Multi-step Method for Resolving Ill-conditioned Problem of Gauss-markov Model

Minwei Jiang
Department of Surveying and Mapping Engineering
School of Earth Sciences and Engineering
Hohai University
Nanjing, China
jmw121@163.com

***Abstract*—Toward the ill-conditioned Problem of Gauss-Markov model, analyze the causes and all kinds of methods, emphatically research the computational process of Two-Step method. Extending and simplifying Two-Step method, give a new method for solving ill-conditioned problem, multi-step method. Detailedly expound the computational process of it. Simulate an ill-conditioned model and compute its parameter estimations by some different methods. Finally through analog computation, verify the feasibility and quality of multi-step method.**

Keywords-multi-step method, ill-conditioned problem, Gauss-Markov model, Two-Step method, regularization theory

I. INTRODUCTION

Ill-conditioned Problem in data processing is drawn attention over a long period of time. Ill-conditioned problem is in relation to well-conditioned problem. A parametric model is ill-conditioned if its parameter estimations have great change with data's subtle change. On the contrary, it is well-conditioned. Gauss-Markov model [1] is the base parametric model used to research ill-conditioned problem. The following is Gauss-Markov model:

$$\begin{cases} L = AX + \Delta \\ E(\Delta) = 0, D(\Delta) = \sigma_0^2 P^{-1} \end{cases} \quad (1)$$

where L is n-dimensional observed data vector, A is a design matrix with n rows and m columns, rank(A)=m, X is n-dimensional unknown parameter vector, Δ is n-dimensional observation error vector, $\Delta \sim N(0, \sigma_0^2 I)$, σ_0^2 is variance of unit weight, and P is weight matrix. For formula (1), $\hat{X}$ the least squares estimation (LSE) of Gauss-Markov model can be easily gotten:

$$\hat{X} = N^{-1}W = (A^T PA)^{-1} A^T PL \quad (2)$$

where:

$$N = A^T PA$$
$$W = A^T PL$$

According to Gauss-Markov theorem, on the premise that observed data obey the law of normal distribution, $\hat{X}$ is the best linear unbiased estimator of X. For this reason, the least squares estimate is a current important estimation method. However, in practice, because of parameter selection, arrangement of the observation, computational method, and so on, LSE is instable and unreliable.

In formula (2), when N is nonsingular, $\hat{X}$ is unique. Consider a case that N and W have small perturbation, δN and δW, $\hat{X}$ will have a deviation, $\delta\hat{X}$. It may be expressed as:

$$\hat{X} + \delta\hat{X} = (N + \delta N)^{-1}(W + \delta W) \quad (3)$$

According to the corollary to lemma banach, $N + \delta N$ should be always nonsingular when $\|N^{-1}\|\|\delta N\| < 1$. Obviously, influence on $\delta\hat{X}$ caused by δN and δW is connected with N^{-1}. Therefore, through changing N, normal matrix, illness of Gauss-Markov model can be weakened or eliminated.

II. TWO-STEP METHOD

There are many methods changing normal matrix. Zhengjie Wang proposed a new method, Two-Step method [2]. It weakens the illness of Gauss-Markov model by changing normal matrix twice. Two-Step method derives from Tikhonov regularization [3]. The following is its estimation criterion:

$$\left\|A\hat{X} - L\right\|^2 + \alpha\Omega(\hat{X}) = \left\|A\hat{X} - L\right\|^2 + \alpha\hat{X}^T R\hat{X} = \min \quad (4)$$

where α is regularization parameter, R is regularization matrix, $\Omega(\hat{X})$ is functional stability and $\|\bullet\|$ is 2-norm.

Essentially, the first step of tow-step method is ridge estimation. R_1, regularization matrix, is unit matrix. By Tikhonov regularization and formula (4), parameter estimation and its mean square error matrix [4] may be gotten:

$$\hat{X}_1 = (A^T A + \alpha_1 R_1)^{-1} A^T L \quad (5)$$
$$MSEM(\hat{X}_1) = \hat{\sigma}_0^2 (A^T A + \alpha_1 I)^{-1} \quad (6)$$

In formula (5) and formula (6), α_1 regularization parameter is obtained by L-curve method [5], and $\hat{\sigma}_0$ is variance of unit weight.

The second step of Two-Step method is generalized ridge estimation in fact. In this step, R_2, regularization matrix, is not unit matrix but the matrix obtained by the solution of the first step. By Tikhonov regularization matrix and formula (4), parameter estimation, X_2, may be gotten:

$$\hat{X}_2 = (A^T A + \alpha_2 R_2)^{-1} A^T L \quad (7)$$

In formula (7), α_2 is obtained by L-curve method.

where:

$$R_2 = diag(MSEM(\hat{X}_1)^{-1}) = diag(\frac{1}{\hat{\sigma}_0^2}(A^T A + \alpha_1 I))$$

It deserves noting that above solution is obtained when P is unit matrix. If P isn't unit matrix, it need be translated to unit matrix [6]. In the second step, for R isn't unit matrix, it need be translated when calculating α_2 by L-curve method [7].

Though analyzing the computational process of two-step method, it can be seen that its second step is based on its first step, optimizes its normal matrix, N, and gets better parameter estimations. Therefore, it may go deeper and further optimize N though similar computation.

III. MULTI-STEP METHOD

Based on two-step method, multi-step method is a solution of solving ill-conditioned problem through optimizes normal matrix many times. Its parameter estimations meet special required precise. Here, the number of computational times depends on the required precise. The number may be two or greater than two. The following is the Computation flow process of multi-step method:

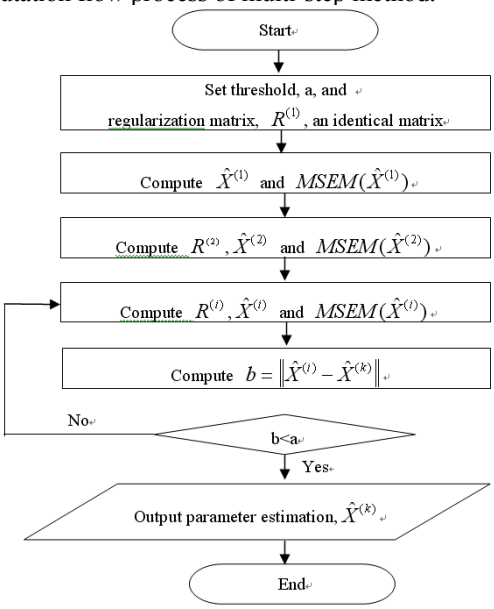

Figure 1. Computation flow process of multi-step method

In Fig.1, $\|\bullet\|$ is infinite norm, b is the maximum in the absolute values of the differences between two adjacent parameter estimations, and $\hat{X}^{(k)}$ is the final estimations.

$$\hat{X}^{(k)} = (A^T A + \alpha^{(k)} R^{(k)})^{-1} A^T L \quad (8)$$

Where:

$$R_i = diag(MSEM(\hat{X}_{i-1})^{-1})$$

$$MSEM(\hat{X}^{(i)}) = \hat{\sigma}_0^2 \ (A^T A + \alpha^{(i)} R^{(i)})^{-1}$$

Multi-step method and two-step method is similar. Both of them estimate the parameters by optimizing the norm matrix, based on regularization theory; R_1 is unit matrix and other regularization matrixs are gotten depending on the solutions of the last computation; All regularization parameters of them are gotten by L-curve method. The differences between two method are that: the former has more computation step number and has more precise parameter estimations than the latter, and the former needs a threshold to control the computation process and the latter doesn't need it. It needs to be noticed that the threshold should be suitable. If the threshold is too little, computation steps would be very many or never end. If it is too large, the computation would be finished rapidly, but its accuracy would be too low. Normally, the threshold may be set a required accuracy. For example, if the required accuracy is 0.002m, the threshold may be 0.002m.

Multi-step has the following better charateristics:

- Multi-step method is a linear transformation of LSE.

The following is the certification of above-mentioned charateristic:

Transform the formula (8), and it can be gotten that

$$\begin{aligned}\hat{X}^{(k)} &= (A^T A + \alpha^{(k)} R^{(k)})^{-1} A^T L \\ &= [(A^T A)^{-1} - (A^T A)^{-1}((A^T A)^{-1} \\ &\quad + (\alpha^{(k)} R^{(k)})^{-1})^{-1} (A^T A)^{-1}] A^T L \\ &= [I - (A^T A)^{-1}((A^T A)^{-1} + (\alpha^{(k)} R^{(k)})^{-1})^{-1}]\hat{X} \\ &= [I - (I + (\alpha^{(k)} R^{(k)})^{-1} A^T A)]\hat{X} \\ &= [(I + (\alpha^{(k)} R^{(k)})^{-1} A^T A)^{-1} (\alpha^{(k)} R^{(k)})^{-1} A^T A]\hat{X}\end{aligned} \quad (9)$$

$$E(\hat{X}^{(k)}) = (A^T A + \alpha^{(k)} R^{(k)})^{-1} A^T A X \neq X \quad (10)$$

According to formula (9), after $\alpha^{(k)}$ is sure, obviously $\hat{X}^{(k)}$ is a linear transformation of $\hat{X}$. Moreover, because $A^T A$ is positive definite matrix and $R^{(k)}$ is diagonal matrix whose diagonal elements are positive number, to arbitrary $\alpha^{(k)}$ ($\alpha^{(k)} > 0$) and $\hat{X}$ ($\hat{X} \neq 0$), it can be gotten that

$$(\alpha^{(k)} R^{(k)})^{-1} < I + (\alpha^{(k)} R^{(k)})^{-1} A^T A$$

According the charateristics of matrix and eigenvalue, it can be gotten that

$$\begin{aligned}\|\hat{X}^{(k)}\| &= \|(I + (\alpha^{(k)} R^{(k)})^{-1} A^T A)^{-1} (\alpha^{(k)} R^{(k)})^{-1} A^T A \hat{X}_{LS}\| \\ &\leq \|(I + (\alpha^{(k)} R^{(k)})^{-1} A^T A\| \|(\alpha^{(k)} R^{(k)})^{-1} A^T A\| \|\hat{X}_{LS}\| \\ &< \|\hat{X}_{LS}\|\end{aligned} \quad (11)$$

Formula (10) shows that $\hat{X}^{(k)}$ is biased estimation of $\hat{X}$, as long as $\alpha^{(k)} \neq 0$. Formula (11) shows that $\hat{X}^{(k)}$ is a compression of $\hat{X}$ to the origin.

- The parameter estimation of Multi-step method has higher anti-interference than LSE's.

The following is the certification of above-mentioned charateristic:

According formula (10) and reference[8], there are the following two formulas.

$$K(\hat{X}_{LS}) = K(A^T A) = K(N^{(k)}) \quad (12)$$

$$K(\hat{X}^{(k)}) = K(A^T A + \alpha^{(k)} R^{(k)}) = K(N^{(k)}) \quad (13)$$

Where:

$$N = A^T A$$

$$N^{(k)} = A^T A + \alpha^{(k)} R^{(k)}$$

Through eigenvalue decomposition of N, it can be gotten that

$$N = Q\Lambda Q^T = Q\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_m \end{bmatrix} Q^T \tag{14}$$

Where Q is orthogonal matrix, and

$$Q = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1m} \\ q_{21} & q_{22} & \cdots & q_{2m} \\ \cdots & \cdots & \ddots & \cdots \\ q_{n1} & q_{n2} & & q_{nm} \end{bmatrix}$$

$\Lambda = diag(\lambda_1, \cdots \lambda_m)$, $\lambda_1 \geq \cdots \geq \lambda_m > 0$

Because of $R^{(i)} = diag(MSEM(\hat{X}^{(i-1)})^{-1})$, $R^{(i)}$ is diagonal matrix. Now assume $R^{(i)}$ is that

$$R^{(k)} = \begin{bmatrix} H_1 & & & \\ & H_2 & & \\ & & \ddots & \\ & & & H_m \end{bmatrix} \tag{15}$$

Where H_i is the No.i diagonal element of $R^{(k)}$.

Combining the equation:

$$MSEM(\hat{X}^{(i)}) = \hat{\sigma}_0^2 \ (A^T A + \alpha^{(i)} R^{(i)})^{-1}$$

It can be gotten that:

$$H_i = \frac{1}{\hat{\sigma}_0^2}\left\{ q_{i1}^2[\lambda_1 + \frac{\alpha^{(k-1)}}{\sigma_0^2}(\lambda_1 + \frac{\alpha^{(k-2)}}{\sigma_0^2}(\lambda_1 + \cdots + \frac{\alpha^{(2)}}{\sigma_0^2}(\lambda_1 + \alpha^{(1)})))] + \right.$$
$$q_{i2}^2[\lambda_2 + \frac{\alpha^{(k-1)}}{\sigma_0^2}(\lambda_2 + \frac{\alpha^{(k-2)}}{\sigma_0^2}(\lambda_2 + \cdots + \frac{\alpha^{(2)}}{\sigma_0^2}(\lambda_2 + \alpha^{(1)})))] +$$
$$\left. \cdots + q_{im}^2[\lambda_m + \frac{\alpha^{(k-1)}}{\sigma_0^2}(\lambda_m + \frac{\alpha^{(k-2)}}{\sigma_0^2}(\lambda_m + \cdots + \frac{\alpha^{(2)}}{\sigma_0^2}(\lambda_m + \alpha^{(1)})))] \right\}$$
$$= \left(\frac{1}{\hat{\sigma}_0^2} + \frac{\alpha^{(k-1)}}{\hat{\sigma}_0^4} + \frac{\alpha^{(k-1)(k-2)}}{\hat{\sigma}_0^6} + \cdots + \frac{\alpha^{(k-1)(k-2)} \cdots \alpha^{(2)}}{\hat{\sigma}_0^{2(k-1)}} \right) \sum_{j=1}^{m} q_{ij}^2 \lambda_j$$
$$+ \frac{\alpha^{(k-1)(k-2)} \cdots \alpha^{(2)} \alpha^{(1)}}{\hat{\sigma}_0^{2(k-1)}}$$

Assume the follwing equations

$$\omega = \left(\frac{1}{\hat{\sigma}_0^2} + \frac{\alpha^{(k-1)}}{\hat{\sigma}_0^4} + \frac{\alpha^{(k-1)(k-2)}}{\hat{\sigma}_0^6} + \cdots + \frac{\alpha^{(k-1)(k-2)} \cdots \alpha^{(2)}}{\hat{\sigma}_0^{2(k-1)}} \right)$$

$$\mathrm{T} = \frac{\alpha^{(k-1)(k-2)} \cdots \alpha^{(2)} \alpha^{(1)}}{\hat{\sigma}_0^{2(k-1)}}$$

then

$$H_i = \omega \sum_{j=1}^{m} q_{ij}^2 \lambda_j + T \tag{15}$$

By formula (14) and formula (15), there are those

$$\begin{aligned} R^{(k)} &= Q(\omega\Lambda + \mathrm{T}I)Q^T \\ N^{(k)} &= N + \alpha^{(k)} R^{(k)} \\ &= Q\Lambda Q^T + \alpha^{(k)} Q(\omega\Lambda + TI)Q^T \\ &= Q[\Lambda + \alpha^{(k)}(\omega\Lambda + \mathrm{T}I)]Q^T \\ &= Q[(1 + \alpha^{(k)}\omega)\Lambda + \mathrm{T}I]Q^T \end{aligned} \tag{16}$$

For Λ and I are diagnal matrix, by formula (16), eigenvalues of $N^{(k)}$ is that

$$\lambda_i^{(k)} = (1 + \alpha^{(k)}\omega)\lambda_i + \mathrm{T} \qquad (i = 1,2,\cdots m)$$

Because of $1 + \alpha^{(k)}\omega > 1$ and $\mathrm{T} > 0$, there are

$$K(N^{(k)}) = \frac{(1 + \alpha^{(k)}\omega)\lambda_1 + \mathrm{T}}{(1 + \alpha^{(k)}\omega)\lambda_m + \mathrm{T}} < \frac{\lambda_1}{\lambda_m} = K(N) \tag{17}$$

Formula (17) shows that condition number of normal matrix, $N^{(k)}$, of multi-step method, is less than LSE's. It means that parameter estimation of multi-step method has higher anti-interference than LSE's.

- Multi-step method is better than LSE in the sense of mean square error.

The following is the certification of above-mentioned charateristic:

It is easy to know that the mean square error of $\hat{X}^{(k)}$ is that

$$\begin{aligned} MSEM(\hat{X}^{(k)}) &= \hat{\sigma}_0^2 (N^{(k)})^{-1} \\ &= \hat{\sigma}_0^2 Q[(1 + \alpha^{(k)}\omega)\Lambda + \mathrm{T}I]^{-1} Q^T \end{aligned}$$

Because of $1 + \alpha^{(k)}\omega > 1$ and $\mathrm{T} > 0$, there are

$$\begin{aligned} &MSE(\hat{X}^{(k)}) = tr(MSEM(X^{(k)})) \\ &= \sigma_0^2 \sum_{i=1}^{m} \frac{1}{(1 + \alpha^{(k)}\omega)\lambda_i + \mathrm{T}} < \sigma_0^2 \sum_{i=1}^{m} \frac{1}{\lambda_i} = MSE(\hat{X}_{LS}) \end{aligned} \tag{18}$$

Formula (18) shows that Multi-step method is better than LSE in the sense of mean square error.

IV. Example Analysis

Here design an ill-conditioned problem. There are 3 parameters whose truth values are $X = [10 \quad 15 \quad 6]^T$. Design matrix is that:

$$A = \begin{bmatrix} 1 & 3.6 & 2.4 & 1 & 3.5 & -1 & 5 & 1 & 4 & 3 \\ 2 & 1 & 1.5 & 2 & 1 & 3 & 0.5 & 2 & 1 & 1 \\ 4 & 2.1 & 3 & 3.9 & 2 & 6 & 1.1 & 4.1 & 1.9 & 2 \end{bmatrix}^T.$$

Weighting matrix of observed values, P, is unit matrix. $\Delta \sim N(0, \sigma^2 I)$, $\sigma = 1$. Δ is obtained by random number generator.

For analyzing the charateristics of multi-step method, examine as many computation steps as possible. Here set a very little threshold, $a = 0.00001$, and choose front eight computed results(see Tab.1).

Analyzing those computed results, get some characteristics:

- The No.3 value of $\left\| \hat{X}_i - X \right\|_2$ is obviously less than the No.2 one. Starting with the No.3 step, the value of $\left\| \hat{X}_i - X \right\|_2$ holds the line basically.
- Starting with the No.2 step, the variation of regularization parameter is very little.
- The value of $b = \left\| \hat{X}^{(k)} - \hat{X}^{(k-1)} \right\|$ decreases gradually as computation steps increase to the general trend, and this variation tends towards stability.

- If a=0.01, when the computation is in the No.4 step, b<a; If a=0.002, the computation would come to an end in the No.8 step.

Therefore, there are these following conclusions that

- After multi-step computation, parameter estimation whose variation is relatively steady is optimal approximate solution.
- The accuracy of parameter estimation of multi-step method is higher than two-step method's.
- The computation steps have business with the threshold. The less the threshold is, the more the computation steps are.

V. CONCLUSION

There are many methods of solving the illness of Gauss-Markov model. Some methods are used to change the normal matrix to weaken or eliminate the model's illness. Tow-step method is the newer one that is based on regularization theory and whose accuracy is higher than ridge estimation by two steps of computations. This paper puts forwards a new method, multi-step method that borrows ideas from two-step, improves the accuracy of parameter estimation by more steps of computations, and is the extension of two-step method. This paper researches on multi-step method's characteristics, based on a lot of analog computations. By example analysis, it can be known that multi-step method has better estimation characteristics and is worth going to deeply.

REFERENCES

[1] Jinkui Zhang. Parameter estimation of linear model and its improvement, National University of Defense Technology Press, Changsha, China,1992

[2] Zhengjie Wang, Jikun OU, Lintao Liu." A new method of solving ill-conditioned problem: Two-Step method," Journal of Wuhan Information Science University: Information Science, Vol.30, No.9, 2005, pp: 821~823.

[3] Tikhonov A N, Arsenin V Y. Solutions of Ill-posed Problems. New York: Wiley, 1997

[4] Wencai Yang.Geophysical inversion and seismic tomography, Geol.Press, Beijin, China,1989.

[5] Hansen P C." Analysis of discrete Ill-posed problems by means of the l-curve." SIAM Revies, Vol.34, No.4, 1992, pp:561～580.

[6] Jiangwen Zhou, etal. A new approaching to theory of measurement error. Seismological Press, Beijin, China, 1999.

[7] Hansen P C. Rank-deficient and discrete Ill-posed problems: numerical aspects of linear inversion. Philadelphia: SIAM, 1998.

[8] Songgui Wang, Zhongzhen Jia. Inequality in the theory of matrix. Hefei: Anhui Education Press, 1994

TABLE I. TAB.1 THE RESULTS OF CALCULATIONS FOR MULTI-STEP METHOD

project	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8
$\hat{X}_i$	10.0260	10.0379	10.0383	10.0377	10.0380	10.0378	10.0380	10.0379
	6.0511	13.7100	13.7419	13.7356	13.7392	13.7365	13.7392	13.7374
	10.4504	6.6165	6.6007	6.6036	6.6019	6.6032	6.6019	6.60279
	0.16470	0.00305	0.00301	0.00308	0.00304	0.00307	0.00304	0.00306
$\left\|\hat{X}^{(k)}-\hat{X}^{(k-1)}\right\|$		7.65886	0.03194	0.00636	0.00364	0.00274	0.00274	0.00183
$\left\|\hat{X}^{(k)}-X\right\|_2$	9.99445	1.43025	1.39463	1.40163	1.39763	1.40064	1.39763	1.39964