Processing Approach of Non-linear Adjustment Models in the Space of Non-linear Models

LI Chaokui ZHU Qing SONG Chengfang

1 Introduction

Bates (1980)\(^1\) pointed out that each non-linear model may contain two vectors. The measuring symbols relating to the vectors are IN (intrinsic nature) and PE (parameter effect). In fact, most of non-linear models in adjustment and data processing are weak non-linear models. Stress non-linear models seldom show themselves, but really exist, such as bi-twist bi-curve function, mantle convection power equation etc. The stress non-linear models are not fit for being linearized because:

1) there exists no derivative of non-linear models at some point, for example

\[ f(x) = x^{1/3} \]  

2) the numeric stability of derivative functions is very bad, and the derivative functions are more sensitive to the variation of initial values of the parameters, leading to the emanation of iteration thereon, for example

\[ y = a_1 x^{\beta_1} + a_2 x^{\beta_2} \]

Linearizing the non-linear models may produce model errors and change the relations between parameters of the non-linear models, thus distorting the estimated value of parameters and the assessment accuracy. Only when the demanded accuracy for estimated parameters is not high and the non-linear model is stable, can we linearize the non-linear models. In order to further improve the accuracy of data processing, scholars have done many researches on non-linear adjustment\(^2\text{-}7\). But most of them focus on the resolution of distinction function.
2 Mathematic features of non-linear models

From the view of geodesy, non-linear models can be divided into two types: relation-structure models and random models. Relation-structure models are also of two types, definite functional models such as polar condition equation in geodetic quadrangle, and indefinite functional models such as the regression model in deformation data processing. There are different processing methods for different non-linear models, and the distinguishing of non-linear models depends on its mathematic features. This section will investigate the mathematic characteristics of non-linear models, in the meantime, analyse the bi-twist bicurve function.

Let 
\[ y = f(x, \theta) + \Delta \] (3)
where \( f \) are non-linear models; \( y \) and \( \Delta \) are the dependent variable vector and error vector, respectively, and \( E(\Delta) = 0; x \) is the independent variable vector; \( \theta \) is the unknown parameter vector.

2.1 Differential curvature
Assuming that \( f(x, \theta) \) has more than second order successive derivates. Selecting randomly \( \theta_i \) from \( \theta \) space, we can obtain a line \( l \) that passes \( \theta_0 \) in the direction of \( h \):
\[ l_i(\theta(b)) = \theta_0 + bh \] (4)
where \( b \) is the real parameter; \( h \) is an arbitrarily probable fixed direction; \( l \) maps to solution locus lying in sample space \( R^n \) by \( \eta = f(\theta) \), and a line \( C \) can be given as
\[ C_i(\eta) = \eta_i(b) = f(\theta_0 + bh) \] (5)
Here, \( C \) is called lift-line. Expanding \( C \) at \( \theta_0 \) into Taylor series, we have
\[ f(\theta_0 + bh) - f(\theta_0) = Vh + (h^T Wh)/2 + O(\|h\|^2) = Vh + (h^T Wh)^N/2 + (h^T Wh)^T/2 + O(\|h\|^2) \] (6)
where \( h = (\theta_0 + bh - \theta_0)/b, V = \partial f/\partial \theta, W = \partial^2 f/\partial \theta \partial \theta \). From above, we can get the first order and second order derivates of \( C \);
\[ \dot{\eta}_i = Vh, \ddot{\eta}_i = \ddot{\eta}_i^N + \ddot{\eta}_i^T \]
\[ h^T Wh = (h^T Wh)^N + (h^T Wh)^T \]
where \( (h^T Wh)^N \) and \( (h^T Wh)^T \) represent vertical component and tangent component of second order derivates, respectively, and \( h^T Wh = (h^T Wh)^N + (h^T Wh)^T \). So \( IN \) and \( PE \) of the model along \( h \) direction at \( \theta_0 \) are defined as
\[ IN = \frac{\|\ddot{\eta}_i^N\|}{\|\ddot{\eta}_i\|} = \frac{\|h^T Wh\|^N}{\|Vh\|^2} \] (7)
\[ PE = \frac{\|\ddot{\eta}_i^T\|}{\|\ddot{\eta}_i\|} = \frac{\|h^T Wh\|^T}{\|Vh\|^2} \] (8)
where \( IN \) and \( PE \) are invariables independent of coordinates. Choosing the maximum of \( IN \) and \( PE \), we can obtain
\[ IN_m = \max(IN)_h \] (9)
\[ PE_m = \max(PE)_h \] (10)
Because the curvature depends on the dimension of data and parameters, in order to eliminate these correlation, Eqs. (9) and (10) may be converted into maximum relative \( IN \) and maximum relative \( PE \) as follows
\[ IN^c = \frac{\sqrt{\max(IN)}}{\sqrt{m}} \] (11)
\[ PE^c = \frac{\sqrt{\max(PE)}}{\sqrt{m}} \] (12)
where \( \sqrt{\left(\frac{\Delta}{\Delta}\right)^2} \) and \( \sqrt{\frac{\Delta}{\Delta}} \) represent vertical component and tangent component of second order derivates, respectively, and \( h^T Wh = (h^T Wh)^N + (h^T Wh)^T \). Because the curvature depends on the dimension of data and parameters, in order to eliminate these correlation, Eqs. (9) and (10) may be converted into maximum relative \( IN \) and maximum relative \( PE \) as follows
\[ IN^c = \frac{\sqrt{\max(IN)}}{\sqrt{m}} \] (11)
\[ PE^c = \frac{\sqrt{\max(PE)}}{\sqrt{m}} \] (12)
where \( \delta = \sqrt{\frac{\Delta}{\Delta}} \) and \( \Gamma^c \) and \( \Gamma^t \) have no dimensions.

Given permitting curvature index \( \Gamma = 1/\sqrt{F} \) (Ratkowsky[5] suggested that \( \Gamma = 1/2/\sqrt{F} \), there is no difference but enlarging the permitting value), where \( F=F(m,n-m,a) \) and \( 1-a \) is the confidence level. By Comparing the permitting curvature indexes, there are three types of models;
1) \( \Gamma^c < \Gamma, \Gamma^t > \Gamma \), this type of model can be linearized;
2) \( \Gamma^c, \Gamma^t > \Gamma \), this kind of model must be reparameterized;
3) \( \Gamma^c > \Gamma, \Gamma^t < \Gamma \), this sort of model can not be linearized.

2.2 Successive and numeric stability
Lemma 1: Assuming that \( f(x, \theta) \) has its definition within neighbor-field of \( x_0 \), when \( \Delta x = x - x_0 \) approximates to zero, \( \Delta f = f(x_0 + \Delta x, \theta) - f(x_0, \theta) \) also slides to zero, then \( f(x, \theta) \) is successive at point \( x_0 \); if limit \( (\Delta f/\Delta x) \) exists when \( x_0 \rightarrow 0 \), this limit value is called
derivates of \( f(x, \theta) \) at point \( x_0 \). Most non-linear models consist of basic elementary functions, so within the definition field of each basic elementary functions, the non-linear model is successive. Being successive does not mean being derivative, but being derivate means being successive. The concept includes two aspects: one is that the non-linear model has no derivatives at certain a point, as for Eq. (1), if \( x = 0 \), the derivates of Eq. (1) does not exist and can not be linearized; the other is that the derivative function is sensitive to initial values, whose small changes may lead to great ones of function values. This is called “butterfly effects”, as for Eq. (2), it is treated as unable to be linearized\(^9\). It is found by tests that Eq. (2) has its arbitrary order derivates at any points.

Lemma 2: Assuming that for the differential equation \( \frac{dy}{dx} = f(x, \theta) \), let \( \epsilon > 0 \), always existing \( \delta > 0 \), when \( |x' - x_0| < \delta \), for all \( x \geq x_0 \), there is constantly inequality as below:

\[
| f(x_0, \theta) - f(x', \theta) | < \epsilon
\] (13)

Eq. (13) is regarded as the numeric stability of equation at point \( x_0 \). If there is no characteristics as above for non-linear derivative function , solutions of the equation will be very bad. For most adjustment models, most parameters have their sufficient approximate priori values. The stability of derivates function helps to investigate further the curvature measurement and solution properties of non-linear models.

2.3 Features test for stress non-linear model

For the bi-twist and bi-curve regression model\(^{[2]}\):

\[
y = \theta_1 + \theta_2 (x - \theta_4) + \theta_3 [(x - \theta_4)^2 + \theta_5]^{1/2} + \epsilon
\] (14)

the observed data is listed in Table 1. Making use of data in Table 1, we may solve curvature measurement index and LSE parameter values which are listed in Table 2.

| Parameter | \( \theta_1 \) | \( \theta_2 \) | \( \theta_3 \) | \( \theta_4 \) | \( \theta_5 \) | IN | PE | 1/2 \( \sqrt{\chi^2} \) |
|-----------|-------------|-------------|-------------|-------------|-------------|-----|-----|------------------|
| Value     | 136.82      | 0.696       | -0.586      | 7           | 18,752      | 6,727 | 0.434 | 1.812            | 0.307 |

From Table 2, we can know that model (14) is a stress non-linear function, whose PE is more evident than IN. Now test the successive derivates and numeric stability of model (14):

First: the parameters in Eq. (14) are replaced by corresponding values in Table 2. Differentiating the function, we can obtain

\[
y^{(1)} = 0.696 - \frac{0.586 (x - 18.752)}{\sqrt{(x - 18.752)^2 + 6.727^2}}
\] (15)

The denominator in Eq. (15) is always beyond zero, so for any \( x \), derivative Eq. (15) is also meaningful, successive thereon;

Second: within changeable field of independent variable, selecting randomly a point \( x_0 = 100 \), choosing a point \( x_0' = 100. 001 \) in the neighbor field of \( \delta = 0. 1 \), the increment of Eq. (15) is:

\[
| y^{(1)}(x'_0) - y^{(1)}(x_0) | = 2.5 \times 10^{-10}
\] (16)

From above, we have the following conclusions:

1) The stress non-linear model may have good numeric stability, that is to say, the stress non-linear model may be linearized;
2) Calculating PE and IN of a non-linear model is complicated;
3) For some non-linear models with complicated construction, especially implicit functions, the derivation is very difficult.

Table 1 Observed data

| \( x \) | \( y \) | \( x \) | \( y \) |
|--------|--------|--------|--------|
| 1      | 113.98 | 10     | 125.28 |
| 2      | 115.29 | 11     | 126.40 |
| 3      | 116.71 | 12     | 127.87 |
| 4      | 117.62 | 13     | 129.17 |
| 5      | 118.78 | 14     | 130.35 |
| 6      | 120.69 | 15     | 131.48 |
| 7      | 131.21 | 16     | 133.04 |
| 8      | 122.76 | 17     | 135.10 |
| 9      | 124.41 | 18     | 135.16 |

Table 2 Curvature measurement and parameter estimation values
3 Approaches to process evident PE and IN

The non-linearity caused by IN and PE can be weakened by increasing observations, decreasing the variance of residual error and parametric transformation.

3.1 Proof of weakening PE and IN by increasing observations and decreasing residual variance

Assuming there is an equation

\[ L = f(\hat{X}) \]  

(17)

where \( f(\cdot) \) is a non-linear function. On purpose of estimating \( X, L \) is observed \( P \) times repeatedly, then we may have

\[ L = (l_1, l_2, \ldots, l_P)^T, \]

\[ f(\hat{X}) = [f_1(\hat{X}), f_2(\hat{X}), \ldots, f_n(\hat{X})]^T. \]

Let \( \hat{X} = X_0 + \Delta X \), expand \( f(\hat{X}) \) into Taylor-series at \( X_0 \), and obtain the first and second order. Considering Eq. (6), we may have

\[ \| Vh \|_c = P(\partial X \partial f)^2 = P \| Vh \|^2. \]

(18)

\[ \| (h^T Wh)^T \|_c = \| P(\partial X \partial f)^T \| = \sqrt{P} \| (h^T Wh)^T \| \]

(19)

thereby we can have

\[ PE_{c} = PE/ \sqrt{P} \]

(20)

Letting \( \Delta_{i} = \Delta/k, k \geq 1 \)

(21)

and putting \( \Delta_{i} \) into \( \hat{\sigma}_{i}^2 \) expression, then we can obtain

\[ \hat{\sigma}_{i}^2 = \sigma_{i}^2/k \]

(22)

Putting Eq. (22) into Eq. (10), we have

\[ \Gamma_{i}^T = \Gamma_i/k \]

(23)

Eq. (20) and Eq. (23) show that increasing observations and decreasing the variance of residual error can weaken PE. According to the same idea, we can give the proof of weakening IN by increasing observations and decreasing residual variance.

3.2 Methods of parametric transformation

The parametric transformation can weaken PE. Some ways of parametric transformation have already been introduced in some works\(^{10}\). We supplement the following three kinds of ways:

- Bi-linear transformation

\[ x' = \frac{x + u}{1 + ux}, \quad |u| \neq 1 \]

(24)

- Box-Cox transformation

\[ x' = \begin{cases} \log u \quad u \neq 0 \\ \frac{x-1}{u} \quad u = 0, u \in (0,1) \end{cases} \]

(25)

- Exponent-logarithm transformation

\[ x' = \begin{cases} (ax + b)^{\lambda}, \quad \lambda \neq 0 \\ \ln(ax + b), \quad \lambda \neq 0 \end{cases} \]

(26)

For some concrete model, which parameters should be replaced? The following is the decision making and computing procedure:

1) According to Eq. (5)-Eq. (10), compute \( IN, PE \) and compare them with \( 1/2 \sqrt{F} (F = F(p, n - m - 1 - a) \) can be consulted in distributing table).

2) When \( IN < 1/2 \sqrt{F}, PE > 1/2 \sqrt{F} \), compute M. J. Box bias of each parameter in models with Eq. (27)-Eq. (28) and compare the results with 1%. An equation is given as:

\[ E(x - x) = -\frac{\sigma_{i}^2}{2} \cdot S \cdot tr[P] \]

(27)

where \( S = R^{-1} \), \( R \) is Schmidt decomposing matrix, and \( P \) is the curvature cubic matrix.

The ways to compute \( S, R \) and \( P \) are given in References\(^{11,12}\).

\[ \text{Bias}(\hat{x}_i) = [\text{Bias}(\hat{x}_i)/\hat{x}_i] \times 100 \% \]

(28)

3) We can only replace the parameters whose M. J. Box bias is bigger than 1%.

4) Keep computing, until PE of new parameters is less than 1%.

Because there are more evident differences between the solution of new parameters and that of original parameters, we should estimate the original parameters by use of estimating values of new parameters as initial values of original parameters.

4 Approach to process non-linear adjustment models

As we have seen, on the one hand, it is not convenient for us to investigate the PE and IN with adjustment model, on the other hand, for the stress non-linear model, transformation
of parameters do not change \( IN \). Therefore, if the non-linearity of a non-linear model stems from \( IN \), and its maximum relative \( IN \) or bias of residual error is beyond a given index, then the non-linear model is not fit for being linearized. For these reasons, we should look for a non-linear processing way in the space of non-linear models without investigating \( PE \) and \( IN \). Some approaches to parameter estimation such as minimum mean square error estimation (MMSEE), least square error (LSE), least absolute square error (LASE) and sum mean square error (SMSE) have been introduced in References [11-14]. However, all of them belong to linear approaches. This paper presents a new approach to process adjustment models in non-linear space from a new viewpoint.

4.1 Approach of parameter estimation

As we know, the error is defined as

\( A = X - \bar{X} \) \hspace{1cm} (29)

where \( A \) are the errors including system error, random error and gross error; \( X \) is the observational value; \( \bar{X} \) is the objective value of \( X \). Due to the difficulties in distinguishing and separating system error, random error and gross error, in addition, when estimating the surveying accuracy, we should consider the general precision which contains precision, accuracy and reliability. Assuming there is an adjustment model:

\[ L = f(X) + \Delta \]

\[ E(\Delta) = 0, \Delta = \Delta_s + \Delta_e \] \hspace{1cm} (30)

\[ \text{cov}(\Delta) = \sigma^2 P_{\Delta} \]

where \( f(\cdot) \) is a non-linear function; \( L \) is a vector of observations; \( \Delta \) is the parameter vector; \( \Delta_s \) is the real error vector, stochastic and undefined; \( P_{\Delta} \) is a weight matrix.

In order to obtain the parameters \( \Delta \) in the space of non-linear models, firstly we comprehend the two concepts as:

\[ \text{GPE} = E(X - \bar{X})^2 \] \hspace{1cm} (31)

where GPE means the general precision estimation; \( E(\cdot) \) is the expectation function. Then, we can give the estimation criterion of adjustment parameters as:

\[ \text{SGPE} = \min E(X - \bar{X})^2 \] \hspace{1cm} (32)

where SGPE means the sum general precision estimation; \( X \) and \( \bar{X} \) are all vectors.

In the discussion of the way to work out \( X \) in non-linear model space under the strict Eq. (32) by combining Eq. (30) with Eq. (32), we can have

\[ E(L - f(X))^2 = \min \] \hspace{1cm} (33)

Then, with the help of iteration, we can obtain \( X \) without derivation of the non-linear function \( f(X) \). Certainly a definite terminating criterion of iteration should be given according to the features of problems.

Here we prepare to discuss the \( \Delta \) and introduce automatic regression model AR(1) into the following formula:

\[ \Delta_i = p \Delta_{i-1} + e_i \] \hspace{1cm} (34)

where \( p \) is the scale factor; \( \{e_i\} \sim \text{iidN}(0, \sigma^2) \), then we have

\[ \Delta_{n+k} = e_{n+k} + \ldots + p^k e_i + p^{k+1} e_{i-1} + \ldots \] \hspace{1cm} (35)

\[ E(\Delta_i) = 0 \]

\[ E(\Delta_i^2) = \sigma^2 (1 + p^2 + p^4 + \ldots) = \sigma^2 / (1 - p^2) \] \hspace{1cm} (36)

\[ E(\Delta_i, \Delta_{i+1}) = p \sigma^2 / (1 - p^2) \] \hspace{1cm} (37)

Assuming that \( X = (x_1, x_2, \ldots, x_n)^T, Y = (x_2, x_3, \ldots, x_n)^T, x_i \) is the observational value, then

\[ \rho = \sum x_i x_{i-1} / \sum x_i^2 = (X^T X)^{-1}X^T Y, \]

\[ i = 2, 3, \ldots, n \] \hspace{1cm} (38)

\[ E(e_i^2) = \sigma^2 \sum (x_i - \rho x_{i-1})^2 / (n - 1), \]

\[ i = 2, 3, \ldots, n \] \hspace{1cm} (39)

Correlation factor

\[ R_i = \sum (x_i x_{i-1}) / \sum x_i^2, \]

\[ i = 2, 3, \ldots, n \] \hspace{1cm} (40)

Combining Eqs. (32), (36), (38) and (39), and assuming that

\[ Q = \sum [x_i - (Xj^T X)^{-1}X^T Y x_{i-1}]^2 / (n - 1) \]

we can have

\[ \sum (L_i - f_i(X))^2 / n = Q \] \hspace{1cm} (42)

where \( f_i(\cdot) \) is still a non-linear function. In order to avoid linearizing \( f_i(\cdot) \), we must make use of the iterative approach, and transform Eq. (32) into

\[ \Phi(X) = \sum (L_i - f_i(X))^2 / n - Q \] \hspace{1cm} (43)
4.2 Example

Assuming there is a non-linear model as
\[ y = \frac{\theta_3 x_1}{(1 + \theta_1 x_1 + \theta_2 x_2)} \]  
(44)
The known observed data are listed in Table 3.

| \(x_1\) | \(x_2\) | \(y\)  |
|--------|--------|------|
| 1.0    | 1.0    | 0.126|
| 2.0    | 1.0    | 0.219|
| 1.0    | 2.0    | 0.076|
| 2.0    | 2.0    | 0.126|
| 0.1    | 0.0    | 0.186|

Assuming that the gross error in original data has been well revised, then we have the error function like Eq. (44)
\[ v_i = y_i - \frac{\hat{\theta}_3 x_{1i}}{1 + \hat{\theta}_1 x_{1i} + \hat{\theta}_2 x_{2i}} \]  
(45)

According to Eq. (32), we can obtain the criterion function
\[ \text{SGEP}(\theta) = \min \sum_{i=1}^{5} (y_i - f(x_i, \theta))^2 \]  
(46)
where \( X = (x_1, x_2)^T, \theta = (\theta_1, \theta_2, \theta_3)^T \). Applying the data in Table 3 to Eq. (46), and model AR(1) to describe \( v_i \), we can obtain the iterative function;
\[ \Phi(\theta) = \sum_{i=1}^{5} (y_i - f(x_i, \theta))^2 / 5 - Q, \]  
where \( Q = 4,603.4 \). Select \( \theta_0 = (3, 5, 1)^T \). The estimated values of parameters calculated by iteration are listed in Table 4, and the LS estimated values in linear model space are also listed in Table 4.

| Parameter | Linear | Non-linear | Difference |
|-----------|--------|------------|------------|
| 0\(_1\)   | 3.13   | 3.10       | 0.03       |
| 0\(_2\)   | 15.16  | 15.24      | -0.08      |
| 0\(_3\)   | 2.44   | 2.39       | -0.05      |

5 Conclusions

1) It is very complicated to investigate the PE and IN of a non-linear model;
2) Not all stress non-linear models can be linearized;
3) It is a good way to find an approach to process non-linear adjustment models in the non-linear models space;
4) Real error may be described in the model of AR(P), according to Akaike's information criterion(AIC) to determine parameter P;
5) To obtain the resolution of model (30), we have to make use of iteration.

References

1 Bates D M, Watts D G (1980) Relative curvature measures of non-linearity. Journal of Royal Statistical Society, 42(1): 1-25
2 Teunissen P J G (1989) First and second moments of non-linear least-squares estimators. Bulletin Geodesique, 63: 253-262
3 Teunissen P J (1985) The geometry of geodetic inverse linear mapping and non-linear adjustment. In: Netherlands Geodetic Commission, eds. Publications on Geodesy, New Series, 8(1)
4 Teunissen P J (1995) Nonlinear and least-squares. CISM Journal ACSCG, 42(4): 321-330
5 Teunissen P J (1991) On the local convergence of the iterated extended Kalman filter. In: Proceedings XX General Assembly of the IUGG, IAG Section IV, 177-184
6 Teunissen P J G (1990) Nonlinear least-squares. Manuscripta Geodaetica, 15: 137-150
7 Teunissen P J (1989) Nonlinear adjustment. Hotine-Marussi Symposium on Mathematical Geodesy, Pisa.
8 Ratkowsky D A (1986) Hong Z J et al. translation, non-linear regression model. Nanjing: Nanjing University Press.
9 Weisberg S (1998) Wang J L et al. translation, applied linear regression. Beijing: China Statistic Press.
10 Li Q H, Tao B Z (1982) Probability statistic theory and its applications in surveying. Beijing: Publishing House of Surveying and Mapping. (in Chinese)
11 Wei B C (1989) Modern non-linear regression analysis. Nanjing: East-Southern University Press. (in Chinese)
12 Wang Z Z, Yang K Q, Zhu J J (1999) Investigating linear approximate condition in non-linear models. Journal of Central South University of Technology, 30(3): 25-30 (in Chinese)
13 Bates D M, Watts D G (1997), Wei B C, et al. translation, Non-linear regression analysis and its application. Beijing: China Statistic Press.
14 Blaha G (1994) Non-iterative approach to non-linear least-squares adjustment. Manuscripta Geodaetica, 19: 199-212