Unknown parameter's variance-covariance propagation and calculation in generalized nonlinear least squares problem*

TAO Hua-xue, GUO Jin-yun
(College of Geo-Information Science and Technology, Shandong University of Science and Technology, Taian 271019, Shandong, China)

1 Introduction

Now suppose there is a nonlinear function

$$\mathbf{L} = f\ (\mathbf{x}, \mathbf{y}) \tag{1}$$

where $\mathbf{L} = (l_1, l_2, \cdots, l_m)^T$ is a group of observing data containing errors; $\mathbf{y} = (y_1, y_2, \cdots, y_m)^T$ is another type of data containing errors and different accuracy. The observing powers of two types data are $\mathbf{P}_1$ and $\mathbf{P}_2$ respectively. The vector $\mathbf{y}$ is the random parameter and $\mathbf{x} = (x_1, x_2, \cdots, x_n)^T$ is the unknown non-random parameter in the function model. Concurrently considering the errors of $\mathbf{L}$ and $\mathbf{y}$, $\mathbf{L} = f(\mathbf{x},\mathbf{y})$ can optimally fit these data to make the square sum of fitted error minimum. This is the generalized nonlinear dynamic least squares problem [1, 2]. The number of generalized nonlinear least squares problem is **m** more than that of common nonlinear least squares problem [3, 4]. So it is more difficult to solve the generalized nonlinear least squares problem than the latter. So, based on the special structure of generalized nonlinear least squares problem, it can be separated to simplify the high-dimensional equation group and reduce the dimension of unknown parameters to be half. With the separated method to solve the generalized dynamic least squares problem, the relation of function $f(\mathbf{x},\mathbf{y})$ dependent on another variable can be removed to satisfy for the one-degree essential condition of optimal solution. Then the problem can be translated into the common nonlinear least squares problem to relate with one variable. Finally the problem (1) is separated into two sub-problems to be solved.
According to the nonlinear variance-covariance propagation theory [5, 6], the generalized variance-covariance propagation [7] of unknown parameter $\mathbf{x}$ in the generalized nonlinear least squares problem can be derived from the errors of $\mathbf{L}$ and $\mathbf{y}$ with the separated method. So the generalized nonlinear probem is divided into two sub-problems. Firstly the variance-covariance model of unknown parameter $\mathbf{x}$ can be derived from the error of observation $\mathbf{L}$ for the first sub-problem. Then the variance-covariance model of unknown parameter $\mathbf{x}$ can be derived from the error of another data $\mathbf{y}$ for the second sub-problem. Finally the integral variance-covariance of unknown parameter $\mathbf{x}$ is again calculated.

2 Variance-covariance propagation of unknown parameter x derived from the error of observation L

Now assuming that the initial value $\bar{\mathbf{y}}$ of $\mathbf{y}$ is fixed, so the variable $\mathbf{y}$ can be separated from problem (1). Then

$$\mathbf{L} = f\ (\mathbf{x}, \bar{\mathbf{y}}) \tag{2}$$

There are m observations $\underset{m\times 1}{\mathbf{L}}$ whose variance-covariance matrix is

$$\sum\nolimits_{\mathbf{L}} = \mathrm{E}[(\mathbf{L} - \mathrm{E}(\mathbf{L}))(\mathbf{L} - \mathrm{E}(\mathbf{L}))^T] \tag{3}$$

There are n unknown parameters $\underset{n\times 1}{\mathbf{x}}$ calculated from the observations. So $\mathbf{x})\psi(l_1, l_2, \cdots, l_m)$ is the function of $\mathbf{L}$, which can be noted as $\mathbf{x} = \mathbf{x}(\mathbf{L})$. So the variance-covariance matrix of unknown parameter $\mathbf{x}$ is

$$\sum\nolimits_{\mathbf{x}}^{1} = \mathrm{E}[\mathbf{x} - \mathrm{E}(\mathbf{x}))(\mathbf{x} - \mathrm{E}(\mathbf{x}))^T] \tag{4}$$

How to determine the relation between $\sum_{\mathbf{x}}^{1}$ and $\sum_{\mathbf{L}}$ $\hat{\mathbf{x}}(\hat{\mathbf{L}})$ is obtained from the adjusted observation. Obviously $\mathrm{E}(\mathbf{x}) \neq \hat{\mathbf{x}}(\hat{\mathbf{L}})$. So the difference between $\mathrm{E}(\mathbf{x})$ and $\hat{\mathbf{x}}(\hat{\mathbf{L}})$ is calculated as $\boldsymbol{\delta} = \mathrm{E}(\hat{\mathbf{x}}(\hat{\mathbf{L}}) - \hat{\mathbf{x}}(\mathbf{L}))$. Then $\mathrm{E}(\mathbf{x}(\mathbf{L}) = \hat{\mathbf{x}}(\hat{\mathbf{L}}) - \boldsymbol{\delta}$. We can get

$$\sum\nolimits_{\mathbf{x}}^{1} = \mathrm{E}[(\mathbf{x}(\mathbf{L}) - \hat{\mathbf{x}}(\hat{\mathbf{L}}))(\mathbf{x}(\mathbf{L}) - \hat{\mathbf{x}}(\hat{\mathbf{L}}))^T] - \boldsymbol{\delta}\boldsymbol{\delta}^T \tag{5}$$

$\mathbf{x}(\mathbf{L}) - \hat{\mathbf{x}}(\hat{\mathbf{L}})$ is expanded into Taylor series at $\hat{\mathbf{L}}$, considering one and two degrees terms and neglecting the higher-degree terms, that is

* **Foundation item:** Project (40174003) supported by the National Natural Science Foundation of China.
Correspondence: TAO Hua-xue, Superior of Ph. D. student, Professor, Foreign academician of Russia Academy of Natural Sciences; Tel.: +86-5 38-6 22 65 34; E-mail: taohx@ta-public.sd.cninfo.net.

$$\mathbf{x}(\mathbf{L})-\hat{\mathbf{x}}(\hat{\mathbf{L}})=\frac{\partial\mathbf{x}}{\partial\mathbf{L}^T}\Big|_{\hat{\mathbf{L}}}(\mathbf{L}-\hat{\mathbf{L}})+\frac{1}{2}\frac{\partial^2\mathbf{x}}{\partial(\mathbf{L}^T)^2}\Big|_{\hat{\mathbf{L}}}(\mathbf{L}-\hat{\mathbf{L}})^2 \quad (6)$$

$$where\ \frac{\partial\mathbf{x}}{\partial\mathbf{L}^T}=\begin{vmatrix}\frac{\partial x_1}{\partial\mathbf{L}^T}\\ \frac{\partial x_2}{\partial\mathbf{L}^T}\\ \vdots\\ \frac{\partial x_n}{\partial\mathbf{L}^T}\end{vmatrix}=\begin{bmatrix}\frac{\partial x_1}{\partial l_1}&\frac{\partial x_1}{\partial l_2}&\cdots&\frac{\partial x_1}{\partial l_m}\\ \frac{\partial x_2}{\partial l_1}&\frac{\partial x_2}{\partial l_2}&\cdots&\frac{\partial x_2}{\partial l_m}\\ \vdots&\vdots&\vdots&\vdots\\ \frac{\partial x_n}{\partial l_1}&\frac{\partial x_n}{\partial l_2}&\cdots&\frac{\partial x_n}{\partial l_m}\end{bmatrix}_{n\times m}=\begin{bmatrix}\mathbf{J}_{11}\\ \mathbf{J}_{11}\\ \vdots\\ \mathbf{J}_{1n}\end{bmatrix}=\mathbf{J}_1,$$

in which $\mathbf{J}_1$ is a Jacobi matrix; and

$$\frac{\partial^2\mathbf{x}}{\partial(\mathbf{L}^T)^2}=\begin{vmatrix}\frac{\partial^2 x_1}{\partial(\mathbf{L}^T)^2}\\ \frac{\partial^2 x_2}{\partial(\mathbf{L}^T)^2}\\ \vdots\\ \frac{\partial^2 x_n}{\partial(\mathbf{L}^T)^2}\end{vmatrix}=\begin{bmatrix}\frac{\partial\mathbf{J}_{11}}{\partial\mathbf{L}^T}\\ \frac{\partial\mathbf{J}_{12}}{\partial\mathbf{L}^T}\\ \vdots\\ \frac{\partial\mathbf{J}_{1n}}{\partial\mathbf{L}^T}\end{bmatrix}=\begin{bmatrix}\mathbf{S}_{11}\\ \mathbf{S}_{12}\\ \vdots\\ \mathbf{S}_{1n}\end{bmatrix}_{n\times m^2}=\mathbf{S}_1,$$

in which $\mathbf{S}_1^T=\mathbf{H}_1$ and $\mathbf{H}_1$ is a Hasse matrix of vector function.
So the equation (6) can be rewritten as

$$\mathbf{x}(\mathbf{L})-\hat{\mathbf{x}}(\hat{\mathbf{L}})=\mathbf{J}_1\mathbf{V}_1+\frac{1}{2}\mathbf{H}_1^T(\mathbf{V}_1\otimes\mathbf{V}_1) \quad (7)$$

where $\mathbf{V}_1=\mathbf{L}-\hat{\mathbf{L}}$, and $\mathbf{V}_1=(V_{11},V_{12},\cdots,V_{1m})^T$; and $\otimes$ is the Kronecker product operator.
Finally we can get

$$\sum\nolimits_{\mathbf{x}}^1=\mathbf{J}_1\sum\mathbf{J}_1^T+\frac{1}{2}\mathbf{H}_1^T\,\mathrm{E}((\mathbf{V}_1\otimes\mathbf{V}_1)\mathbf{V}_1^T)\mathbf{J}_1^T+$$
$$+\frac{1}{2}\mathbf{J}_1\mathrm{E}(\mathbf{V}_1(\mathbf{V}_1\otimes\mathbf{V}_1))\mathbf{H}_1 \quad (8)$$

where $\mathrm{E}(\mathbf{V}_1\otimes\mathbf{V}_1)=Vec\sum_{\mathbf{L}}$, in which Vec is a vector operator indicating the straightening operation.

When there are no dependent relations among observation

$$\mathbf{L},\ \text{then}\sum\nolimits_{\mathbf{L}}=\begin{bmatrix}\sigma_{l_1}&&&\\&\sigma_{l_2}&&\\&&\ddots&\\&&&\sigma_{l_m}\end{bmatrix},$$

$$\mathrm{E}(\mathbf{V}_1\otimes\mathbf{V}_1)=Vec\sum\nolimits_{\mathbf{L}}=\begin{bmatrix}\sigma_{l_1}^2\\ \sigma_{l_2}^2\\ \vdots\\ \sigma_{l_m}^2\end{bmatrix},\ \mathrm{E}((\mathbf{V}_1\otimes V_1)\mathbf{V}_1^T)=0$$

and $\mathrm{E}(\mathbf{V}_1(\mathbf{V}_1\otimes\mathbf{V}_1))=0$.

$$\text{So}\quad \sum\nolimits_{\mathbf{x}}^1=\mathbf{J}_1\sum\nolimits_{\mathbf{L}}\mathbf{J}_1^T \quad (9)$$

There appears the Jacobi matrix $\mathbf{J}_1$ in the above equation. How to calculate it? From equation (2), the error equation is $\mathbf{V}_1=f(\mathbf{x},\bar{\mathbf{y}})-\mathbf{L}$. Based on the least squares principal, that is $\mathbf{V}_1^T\mathbf{P}\mathbf{V}_1$ = min, then

$$\mathbf{A}^T\mathbf{P}[f(\mathbf{x},\bar{\mathbf{y}})-\mathbf{L}]=0 \quad (10)$$

Obviously the equation (10) is the function of the observation $\mathbf{L}$ and the unknown parameter $\mathbf{x}$. Let

$$\psi(\mathbf{x},\mathbf{L})=\mathbf{A}^T\mathbf{P}[f(\mathbf{x},\bar{\mathbf{y}})-\mathbf{L}]=0 \quad (11)$$

The complete differential is made to Equation (11), then

$$\frac{\partial\psi}{\partial\mathbf{x}}d\mathbf{x}+\frac{\partial\psi}{\partial\mathbf{L}}d\mathbf{L}=0 \quad (12)$$

$$\text{So }\mathbf{J}_1=\frac{\partial\mathbf{x}}{\partial\mathbf{L}}=-\left(\frac{\partial\psi}{\partial\mathbf{x}}\right)^{-1}\left(\frac{\partial\psi}{\partial\mathbf{L}}\right) \quad (13)$$

From equation (11) then

$$\frac{\partial\psi}{\partial\mathbf{L}}=-\mathbf{A}^T\mathbf{P}\ \text{and}\ \frac{\partial\psi}{\partial\mathbf{x}}=\mathbf{A}^T\mathbf{P}\mathbf{A}+Vec_n^{-1}(\mathbf{Q},\mathbf{W}).$$

$$\text{So }\mathbf{J}_1=((\mathbf{A}^T\mathbf{P}\mathbf{A})+Vec_n^{-1}(\mathbf{Q},\mathbf{W}))^{-1}\mathbf{A}^T\mathbf{P} \quad (14)$$

where $\mathbf{Q}$ is a $n^2\times m$ Hesse matrix of $f(\hat{\mathbf{x}},\bar{\mathbf{y}})$, $\mathbf{W}$ is a column vector whose dimension is $m\times 1$, $\mathbf{W}=\mathbf{P}\mathbf{V}_1$, Vec_n^{-1} is the inverse operation of operator Vec and the lower script n is the column number of matrix $Vec_n^{-1}(\mathbf{Q},\mathbf{W})$. Substitute equation (14) for equation (9), the variance-covariance of unknown $\mathbf{x}$ can be computed.

3 Variance-covariance propagation of unknown parameter x derived from the error of another observation y

Let a suitably little vector $\bar{\mathbf{x}}$ be the initial value of unknown parameter $\mathbf{x}$. Fixed $\bar{\mathbf{x}}$, then $\mathbf{x}$ can be separated from problem (11) [8, 9]. First the relation between $\sum_{\mathbf{L}}$ and $\sum_{\mathbf{y}}$ is calculated and then the relation between $\sum_{\mathbf{y}}$ and $\sum_{\mathbf{x}}^2$ is also obtained.
Now $\mathbf{L}=f(\bar{\mathbf{x}},\mathbf{y})$. Same as the above the conclusion, we can get $\sum_{\mathbf{L}}=\mathrm{E}[(\mathbf{L}-\mathrm{E}(\mathbf{L}))(\mathbf{L}-\mathrm{E}(\mathbf{L}))^T]$ and $\sum_{\mathbf{y}}=\mathrm{E}[(\mathbf{y}-\mathrm{E}(\mathbf{y}))(\mathbf{y}-\mathrm{E}(\mathbf{y}))^T]$. $\mathbf{L}$ is a nonlinear function of $\mathbf{y}$, which is noted as $\mathbf{L}=\mathbf{L}(\mathbf{y})$. Obviously $\mathrm{E}(\mathbf{L}(\mathbf{y}))\neq\hat{\mathbf{L}}(\hat{\mathbf{y}})$, where $\hat{\mathbf{L}}(\hat{\mathbf{y}})$is solved from the adjusted values of $\mathbf{y}$. Let $\sigma=\mathrm{E}(\hat{\mathbf{L}}(\hat{\mathbf{y}})-\mathbf{L}(\mathbf{y}))$. Then

$$\sum\nolimits_{\mathbf{L}}=\mathrm{E}(\mathbf{L}(\mathbf{y})-\hat{\mathbf{L}}(\hat{\mathbf{y}}))(\mathbf{L}(\mathbf{y})-\hat{\mathbf{L}}(\hat{\mathbf{y}}))^T]-\sigma\sigma^T \quad (15)$$

$\mathbf{L}(\mathbf{y})-\hat{\mathbf{L}}(\hat{\mathbf{y}})$ can be expanded into the Taylor series at $\bar{\mathbf{y}}$, considering the first and second degrees terms and neglecting the higher-degree terms. Then

$$\mathbf{L}(\mathbf{y})-\hat{\mathbf{L}}(\hat{\mathbf{y}})=\frac{\partial\mathbf{L}}{\partial\mathbf{y}^T}\Big|_{\hat{\mathbf{y}}}(\mathbf{y}-\hat{\mathbf{y}})+\frac{1}{2}\frac{\partial^2\mathbf{L}}{\partial(\mathbf{y}^T)^2}\Big|_{\hat{\mathbf{y}}}(\mathbf{y}-\hat{\mathbf{y}})^2,$$

$$\text{where }\frac{\partial\mathbf{L}}{\partial\mathbf{y}^T}=\begin{bmatrix}\frac{\partial l_1}{\partial\mathbf{y}^T}\\ \frac{\partial l_2}{\partial\mathbf{y}_T}\\ \vdots\\ \frac{\partial l_m}{\partial\mathbf{y}_m}\end{bmatrix}=\begin{bmatrix}\mathbf{J}_{21}\\ \mathbf{J}_{22}\\ \vdots\\ \mathbf{J}_{2m}\end{bmatrix}=\mathbf{J}_2\ \text{and}$$

$$\frac{\partial^2 \mathbf{L}}{\partial(\mathbf{y}^T)^2} = \begin{bmatrix} \frac{\partial^2 l_1}{\partial(\mathbf{y}^T)^2} \\ \frac{\partial^2 l_2}{\partial(\mathbf{y}^T)^2} \\ \vdots \\ \frac{\partial l_m}{\partial(\mathbf{y}^T)^2} \end{bmatrix} = \mathbf{S}_2. \text{ Let } \mathbf{S}_2^T = \mathbf{H}_2, \text{ where } \mathbf{H}_2 \text{ is a}$$

Hasse matrix. When there is no relation among observations $\mathbf{y}$, then

$$\sum\nolimits_{\mathbf{L}} = \mathbf{J}_2 \sum\nolimits_{\mathbf{y}} \mathbf{J}_2^T \quad (16)$$

where $\mathbf{J}_2 = ((\mathbf{B}^T\mathbf{P}\mathbf{B}) + Vec_m^{-1}(\mathbf{Q}_1\mathbf{W}_1))^{-1}\mathbf{B}^T\mathbf{P}$ in which $\mathbf{Q}_1$ is a $m^2 \times m$ Hesse matrix of $f(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ and $\mathbf{W}_1$ is a $m \times 1$ column vector, $\mathbf{W}_1 = \mathbf{P}\mathbf{V}_2$ and $\mathbf{V}_2 = f(\hat{\mathbf{x}}, \hat{\mathbf{y}}) - \mathbf{L}$; and $\mathbf{B} = \left.\frac{\partial f(\bar{\mathbf{x}}, \mathbf{y})}{\partial \mathbf{y}}\right|_{\hat{\mathbf{y}}}$.

Substitute $\mathbf{J}_2$ for equation (16), $\sum_{\mathbf{L}}$ can be calculated. Then substitute equation (16) for equation (9), we can get from the relation between $\sum_{\mathbf{L}}$ and $\sum_{\mathbf{y}}$ as

$$\sum\nolimits_{\mathbf{x}}^2 = \mathbf{J}_1\mathbf{J}_2 \sum\nolimits_{\mathbf{y}} \mathbf{J}_2^T\mathbf{J}_1^T \quad (17)$$

4 Variance-covariance of unknown parameter x

From equation (9) and (17), the variance-covariance of unknown parameter $\mathbf{x}$ in the generalized nonlinear least squares problem (1) can be calculated as

$$\sum\nolimits_{\mathbf{x}}^1 + \sum\nolimits_{\mathbf{x}}^2 = \sum\nolimits_{\mathbf{x}} \quad (18)$$

Based on the nonlinear variance-covariance propagation theory, the variance-covariance propagation formula of generalized nonlinear least squares problem is concluded in the paper. Because of considering the first and second degrees terms of Taylor series, the variance-covariance formula of unknown parameter for the generalized nonlinear least squares problem can improve the accuracy estimation of unknown parameter. It opens up a new way to estimate the accuracy of unknown parameter for the generalized nonlinear least squares problem. So it is very significant.

References

[1] TAO Hua-xue, GUO Jin-yun: Separating iterative solution model of generalized nonlinear dynamic least Squares for data processing in building of digital earth. Transactions of Nonferrous Metals Society of China, 2003, 13(3): 720–723.

[2] MA Yang-ming: Analysis on effect of generalized nonlinear model. Mathematical Statistics and Applied Probability, 1996, 11(3): 195–201. (in Chinese)

[3] SUN Fu-guo: Study and achievement of generalized least squares method. Journal of Changsha Railway College, 1999, (3): 102–107. (in Chinese)

[4] MA Xiao-fen: Numerical method to solve nonlinear least squares problem. Dissertation for Ph. D. Chengdu, Southwest Jiaotong University, 1996. (in Chinese)

[5] HU Sheng-wu. Error propagation of nonlinear model and ist application in GIS. Journal of Wuhan Technical University of Surveying and Mapping, 1997, 22(2): 129–131. (Chinese)

[6] XU Pei-liang: Covariance propagation formula of nonlinear function. Journal of Wuhan Technical University of Surveying and Mapping, 1996, 21(2): 92–99. (in Chinese)

[7] ZHOU Shi-jian: Generalized variance-covariance propagation. Beijing, Press of Surveying and Mapping, 1996. (in Chinese)

[8] TAO Hua-xue, WANG Tong-xiao, WANG You-xuan: Variance-covariance propagation and calculation of the nonlinear least squares adjustment parameters. Site Investigation Science and Technology, 1998, (1): 44–48. (in Chinese)

[9] LIU Guo-lin, TAO Hua-xue: Covariance and cofactor propagation and weight reciprocal for nonlinear observation function. Surveying and Mapping Engineering, 1997, 6(2): 8–16. (in Chinese)

[10] GONG Shou-wen: Variance estimation of dynamic dispersed observation. Crust Deformation and Earthquake. 1987, 7(2): 89–96. (in Chinese)

Adress of the author:
Prof. Hua-xue TAO,
College of Geo-information Sci. and Tech.,
Shandong University of Sci. and Tech.,
Daizong Street 223#, Taian 271019, Shandong Prov., PR China

Abstract

The unknown parameter's variance-covariance propagation and calculation in the generalized nonlinear least squares remain to be studied now, which didn't appear in the internal and external referencing documents. The unknown parameter's variance-covariance propagation formula, considering the two-power terms, is concluded used to evaluate the accuracy of unknown parameter estimators in the generalized nonlinear least squares problem. It is a new variance-covariance formula and opens up a new way to evaluate the accuracy when processing data which have the multi-source, multi-dimensional, multi-type, multi-time-state, different accuracy and nonlinearity.

Keywords

Generalized nonlinear least squares problem, unknown parameter, variance-covariance propagation.

CLC number: TB22 Document code: A