

# Study on the Model and Errors in Data Fitting

Ying Cao

Northeast Petroleum University  
Qinhuangdao, Hebei, 066004 China

**Abstract:** To linearize the several commonly nonlinear function by using variable substitution method, using the least square method and the total least squares method to observed curve fitting data after linearized the function, results show that the fitting precision of curve function that use the total least squares method is higher than that of the least square method. Therefore, the fitting error is very important, and the least square method for curve fitting is the most commonly used method, because of its more accurate, practical and widely used.

**Keywords:** *Least square method; curve fitting; Matlab; accuracy;*

## I. INTRODUCTION

In many practical engineering application and scientific experiments, people need from a set of measured Co. data points, try to find a continuous smooth curve, to make sure all the value simulation function. At present, seek common analog function mainly has the interpolation and fitting. Among them, the interpolation method is applied to the data accuracy and data volume small, otherwise it may cause error in the local large. Fitting method does not require simulation function must be known by all, it is the pursuit of all point to the simulation function to some error minimization, and therefore it is suitable for data error and the larger of data.

### *The basic idea of total least squares*

At present, the least square method is the most commonly used method for processing data. However, this

algorithm itself has some defects, which only considers the processing error of observation vector. However, the error is inevitable. The observation vector, the coefficient matrix may contain errors, errors in data processing how to deal with the coefficient matrix is a very meaningful work. How to simultaneously consider variables, the dependent variable error, error and how to deal with the coefficient matrix and observation vector, the suitable methods to solve this kind of problem is the total least squares (3-4). In a linear regression model as an example to illustrate the basic framework of the total least squares. Consider the linear observation system is:

$$y_i = a + bx_i \quad (1)$$

Type: A, B is theory of regression coefficient Xi is variables. Considering the variable Xi and dependent variable Yi with error, then the observation equation corresponding is:

$$y_i + v_{yi} = a_0 + \delta_a + (b_0 + \delta_b)(x_i + v_{xi}) \quad (2)$$

Wherein:  $a_0, b_0$  is approximate value of  $a, b$ , can select two or three data optionally known observation data to

obtained, in order to meet the  $\min_{v_{xi}, v_{yi}} \|(v_{xi}, v_{yi})\|_F$  ( $\|\cdot\|_F$

represent norm). Get the value of  $\delta_a, \delta_b$ . Type (2) is the basic model of total least squares.

## II. LS, TLS CURVE FITTING

### 2.1 Linearization of nonlinear function

The nonlinear function into a linear function, usually by Taylor series expansion and variable substitution method. By linearizing the nonlinear current curve equation by variable substitution method in general, and the curve equation of general index equation  $y = ae^{bx}$ , gauss equation  $y = ae^{-2(x-b)/c}$ , polynomial equation  $y = ax^3 + bx^2 + cx + d$  and power function equation ( $y = ae^b$ ). By using variable substitution method, can convert the nonlinear model is the linear function model.

$$\begin{cases} \ln y = \ln a + bx \rightarrow Y = A + BX \\ \ln y = -\frac{1}{c^2}x^2 + \frac{2b}{c^2}x + \ln a - \frac{b^2}{c^2} \rightarrow Y = AX + BZ + C \\ y = ax^3 + bx^2 + cx + d \rightarrow Y = AX_3 + BX_2 + CX_1 + D \\ \ln y = \ln a + b \ln x \rightarrow Y = A + BX \end{cases} \quad (3)$$

Through variable substitution, curve equation can be changed into a univariate or multivariate linear equations.

### 2.2 LS curve fitting

The least square method is the most commonly used method of curve fitting [1]. Assume that there is a set of discrete points  $(x_i, y_i)$ , In do not completely determine the curve model, the general choice of polynomial curve fitting  $y = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$ , By type (3) variable substitution, the linear model:

$$y = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n \quad (4)$$

Wherein:  $X_1 = x, X_2 = x^2, \dots, X_n = x^n$

Type (4) (1) by the indirect adjustment model to get parameter estimation:

$$\begin{bmatrix} a_0 & a_1 & a_2 & \dots & a_n \end{bmatrix}^T = (B^T B)^{-1} B L \quad (5)$$

$$B = \begin{bmatrix} 1 & X_{11} & \dots & X_{1n} \\ 1 & X_{21} & \dots & X_{2n} \\ \dots & \dots & \dots & \dots \\ 1 & X_{n1} & \dots & X_{nn} \end{bmatrix}; \quad L = y$$

Wherein:

The above is the polynomial curve fitting based on least square algorithm, this algorithm uses simple, which have been widely used.

### 2.3 TLS curve fitting

The total least squares method considering variable and dependent variable error, In the type  $y = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$ ,  $X, y$  also contains errors

$V_x, V_y$ , So the type (4) turn into:

$$y + V_y = a_0 + a_1(X_1 + V_{x1}) + a_2(X_2 + V_{x2}) + \dots + a_n(X_n + V_{xn}) \quad (6)$$

The type (6) transform into orthogonal form:

$$A_1(X_1 - \bar{X}_1) + A_2(X_2 - \bar{X}_2) + \dots + A_n(X_n - \bar{X}_n) + A_0(y - \bar{y}) = 0 \quad (7)$$

Wherein:

$$\bar{X}_1 = \frac{1}{n} \sum_{i=1}^n X_1^i, \quad \bar{X}_2 = \frac{1}{n} \sum_{i=1}^n X_2^i, \dots, \quad \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_n^i;$$

$$y = \frac{1}{n} \sum_{i=1}^n y_i;$$

$$a_1 = -\frac{A_1}{A_0}, \quad a_2 = -\frac{A_2}{A_0}, \quad a_n = -\frac{A_n}{A_0}$$

$$a_0 = \frac{A_1}{A_0} \bar{X}_1 + \frac{A_2}{A_0} \bar{X}_2 + \dots + \frac{A_n}{A_0} \bar{X}_n + \bar{y}$$

The type (7) eigenvalue decomposition by using the total least squares solution formula:

$$M^T M = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix} \begin{bmatrix} u_1^T \\ u_2^T \\ \dots \\ u_n^T \end{bmatrix} \quad (8)$$

$$\sigma_1 > \sigma_2 > \dots > \sigma_n;$$

$$\text{Wherein: } M = \begin{bmatrix} X_1 - \bar{X}_1 & X_2 - \bar{X}_2 & \dots & X_n - \bar{X}_n \end{bmatrix}$$

$[A_1, A_2, \dots, A_n]^T$  take the corresponding feature vector of  $\sigma_n$ , Thus get type (7) parameters, the parameters turn the inverse substitution, can be obtained coefficients of polynomial curve equation  $a_0, a_1, a_2, \dots, a_n$ . Finally, we get the polynomial curve fitting equation based on total least squares.

### III. THE APPLICATION EXAMPLE OF CURVE FITTING

In the friction experiments, the equivalent voltage is -285V, servo motor after opening, the relationship between time and speed as shown in Table 1, the fitting curves and 200ms for speed and time of acceleration of the.

Table 1.10 to 500ms value of speed

Time/ms <sup>+</sup>	10 <sup>+</sup>	30 <sup>+</sup>	50 <sup>+</sup>	70 <sup>+</sup>	90 <sup>+</sup>	110 <sup>+</sup>	130 <sup>+</sup>	150 <sup>+</sup>	170 <sup>+</sup>
Speed/(m/s) <sup>+</sup>	0 <sup>+</sup>	0.066 <sup>+</sup>	0.088 <sup>+</sup>	0.1539 <sup>+</sup>	0.1979 <sup>+</sup>	0.2419 <sup>+</sup>	0.2639 <sup>+</sup>	0.3079 <sup>+</sup>	0.3299 <sup>+</sup>
Time/ms <sup>+</sup>	190 <sup>+</sup>	210 <sup>+</sup>	230 <sup>+</sup>	250 <sup>+</sup>	270 <sup>+</sup>	290 <sup>+</sup>	310 <sup>+</sup>	330 <sup>+</sup>	350 <sup>+</sup>
Speed/(m/s) <sup>+</sup>	0.3738 <sup>+</sup>	0.3958 <sup>+</sup>	0.4398 <sup>+</sup>	0.4398 <sup>+</sup>	0.4838 <sup>+</sup>	0.4838 <sup>+</sup>	0.5498 <sup>+</sup>	0.5498 <sup>+</sup>	0.5718 <sup>+</sup>
Time/ms <sup>+</sup>	370 <sup>+</sup>	390 <sup>+</sup>	410 <sup>+</sup>	430 <sup>+</sup>	470 <sup>+</sup>	470 <sup>+</sup>	490 <sup>+</sup>	<sup>+</sup>	<sup>+</sup>
Speed/(m/s) <sup>+</sup>	0.5938 <sup>+</sup>	0.6158 <sup>+</sup>	0.6158 <sup>+</sup>	0.6377 <sup>+</sup>	0.6597 <sup>+</sup>	0.6597 <sup>+</sup>	0.6597 <sup>+</sup>	<sup>+</sup>	<sup>+</sup>

(1) According to the known data, take the time as the abscissa, speed as ordinate, to draw the schematic diagram of data points in Matlab, as shown in figure 1. The code is:

$X=0.01, 0.02, 0.49$

$Y=[0, 0.0660, 0.0880, 0.1539, 0.1979, 0.2419, 0.2639, 0.3079, 0.3299, 0.3738, 0.3958, 0.4398, 0.4398, 0.4618, 0.838, 0.5498, 0.5498, 0.5718, 0.5938, 0.6158, 0.6377, 0.6597, 0.6597]$

Plot(x,y,'r \*')

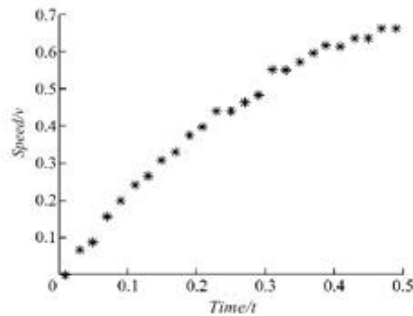


Fig.1 Schematic diagram of data points

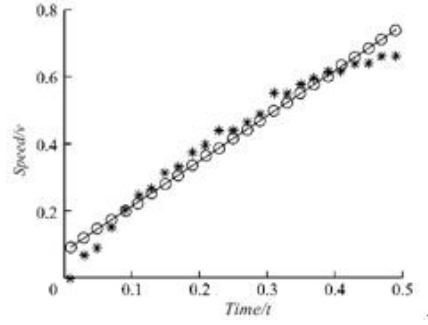


Fig.2 The chart of linear fitting

According to physical knowledge, when the acceleration is constant, time and speed is proportional to. But you don't know how to change the acceleration, so the time and speed may not be linear, also from the data points in Figure 1 did not show obvious mathematical rules, so this was linear, polynomial in two times, three times, and five times, the power function, logarithm function curve fitting, to obtain the optimal fitting from.

(2) For curve fitting to each function model

A) Linear fitting: Mat lab code is as follows:

Hold on

$aa=polyfit(x,y,1);$

$plot(x,polyval(aa,x),'b-o');$

$a=y-polyval(aa,x);$

$e=a*a';$

So, the linear fitting function:  $y=1.3524x+0.0753$

Error sum of square:  $e=0.0393$

B) polynomial two fitting, the fitting function:

$y=-2.0524x^2+2.3786x-0.0103$ , Error sum of square:  $e=0.003$ , fit the pattern is as shown in figure 3.

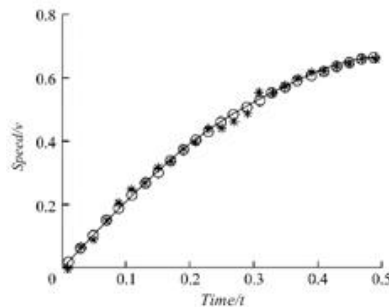


Fig.3 The polynomial 2 times fitting

C) By the polynomial 3 times fitting, we get fitting function:  $y=0.8401x^3-2.6825x^2+2.5048x-0.0155$ , error sum of square:  $e=0.0294$ , fit the pattern is as shown in figure 4.

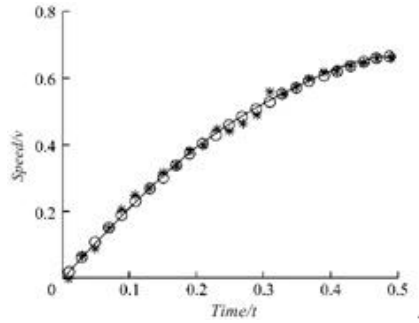


Fig.4 The polynomial 3 times fitting

D) For the logarithmic function fitting, model function is  $y=a+b*\log x$ , set  $t=\log x$ . Therefore, the logarithm model function turn to fit the linear function  $y=a+b*t$ . In the Matlab implementation, fitting function  $y=0.2028*\log x+0.7539$ , error  $e=0.0885$ , fit pattern as shown in figure 5.

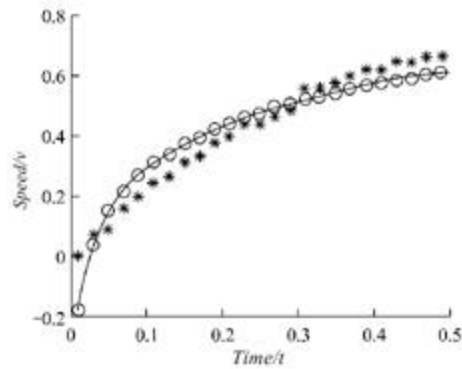


Fig.5 The logarithmic function fitting

E) power function fitting, model function  $y=ax^b$ , both logarithmic:  $\log y=\log a+b*\log x$ ,  $k=\log y$ ,  $t=\log x$ . Therefore the exponential function is converted into linear function of the index function to fit linear functions of  $k=\log a+b*t$ , the realization in Matlab, the fitting function  $y=1.3647x^{0.8339}$ , graphical error square and  $e=0.0273$  fit in figure 6.

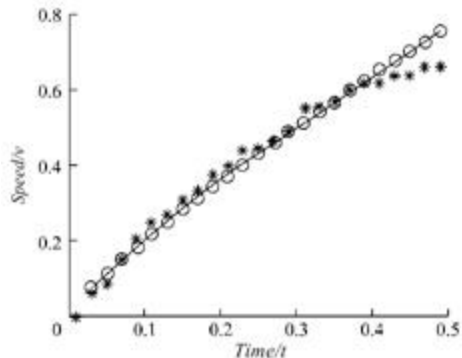


Fig.6 The exponential function fitting

### (3) The fitting function evaluation

You can see from the fitting error square polynomial in two times, three times is the smallest. Therefore, two kinds of curve fitting is optimal, but because the polynomial fitting the higher the number, the greater the amount of calculation, if too high, the noise is incorporated in the model, result in more error combined with the actual! And in error is very small circumstances, tend to choose the two functions to fit polynomial. So the fitting function derivative motor acceleration equation:

$$y' = -4.1048x + 2.3786$$

Therefore, acceleration can be obtained in 200ms is 1.5576m/s<sup>2</sup>.

## IV. CONCLUSIONS

The least square method is widely used in engineering application and scientific experiments, the accuracy of fitting curve is very important. In this paper, through an example of using least squares linear and nonlinear fitting of the known data, with the help of Matlab during the powerful computer tools to achieve the whole process of fast, accurate, and lays a foundation for further study of curve fitting.

## REFERENCE:

- [1] Li Hongwei, Wei Shaochun, Chen Anping, etc.. Application of total least squares in straight line fitting [J]. Geological mineral surveying and mapping, 2010,26 [2] 4-5
- [2] Ding Keliang, ou Ji Kun., Zhao Chunmei. Orthogonal least squares curve fitting [J]. Science of Surveying and mapping, 2007,32 [3]: 17-19.
- [3] Liu Yonghui Wei Musheng. Comparison of TLS and LS problems[J]. Computational mathematics, 2003,25 [4]: 479-492
- [4] Zhu Xiaodong, Lu will, Chen Xijiang. The orthogonal polynomial curve fitting[J]. Journal of Donghua Polytechnic University: Natural Science Edition, 2010,33 [4]: 398-400.