

Use of Regularization Method in the Determination of Ring Parameters and Orbit Correction *

Y.N. Tang and S. Krinsky

National Synchrotron Light Source, Brookhaven National Laboratory, Upton, NY 11973

Abstract

We discuss applying the regularization method of Tikhonov to the solution of inverse problems arising in accelerator operations. This approach has been successfully used for orbit correction on the NSLS storage rings, and is presently being applied to the determination of betatron functions and phases from the measured response matrix. The inverse problem of differential equations often leads to a set of integral equations of the first kind which are ill-conditioned. The regularization method is used to combat the ill-posedness.

1 Introduction

Inverse problems of differential equations, which determine the physical properties (*e.g.* betafuncions, phases) and causes (corrector strengths) from measured data (orbit displacements) have wide applications in many fields. Unfortunately, these problems are inherently ill-posed computationally [3]. Their solutions do not smoothly depend on the input data. Small errors in the input data will cause big changes in their solutions. If we solve them directly on finite precision computers without applying special techniques and if the dimension of the problem exceeds some limit, oscillations will occur and the results may diverge if under iteration.

This paper gives a simple illustration of why the inverse problem is ill-posed, and introduces the regularization method, which is used in the orbit correction of NSLS storage rings and a simulation model for determination of ring parameters from the measured local bump ratios.

2 Inverse Problems and Ill-Posedness

To study a dynamic system, people solve differential equations. The motion of a particle in the storage ring is governed by a second-order differential equation [1]. If the coefficients (*i.e.* the system parameters) of the equation and the driving force satisfy certain conditions, finding its solution is well-defined and conditioned. To solve a differential equation is to integrate the effects of all the sources. However, the reverse process of finding the distributed sources from the cumulative effects is a totally different matter. The immediate question is the uniqueness of its solution

*Work performed under the auspices of the U.S. Dept. of Energy under contract no. DE-AC02-76CH00016.

because the effects caused by some sources could cancel each other. Generally speaking, the inverse problem of differential equations leads to integral equations [3]. This is expected because the differentiation and integration are inverse operations. However, to solve the produced integral equations is far more difficult than solving the original differential equations.

Let's take a simple example of solving the inverse problem of the following equation:

$$\frac{d^2x}{ds^2} + K(s)x = F(s) \quad (1)$$

with some boundary or initial conditions, where $a \leq s \leq b$. The inverse problem is to determine $K(s)$ from the known solution $x(s)$, $a \leq s \leq b$.

It is not possible to find analytical solutions of inverse problems in most of the cases. In practice, the following iterative procedure is employed.

Let

$$K_{n+1} = K_n + \delta K_n \quad \text{and} \quad x = x_n + \delta x_n \quad (2)$$

be the $(n+1)$ th iterates of K and x . K_0 is the initial guess. Substituting Eq. (2) into Eq. (1) and neglecting terms of $O\{\delta x_n \delta K_n\}$, one obtains

$$\frac{d^2x_n}{ds^2} + K_n(s)x_n = F(s) \quad (3)$$

and

$$\frac{d^2(\delta x_n)}{ds^2} + K_n(s)\delta x_n = -x_n \delta K_n(s) \quad (4)$$

By applying the Green function to Eq. (4),

$$\int_a^b G_n(s, s')(-x_n(s')\delta K_n(s'))ds' = \delta x_n(s) \quad (5)$$

Eq. (5) tells that what relates δK_n to δx_n is a Fredholm integral equation of the first kind, which is notorious for its ill-posedness. It presents a great challenge to obtain $\delta K_n(s)$ from the known $\delta x_n(s)$, especially in the case when $\delta x_n(s)$ is derived from the measured data and inevitably has errors.

To make description simple, let's rewrite Eq. (5) in a more general form

$$\int_a^b G(s, s')\theta(s')ds' = p(s), \quad c \leq s \leq d \quad (6)$$

where $p(s)$, $c \leq s \leq d$ is the known function and $\theta(s)$, $a \leq s \leq b$ is unknown. Suppose $\theta_1(s)$ is a solution of Eq. (6). Let

$$\theta_2(s) = \theta_1(s) + N \sin ns \quad (7)$$

where N and n are arbitrary numbers. Using Eq. (6) one finds

$$\int_a^b G(s, s')(\theta_1(s') + N \sin ns') ds' = p(s) + N \int_a^b G(s, s') \sin ns' ds'$$

Theoretically, in order to make $\theta_2(s)$ the solution of Eq. (6), the following equation must hold,

$$N \int_a^b G(s, s') \sin ns' ds' = 0, \quad c \leq s \leq d \quad (8)$$

On a computer with finite precision, Eq. (8) always holds for sufficiently large n .

Suppose we solve Eq. (6) numerically on a grid $s_i, i = 1, 2, \dots, M$, where $s_1 = c$ and $s_M = d$. By using a well-known theorem,

$$\lim_{n \rightarrow \infty} \int_a^b f(s) \sin ns ds = 0, \quad (9)$$

it is easy to show that we can make $N \int_a^b G(s_i, s') \sin ns' ds'$ arbitrarily small for $i = 1, 2, \dots, M$ if n is large enough. If they are smaller than the precision of the digital computer on this grid, $\theta_2(s)$ becomes a solution of Eq. (6).

Note that it is very easy to apply the previous analysis to the orbit correction problem and obtain the same conclusion. As a matter of fact, the analysis is much simpler. The response matrix in the orbit correction is the discretized form of the Green's function.

3 The Regularization Method

The regularization method was devised and developed by Tikhonov et al. [2] for the purpose of reformulating ill-posed problems into problems of a more well-posed nature. The revised problems do not have the same exact solution as the original problems, but in most practical cases, the data is inexact and thus no method can possibly extract the exact solution. However, as the errors in the original posed problem tend to vanish, the regularized solutions are designed to converge to the exact solution. The convergence can be weak, uniform, or of higher orders depending upon the order of regularization chosen.

The regularization method uses stabilizers to form a new regularizing operator to replace the original ill-posed operator. Different stabilizers produce different regularizing operators. Which type of stabilizers to use depends on the nature of the ill-posed problem. This method is widely used to solve many kinds of ill-posed problems such as the integral equations of the first kind and of the convolution type, optimum control, linear algebraic equations *etc.* Tikhonov used functional analysis and gave this method a rigorous mathematical deduction and proof. To introduce it in detail goes far beyond the scope of this paper. What

interests us here is how to use it to solve the ill-posed linear algebraic system:

$$A\theta = x. \quad (10)$$

If we use $\|\theta\|_2^2$ as the stabilizer, the regularization method minimizes the following functional

$$\|A\theta - x\|_2^2 + \alpha \|\theta\|_2^2 \quad (11)$$

where $\|x\|_2 = (\sum x_i^2)^{1/2}$ stands for the geometric norm (or 2-norm) of the vector x , and α is a positive constant and called the regularization parameter.

After some manipulations, the new equation used in the regularization is

$$(A^T A + \alpha I)\theta = A^T x, \quad (12)$$

where A^T is the transpose of matrix A and I is the unit matrix.

Note: If we use different stabilizers, we will get different equations. However, we want to minimize the corrector strengths in our case and hence choose $\|\theta\|_2^2$ as the stabilizer.

The matrix $A^T A$ in Eq. (12) is non-negative symmetric and its eigenvalues are non-negative. If we add a positive constant α to its diagonal, its eigenvalues will be greater or equal to α . If we select suitable α , the system of Eq. (12) becomes well-conditioned.

How to select α is very important in the real computations. The bigger the α is, the more stable Eq. (12) becomes and the farther the regularized solution is from the true solution. However, if the α is chosen too small, Eq. (12) would not be well-conditioned. If α equals zero, Eq. (12) becomes the normal equation of the least square method, which provides a way to solve an overdetermined or underdetermined linear system but does not address or control the ill-posedness of the problem.

4 Orbit Correction

In a circular machine, the orbit change due to a change in the corrector strengths is expressed as

$$A\theta = x. \quad (13)$$

where $\theta = [\theta_i], 1 \leq i \leq N_c, x = [x_j], 1 \leq j \leq N_m$, and matrix A is the $N_m \times N_c$ response matrix, N_m and N_c are number of monitors and correctors, respectively. Eq. (13) is widely used in the orbit correction.

Eq. (13) could be overdetermined ($N_c > N_m$), determined ($N_c = N_m$) or underdetermined ($N_c < N_m$). The least square method yields

$$A^T A\theta = A^T x. \quad (14)$$

Theoretically, Eq. (14) may have one solution or infinitely many solutions.

According to the analysis in the previous section, Eq. (14) is not well-conditioned. Since the dimension of matrix

$A^T A$ is N_c by N_c , how ill-posed is Eq. (14) mainly depends on how large N_c is. When doing study on our VUV and X-ray rings, it was found that in many cases if no more than some numbers of correctors were used, we could solve Eq. (14) directly. Once we used more than this number of correctors, bad oscillation started to occur and Eq. (14) started to give unreasonably large kick values. The more correctors were used, the worse the results became.

In order to combat the ill-posedness, several algorithms have been developed in the accelerator community.

- The Micado method [4] selects N (normally 2 to 4 in our operations) “most effective” correctors. That is, it does not solve the whole Eq. (14). Instead, it selects only a small number N equations in Eq. (13) and the dimension of the resulted normal equation Eq. (14) is N (2 to 4 in our case). Because the number of correctors used are very small, the ill-posedness is effectively controlled.
- Recently, the singular value decomposition (SVD) method has entered the orbit correction field [6, 7]. The SVD method is very powerful tool to solve least square problems. It not only diagnoses the problem, it also solves it. If the dimension of matrix A is large, its singular value set has very small numbers and possibly zeros. The SVD method simply ignores these small and zero singular values to control the ill-conditionedness.

The regularization method has been used successfully in the orbit correction in the NSLS storage rings both for study and operation.

According to the analysis in the previous section, This method solves the equation

$$(A^T A + \alpha I)\theta = A^T x, \quad (15)$$

instead of solving Eq. (14).

The key to a successful execution of the regularization method is choosing a good α , which depends on the degree of the ill-posedness of the problem and also depends on the magnitude of the elements in the matrix A . For example, suppose we selected a very good α to solve Eq. (15) and suppose now we use different units to measure the response matrix A and all elements in A become 10 times in magnitude of the original elements. The elements in the matrix $A^T A$ will be 100 times of the original values. In order to solve the new Eq. (15), the α has to be chosen 100 times larger.

In practice, we simply use the trial-and-error method to find the best α . At first, we pick a typical number in the matrix A and divide it by 3 or 4 and use the quotient as the initial guess of α . Normally, after a few trials, the α could be determined. Fortunately, once a good α is chosen, it could be used in successive sessions unless the lattice of the accelerator changed. Generally speaking, the resulting θ values are not very sensitive to α if it is not chosen too small. For example, when doing orbit correction on the

VUV ring, we use 0.2 as α for the vertical plane and 0.25 for the horizontal. While working on the X ray ring, 0.1 is used for both the vertical and horizontal planes. However, if we use 0.05 or 0.2 as α for the X ray ring, the results do not differ much. Let’s take a case in the horizontal correction (55 correctors were used) of the X ray ring as an example. If $\alpha = 0.05$, the corrector kicks ranged from 3 to 460 digits, and the RMS difference after correction was 0.0605mm; if $\alpha = 0.10$, kicks 1 to 291 and RMS 0.0675; if $\alpha = 0.20$, kicks 3 to 185 and RMS 0.0750. Zero α gave unreasonably large kick values. The SVD method with optimization gave similar results. However, the Micado and harmonic methods were able to reduce the RMS to only 0.10 and 0.16, respectively. We have many similar data sets.

The regularization method not only makes the orbit correction process more stable, but also minimizes the corrector strengths. All the available correctors are used. The results show that it uses much smaller corrector strengths than the traditional methods and gives very good corrections. Generally speaking, The results generated by the regularization method are comparable with the SVD method with optimization, and better than the harmonic [5] and Micado method.

5 Concluding Remarks

We have discussed applying the regularization method of Tikhonov to the solution of inverse problems arising in accelerator operations. This approach has been successfully used for orbit correction on the NSLS storage rings. We have obtained high precision correction with weak corrector strength, eliminating the “fighting” of corrector against one another. In this case, the regularization method corresponds to a least square minimization of the orbit deviations, with an additional constraint to minimize a weighted measure of the corrector strengths.

At present, we are applying this method to the determination of betatron functions and phase from the measured response matrix

$$A_{ij} = \sqrt{\beta_i \beta_j} \cos \nu (|\phi'_i - \phi_j| - \pi) \quad (16)$$

Preliminary results seem promising and we hope to report on this work in the future.

I. REFERENCES

- [1] E.D. Courant and H.S. Snyder, *Annals of Phys.*, **3**, No. 1, p. 1–48, (1958).
- [2] A. N. Tikhonov *etc*, Solutions of Ill-Posed Problems, John Wiley and Sons, New York (1977).
- [3] Y. M. Chen, Inverse Problems, Lecture notes, SUNY at Stony Brook, New York (1984).
- [4] B. Autin and Y. Marti, CERN ISR-MA/73-17 (1973).
- [5] A. Jackson, SRS/NS/76/103 (1976).
- [6] E. Bozoki and A. Friedman, In this proceeding.
- [7] Y. Chung, G. Decker, In this proceeding.