

Parameter Estimation of Least Squares Collocation

Lihong Jin

Department of Basic, City College, Wuhan University of Science and Technology, Wuhan, 430083, China

Abstract—In this paper, based on penalized least squares, the penalized weighted sum of squares is set up, we deduce the calculation method of a positively definite regularity matrix and obtain the corresponding results in least squares collocation model. According to the foundational properties of the random errors, we study how to choose a reasonable smoothing parameter and regularizer matrix. Concepts of confidence region in probability and method of region estimation are used. The adaptation of this technique is more flexible than the traditional model.

Keywords- regularizer matrix; parametric estimation; least squares collocation; smoothing parameter

I. INTRODUCTION

Least squares collocation (LSC) is the traditional method for gravity field transformations, being widely used in many geodetic branches [1-6]. Least squares collocation (LSC) is widely used when one is interested in the integration of spatially connected observations of different types in order to compute a related quantity of which, in the case of physical geodesy, they are functional. In the 1980s, statistics brought forward semi-parametric regression models comprising a parametric part and a non-parametric part [7]. It has some significant merits over parametric models in dealing with complicated relationships between observations and estimated variables and offering useful information. This model has been discussed by a number of authors [7]. See, for example, Fischer and Hegland [8], Hong Shengyan [9], Liang hua [10], P.J. Green and B.W. Silverman [11]. This paper tries to introduce the semi-parametric regression method for surveying adjustment, and to develop a suitable method that can estimate the unknown parameters and can extract method errors simultaneously.

II. ESTIMATE METHOD AND MODELS SOLUTION

Least squares collocation is based on the adjustment within a mixed model where random and fixed parameters are treated simultaneously. The basic linearized observation equations for LSC can be written as:

$$L = B X + G Y + \Delta$$

$$E(\Delta) = 0, \quad D_{\Delta} = \sigma_0^2 Q \quad (1)$$

With $L \in R^n$ is a vector of observations, Y is a vector of fixed parameters to be estimated, X is signal and a vector of

spatially connected random parameters to be predicted, $X = (X_1, X_2)^T$, $EX = E(X_1, X_2)^T = 0$, X_1 is a signal to the observations, X_2 is a signal to be predicted, $B \in R^{n \times m}$ is the designed matrix relating the fixed parameters to the observations with $rank(B) = m$, $rank(G) = t$, $\Delta \in R^n$ is the vector of measurement errors, σ^2 is the unit-weighted variance.

Corresponding to model (1), its error equation is

$$V = B\hat{X} + G\hat{Y} - L \quad (2)$$

where \hat{X} and \hat{Y} denoting the estimates of parametric X and non-parametric Y , $V \in R^n$ is the residues.

Because of the unknown parametric have a total number of $m + t$, but the numbers of equations are only t . Therefore, the unique solution cannot be obtained by the standard least squares technique. So the principle of least squares should be modified. A logical choice [4] is

$$V^T P V + \alpha \hat{X}^T R \hat{X} = \min \quad (3)$$

Where $P = Q^{-1}$ is symmetric positive-definite matrix, and also is the weighted matrix of the observation L . The matrix R is a given positive-definite matrix, which is named as regularizer matrix, α is the smoothing parametric and governs the balance between the two terms of V and \hat{X} .

Construct Lagrange function:

$$\Phi = V^T P V + \alpha \hat{X}^T R \hat{X} + 2K^T (B\hat{X} + G\hat{Y} - L - V) \quad (4)$$

where $K^T = [k_1, k_2, \dots, k_n]$, let $\frac{\partial \Phi}{\partial V} = 0$, $\frac{\partial \Phi}{\partial \hat{Y}} = 0$ and

$\frac{\partial \Phi}{\partial \hat{X}} = 0$, respectively. We can obtain

$$K = P V \quad (5)$$

$$G^T K = 0 \tag{6} \quad = \left(M + G(I - M)(G^T P(I - M)G)^{-1} G^T P(I - M) \right) L \tag{14}$$

$$\alpha R \hat{X} + B^T K = 0 \tag{7}$$

Let $J(\alpha) = \left(M + G(I - M)(G^T P(I - M)G)^{-1} G^T P(I - M) \right)$, then

Substituting (5) into (6), and considering (2),

$$G^T P B \hat{X} + G^T P G \hat{Y} = G^T P L \tag{8} \quad \hat{L} \triangleq J(\alpha) L \tag{15}$$

Substituting (5) into (7), and considering (2),

$$(\alpha R + B^T P B) \hat{X} + B^T P G \hat{Y} = B^T P L \tag{9}$$

By simple calculus (8) and (9), it follows that (4) is minimized when \hat{X} and \hat{Y} satisfy the block matrix equation:

$$\begin{pmatrix} G^T P B & G^T P G \\ \alpha R + B^T P B & B^T P G \end{pmatrix} \begin{pmatrix} \hat{X} \\ \hat{Y} \end{pmatrix} = \begin{pmatrix} G^T P L \\ B^T P L \end{pmatrix} \tag{10}$$

We considering (8), (9) and (10), we can obtain

$$\hat{X} = (B^T P B + \alpha R)^{-1} B^T P (L - G \hat{Y}) \tag{11}$$

$$\hat{Y} = (G^T P(I - M)G)^{-1} G^T P(I - M)L \tag{12}$$

where $M = B(B^T P B + \alpha R)^{-1} B^T P$, by select suitable smoothing parametric α and regularize matrix, R , \hat{X} and \hat{Y} can be calculated by (11) and (12).

III. STATISTIC PROPERTIES OF ESTIMATORS

Some properties of estimators are discussed in the following.

Remark 1 \hat{Y} is unbiased estimation of Y .

Proof: using (12), we can get

$$\begin{aligned} E(\hat{Y}) &= (G^T P(I - M)G)^{-1} G^T P(I - M)E(L) \\ &= (G^T P(I - M)G)^{-1} G^T P(I - M)GY = Y \end{aligned} \tag{13}$$

Remark 2. By simple calculus (11) and (12), the estimator of observations is

$$\hat{L} = B \hat{X} + G \hat{Y}$$

$J(\alpha)$ called the hat- or smoother matrix.

Remark3. Variance matrix of \hat{X} is

$$D_{\hat{X}} = M_1 \cdot D_{\Delta} \cdot M_1^T \tag{16}$$

Where $M_1 = (B^T P B + \alpha R)^{-1} B^T P (I - G(G^T P(I - M)G)^{-1} G^T P(I - M))$; Variance matrix of \hat{Y} is

$$D_{\hat{Y}} = (G^T P(I - M)G)^{-1} G^T P(I - M) \cdot D_{\Delta} (I - M)^T P (G(I - M)^T P G)^{-1} \tag{17}$$

Based on (15) and (16), we can obtained the Variance matrix of \hat{V} is

$$D_{\hat{V}} = J(\alpha) \cdot D_{\Delta} \cdot J(\alpha)^T \tag{18}$$

IV. SMOOTHING PARAMETER SELECTION

As is well known, the smoothing parameter α controls the trade-off between the goodness of fit and the roughness, and it is important to choose an appropriate α .

The residual sum of squares (RSS) of a model is a measure of predictive ability, since a residual is the difference between an observation of a response and its fitted or predicted value:

$$e_i \triangleq L_i - \hat{L}_i \tag{19}$$

However, RSS is not satisfactory as a model selector. The problem is that \hat{L}_i uses L_i as well as the other observations to predict L_i . The result is that the most complex model – that is, the model with the largest degrees of freedom and containing the other models as special cases – always has the smallest RSS. On the other hand, because the observation L_i is being used as part of its own predictor, therefore, a small amount of smoothing, which gives high weight to L_i , appears optimal for prediction according to RSS.

There is a simple remedy to this problem: when predicting

L_i , use all the observations except the i th one. Thus, define $\hat{L}^{-i}(t_i, \alpha)$ is the semi-parametric regression estimator applied to the data but with (t_i, L_i) deleted. Then, let $e_{(i)} \hat{=} L_i - \hat{L}^{-i}(t_i, \alpha)$ be the i th deleted residual. The predicted residual sum of squares (PRESS) is defined by

$$PRESS = \sum_{i=1}^n e_{(i)}^2 \tag{20}$$

This method is the technique of model validation that splits the data into two disjoint sets, fits the model to one set, predicts the data in the second set using only the fit to the first set, and then compares these predictions to the actual observations in the second set. This “leaving one out” strategy is a way of guarding against the wiggly answer that RSS gives. The choice of α is the one that minimizes PRESS over $\alpha > 0$.

The PRESS criterion is not as difficult to calculate as it might first appear. One does not need to fit the model n times, thanks to an important identity. Let $J_{ii}(\alpha)$ be the i th diagonal element of the hat matrix $J(\alpha)$. Then the i th deleted residual is related to the i th ordinary residual by

$$L_i - \hat{L}^{-i}(t_i, \alpha) = \frac{L_i - \hat{L}_i}{1 - J_{ii}(\alpha)} \tag{21}$$

Using (20), we can obtain

$$PRESS(\alpha) = \sum_{i=1}^n \left(\frac{L_i - \hat{L}_i}{1 - J_{ii}(\alpha)} \right)^2 \tag{22}$$

Thus, the predicted residual sum of squares (PRESS) can be computed using only ordinary residuals and the diagonal elements of the hat matrix.

V. HYPOTHESIS TESTING

First consider the problem of hypothesis testing for y_i , the i th entry of Y .

By simple calculus (13), we can obtain the estimator of variance $\hat{\sigma}^2$ is

$$\hat{\sigma}^2 = \frac{(L - \hat{L})^T (L - \hat{L})}{tr(I - J(\alpha))} \tag{23}$$

Where $I = I_n$ being $n \times n$ identity matrix, and $J(\alpha)$ is the hat- or smoother matrix in (14).

Specifically, for the hypothesis set-up

$$H_0: y_i = 0 \quad \text{versus} \quad H_1: y_i \neq 0, \quad i = 1, 2, \dots, t$$

In order to distinguish with the smoothing parameter, we may assume the significance level to be α_1 here. Set the estimates of observation L resulting from adjustment as \hat{L} , owing to the estimation of variance is formula (22), $J(\alpha)$ is the hat matrix. Thus, we get the test statistics:

$$T = \frac{\hat{y}_i}{\hat{\sigma} \sqrt{c_{ii}}} \sim t(f),$$

Where c_{ii} is the i th diagonal element of the matrix $(G^T P(I - M)G)^{-1}$, T is t distribution statistics with degree of freedom as $f = tr(I - J(\alpha))$. The rejection region is $|T| \geq t_{\frac{\alpha_1}{2}}$; Through the test, we can give specific analysis, correct the observation equation of observe value, so as to achieve the aim of improving precision.

VI. EXAMPLE COMPERISON

For many areas, the gravity anomaly not only contains a random part, but also contains the system part, the system part usually expressed as a linear function of each point coordinates (x_i, y_i) :

$$T_i = a_0 + a_1(x_i - \bar{x}) + a_2(y_i - \bar{y}), \quad i = 1, 2, \dots, n \tag{24}$$

Where \bar{x} 、 \bar{y} respectively represent the average of observation point coordinates, a_0 、 a_1 and a_2 are the undetermined parameter in the function, all which are undetermined non-random parameters, that is tendency parameters, cause:

$$GY = \begin{bmatrix} 1 & x_1 - x_0 & y_1 - y_0 \\ 1 & x_2 - x_0 & y_2 - y_0 \\ \dots & \dots & \dots \\ 1 & x_n - x_0 & y_n - y_0 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} \tag{25}$$

Set the random part(sign) of gravity anomaly as X , we may assume that $B_1 = I$ (I is the unit matrix),the gravity anomaly value Δg on the observed points should be

$$\Delta g = L - \Delta = GY + X \tag{26}$$

But for the unobserved points, if use X' represent the sign on these points, interpolating the gravity anomaly value Δg_p according to the following formula

$$\Delta g_p = G_p Y + X' \tag{27}$$

Then the observation equation is

$$L = GY + X + \Delta \tag{28}$$

This example taken from literature[4]. Assume that the gravity anomaly value L of four observation points $P_1 \sim P_4$ and their coordinates x_i, y_i is known as table 1. The variance of observation error (noise) is $D_\Delta = (0.03)^2 I$, try to calculate the gravity anomaly estimations of the observed points $P_1 \sim P_4$ and unobserved points P'_1 and P'_2 .

TABLE I. GRAVITY ANOMALY VALUES OF OBSERVATIONS AND ITS LOCATIONS

Points	$L(mGal)$	$x(m)$	$y(m)$
P_1	-0.55	+640	+480
P_2	-0.23	+440	+400
P_3	+0.58	+140	+140
P_4	-1.80	+620	+180
P'_1		+500	+300
P'_2		+460	+300

To calculate the distances from each point according to the known points, we can obtain

$$R_{x_i} = \begin{pmatrix} 0 & 2.15 & 6.05 & 3.01 \\ 2.15 & 0 & 3.97 & 2.84 \\ 6.05 & 3.97 & 0 & 4.82 \\ 3.01 & 2.84 & 4.82 & 0 \end{pmatrix},$$

$$R_{x_1, x_2} = \begin{pmatrix} 2.28 & 1.17 & 3.94 & 1.70 \\ 2.55 & 1.02 & 3.58 & 2.00 \end{pmatrix},$$

$$R_{x_2} = \begin{pmatrix} 0 & 0.40 \\ 0.40 & 0 \end{pmatrix},$$

$$R = \begin{pmatrix} R_{x_1} & R_{x_1, x_2}^T \\ R_{x_1, x_2} & R_{x_2} \end{pmatrix}^{-1} G = \begin{pmatrix} 1 & +1.8 & +1.8 \\ 1 & -0.2 & +1.0 \\ 1 & -3.2 & -1.6 \\ 1 & +1.6 & -1.2 \end{pmatrix},$$

$$G_p = \begin{pmatrix} 1 & +0.4 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

Using the cross verification method here, the smoothing parameter $\alpha = 1$ can be calculated by (21). We can obtain by (11),(12)

$$\hat{Y} = (-0.3888, -0.4879, 0.3809)^T,$$

$$\hat{X}_1 = (0.0315, -0.3198, 0.017, -0.1736)^T,$$

$$\hat{X}_2 = (-0.2204, -0.2386)^T,$$

Then, the gravity anomaly estimations of $P_1 \sim P_4$ is:

$$\Delta g = G\hat{Y} + \hat{X}_1 = (-0.5499, -0.2301, 0.5800, -1.800)^T$$

The gravity anomaly estimations of unobserved points P'_1 and P'_2 is

$$\Delta g_p = G_p \hat{Y} + \hat{X}_2 = (-0.8043, -0.6274)^T,$$

The results calculated by least square collection are shown in literature [4], we found that the estimated result is extremely close to the real value, after comparing the results with that in literature [4]. It shows that the least square collection has strong adaptability.

REFERENCES

- [1] SHA Yuejing, The Application of Least Square Collection Method in the GPS Elevation [J],Journal of Geomatics,2000,3. 3-5.
- [2] HUANG Motao,ZHAI Guojun, et al, The Recovery of Bathymetry from Altimeter Data [J],Geomatics and Information Science of Wuhan University, 2002, 2(27):133-137.
- [3] Wang Jiexian, Using the GPS Observations and Gravitation Fields Information to Estimate the Elevation [J], .Annals of Shanghai Observatory Academia Sinica,1992,13,67~73
- [4] YU Zhong chou,Tao Benzao, et al, Generalized measurement adjustment (New version) [M],Wuhan:Wuhan Technical University of Surveying and Mapping, ,1996.83-95.
- [5] SUN Hai yan,WU Yun. Semiparametric Regression and Model Refining[J], Geo-spatial Information Science (Quarterly), 2002,4(5):pp10~13 .
- [6] LIU Nian, Least-Squares Interpolation and Collection [J],Science of Mapping and Surveying,2002, 3(27):pp568~ 581.

- [7] Green P J, Silverman B W, Nonparametric Regression and Generalized linear Models[M]. London: Chapman and Hall, 1994.
- [8] Beat Fischer and Markus Hegland. Collocation, Filtering and Nonparametric Regression, Part 1. ZfV, 1999, 1, 17-24.
- [9] Hong Shengyan, Automatic bandwidth selection and data-driven estimators in a semiparametric regression model (Doctor dissertation), USA, 1998.
- [10] Liang hua, Related topics in partially linear models (Doctor dissertation), USA, 2001.
- [11] P.J.Green and B.W.Silverman. Nonparametric Regression and Generalized Linear Models. London: Chapman & Hall, 1994