector. All the vectors can be considered as $(m \times 1)$-matrices.

On the other hand, the standard inverse problem is viewed mathematically as

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

where A^{-} is an inverse of the A matrix. Instabilities in the t-solution (2) arise due to errors inherent in both A and f. The essential qualitative point here concerns relatively small errors in the "direct" matrix A which transform into relatively large errors in the inverse matrix, A^{-}.

One of the important phenomenological points is the fact that the direct problem corresponds to integration, while the inverse one corresponds to differentiation. In digitized forms, when the digitization argument interval becomes to be infinitely small, the numerical integral converges to a finite value, while the numerical derivative becomes infinitely large, i.e. it diverges.

The inverse problem can be solved in the original Fourier space by means of different techniques from Gauss Elimination Method (GEM) to Singular Value Decomposition (SVD) within the floating point arithmetic (FPA) basis. An alternative solution can be obtained in the Fourier image space by using Fast Fourier Transform (FFT). Unfortunately, the latter is subject to the same kind of instabilities as the object of solution, i.e. an initial SLAE.

Another crucial point concerns the standard solution technique used on computers, i.e. local optimization. In solving the inverse problem (2) one usually tries to find out a minimum of the following objective function

$$F = (At - f)^2 = n^2 = MIN! \tag{3}$$

or

$$U = \frac{(At - f)^2}{n^2} = MIN! \tag{4}$$

corresponding to statistical weights $W = 1$ and $W = 1/n^2$, respectively. Here the general assumptions are as follows:

1. The additive noise is the only error source.
2. The matrix A is known exactly, i.e. without any errors.
3. The objective functions, i.e. F and U are smooth and unimodal, i.e. their local minima coincide with the global ones.

Unfortunately, all the above general crucial assumptions concerning statement of the problem, error model and mathematical properties of the relevant objective functions are improper to some or another degree, thus leading to catastrophic instabilities in trying to solve (2).

3 The novel statement of the inverse SLAE problems

Our novel approach involves the formulations of two inverse problems in the all-matrix formulation instead of the single standard one (2) in the mixed matrix-vector form. The first inverse SLAE problem is formulated like (2)

$$T = A^{-}(F + N) \tag{5}$$

while the second inverse SLAE problem looks like

$$A = (F + N)T^{-} \tag{6}$$

where T^{-} is a T — inverse. In case of the mixed matrix-vector formulation the second inverse problem can be written down as

$$A = (f + n)t^{-} \tag{7}$$

However, as opposed to the first inverse problem (2) here we deal with a Diophantine underdetermined problem. Thus, at the very beginning, in stating the inverse SLAE problems, we discover that the coefficient matrix A in the mixed matrix-vector formulation (1) is nonunique. This leads, in particular, to the noncontrollable behaviour of the A^{-} due to "negligibly small" rounding-off errors in the A matrix within the FPA basis presentation.

Moreover, the underdeterminated nature of the A matrix allows us to conclude that, in the mixed matrix-vector formulation, even the direct SLAE problem (1) is stated improperly, i.e. it can be solved only in an unstable (nonunique) way.

This being true, the relevant first (5) and second (6) inverse problems must be improper and unstable so to say from the very origin.

4 Definition of the SLAE solution and the so-called PTT-regularization technique

The standard inverse problem (2) with noisy input data is known to be solved by the PTT-regularization technique [1–3] where the minimized objective function is defined as

$$F = \|At - f\|^2 + \alpha\|t\|^2 = R + \alpha N = MIN! \tag{8}$$

with $\| \quad \|^2$ to stand for Euclidean metrics and norms and α being the regularization parameter. However, computational experience shows this technique to be valid only up to relative errors $e(f) \leq 1\%$.

From our point of view, here we deal with a mathematically improper formulation, when the vector objective function $F = F(R, N)$ is improperly transformed into a scalar-like object, $F = R + \alpha N$.

The proper formulation must look like

$$F_1 = \|AT - F\|_p = MIN! \tag{9}$$

$$F_2 = \|T\|_p = MIN! \tag{10}$$

where p is the index of any Holder norm able to produce robust statistics and robust final results.

5 Local vs global optimization

The objective functions like (7–9) are smooth, monotone and differentiable, in short, analytic ones, only in the absence of errors. The introduction of errors $e(f) \geq 1\%$ results in fractal-like function patterns specified by multiple discontinuities, i.e. multiple local minima. Any real objective function becomes multimodal.

The second reason for this implicit unimodal-multimodal transform is the generally noncritical use of the weighted least squares technique (WLS). The analysis of two-parametric Holder norm patterns with $p = \infty, 2$ and 1 shows geometric figures like an outer square, an inscribed circle and an inner inscribed square, respectively, while the WLS corresponds to an innermost inscribed ellipse. A part of this latter is located inside the $p = 1$ square, thus corresponding to $p < 1$. The objective functions in the form of the Holder $p < 1$ metrics are known to be multimodal by definition.

Thus, any attempt to solve the inverse SLAE problem (2) in the standard mixed matrix-vector formulation is doomed to failure from pure mathematical considerations.

The only constructive remedy to this situation is a transfer to global optimization [4–5].

6 Error models

In solving unstable inverse SLAE problems we do not know a priori what is the specific error (noise) model we need. The trivial assumption about this to be an additive one is convenient only from the WLS point of view. However, this model is based on the fundamental hypothesis about the spectral part (f) and the noise part (n) to be statistically independent. The discretization process of the analyzed data sample, however, introduces a very pronounced correlation between the spectral and noise parts at bin level.

By acting within a trial-and-error model, here again it is possible to introduce a few error models instead of the single standard (additive) one.

Let us consider the additive error model as the first one. Then the second error model can be multiplicative, so that the relevant objective function will look like

$$F_1 = \|lnAT - lnF\| = MIN! \qquad (11)$$

$$F_2 = \|lnT\| = MIN! \qquad (12)$$

The third error model can be formulated as

$$F_1 = \|exp(AT) - exp(F)\| = MIN! \qquad (13)$$

$$F_2 = \|expT\| = MIN! \qquad (14)$$

corresponding to the hypoadditive noise source.

7 Structure of the A-matrix

Let us consider an unstable inverse SLAE problem for a spectral reconstruction case, when, e.g. a particle spectrometer is irradiated by some particle source. The general form of the A-matrix can be viewed as

$$A = A_i \cdot A_r \cdot A_a \cdot A_u \qquad (15)$$

where A_i — an identification factor, A_r — a resolution factor, A_a — an acceptance (registration efficiency) factor and A_u — a factor accounting for unknown information about the whole system composed of a particle source and a particle spectrometer.

In most practical cases presently is used the trivial version corresponding to $A = A_r$ and $A_i = A_a = A_u = I$, where I is the diagonal identity matrix. This standard oversimplification results in a distorted structure of both direct and inverse matrices.

8 Error-free solutions

Let us consider the all-matrix SLAE form

$$AT = F + N \qquad (16)$$

with all the previously noted mathematical defects eliminated. In computing the relevant unstable inverse problem within the FPA basis the only remaining **destabilizing** factor will be the computer rounding-off error. Its catastrophic action is easily identified qualitatively in analyzing the process of the $A \to A^-$ transform supported by practical computations.

The known palliative measures like the sophisticated control of singular value spectrum in SVD technique, the FFT with the subsequent intricate separation of spectral and noise components etc are followed only by a partial success.

The drastic change in solution stability is due to the transfer into integer modular arithmetic (IMA) basis or integer nonmodular arithmetic (INA) basis [6].

8.1 IMA basis

The transfer from the infinite real number field Q to a finite (modular) integer number field $GF(m)$ can be performed in an easy-to-do way. However, even for relatively small $m \times m$ -matrices A with $m > 3$ inverted on IBM computers in the single precision mode of operation one is soon limited by the insurmountable problem of overflow. An increase in the precision of numbers does not lead to any substantial progress.

The only remedy seems to be the transfer to a factorized modulus

$$m = m_1 \cdot m_2 \cdot \ldots \cdot m_i \tag{17}$$

and the use of the Chinese Remainder Theorem at the reconstruction step.

Here again arises a fundamental disadvantage of the IMA basis due to the implicit improper reduction process for some partial moduli m_i. The search for a proper set of m_i transforms into an iteration process with a priori unknown outcome.

Thus, we need an integer basis devoid of above intrinsic disadvantages.

8.2 INA basis

This can be done by using the well-known Hermite and Smith Normal Forms, HNF and SNF, respectively.

The computation with the HNF and SNF algorithms can be arranged in such a way as to minimize the overflow problem in principle. On the other hand, the final SNF is usually obtained in a diagonal form with automatic multiple computations of the most important matrix parameter, i.e. the matrix rank $R(A^-)$. This latter cannot be computed exactly within the FPA basis even by means of such a sophisticated and intricate technique as SVD and its modifications.

The principal point here concerns the transfer from the infinite real number field Q to the finite integer number ring R, with accompanying division problems. However, the thorough analysis demonstrates the absence of division operations in the course of matrix inversion within R. In solving the SLAE problem the final step of matrix multiplication can be easily performed within the final FPA (rational) basis.

Such a solution technology avoids both the IMA defects and the complications inherent in p-adic versions. The slightly increased computation time in the INA basis as compared to the FPA one is favorably compensated by the overall stability of solution process with arbitrary input errors.

9 Statistical vs systematic errors

Our interest is solving the unstable inverse problems have been provoked by specific experimental problems, in particular, those encountered in the EMC-NA4 studies of deep inelastic muon scattering [7]. Here the involved magnetic spectrometers have been located one after another in a high energy muon beam. The final results gained with both spectrometers have pertained to the same object — the so-called structure functions. However, with statistical errors $e(f) \approx 1\%$, the systematic ones reach $s(f) \approx 10 - 50\%$. This situation is quite typical, as can be seen from the summary data plots published in [7].

On the other hand, there are no published data on the apparatus matrix A not only for the above EMC-NA4 twin spectrometers, but for any other either. The above indicated nonuniqueness of A-matrix makes these data to be useless within the standard mixed matrix-vector formulation of both direct and inverse SLAE problems.

We will not consider here particular problems of vector-matrix transform essential for the transfer from the old mixed matrix-vector form to the novel all-matrix one to be published elsewhere [8].

10 Conclusion

The highly traditional theme of solving SLAE could not be critically reconsidered without the implicit support provoked by the works of such mathematicians as F. Klein and C. Lanczos. Multiple discussions with staff members of the JINR High Energy Lab and other JINR departments are highly appreciated.

References

[1] Phillips D.L., *J.ACM*, 9, 1 (1962).

[2] Twomey S., *J.ACM*, 10, 1 (1963).

[3] Tikhonov A.N., *DAN SSSR* 151, 3 (1963) (in Russian).

[4] Torn A. and Zilinskas A., Global Optimization, *Springer*, Berlin (1989).

[5] Ilyushchenko V.I., *Preprint JINR E10-90-410*, Dubna (1990).

[6] Gregory R.T. and Krishnamurthy E.V., Methods and Applications of Error-Free Computation, *Springer*, New York (1984).

[7] The Review of Particle Properties, *Phys. Rev. D45*, 11, III — 75 (1992).

[8] Ilyushchenko V.I., *Preprint JINR* (to be published).

Received by Publishing Department
on November 5, 1993.

Илющенко В.И. E10-93-403

Новый подход к решению систем линейных алгебраических уравнений (СЛАУ)

Стандартный подход определяет проблему для систем линейных алгебраических уравнений (СЛАУ) в рамках смешанной матрично-векторной формулировки и действительного арифметического базиса как $At = (f+n)$, где A — $m \times m$-матрица и t, f, n — столбцовые векторы истинного решения, входных данных и ошибок входных данных соответственно. Обратная СЛАУ-проблема, которая обычно является неустойчивой, формулируется как $t = A^{-}(f+n)$, где A^{-} является обратной матрицей для А. Наиболее существенный вычислительный параметр, а именно ранг $R(A^{-})$, здесь не может быть вычислен в принципе. Новый безошибочный подход использует полноматричную формулировку и целочисленный (модулярный или немодулярный) базис для решения прямой проблемы $AT = F+N$, а также первой обратной проблемы $T = A^{-}(F+N)$ и второй обратной проблемы, $A = (F+N)T^{-}$, при этом соответствующие ранги матриц вычисляются точно в течение всего процесса вычислений.

Работа выполнена в Лаборатории высоких энергий ОИЯИ.

Препринт Объединенного института ядерных исследований. Дубна, 1993

Ilyushchenko V.I. E10-93-403

The Novel Approach to Solving Systems of Linear Algebraic Equations (SLAE)

The standard approach specifies the direct problem for systems of linear algebraic equations (SLAE) within the mixed matrix-vector formulation and a real arithmetic basis as $At = (f+n)$, where A — $m \times m$-matrix and t, f, n — column vectors of a true solution, input data and input data error, respectively. The inverse SLAE problem (generally unstable) is formulated as $t = A^{-}(f+n)$, where A^{-} is an inverse of A. The crucial computational parameter, i.e. the rank $R(A^{-})$, cannot be computed here. The novel error-free approach uses the all-matrix formulation and integer (modular and nonmodular) arithmetic bases to solve the direct problem $AT = F + N$ as well as the first, $T = A^{-}(F+N)$ and second, $A = (F+N)T^{-}$, inverse problems, with the continuous rank crosscheck during the whole computation process.

The investigation has been performed at the Laboratory of High Energies, JINR.

Preprint of the Joint Institute for Nuclear Research. Dubna, 1993