Piecewise Curve Fitting Based on Least Square Method in 3D Space

Lihong Xue*

Department of Mathematics, Jining Normal University, Wulanchabu City, Inner Mongolia Autonomous Region 012000, China. E-mail: 997140306@qq.com

Abstract: Least squares curve fitting is widely used in various neighborhoods, such as industry, agriculture, and economy. The method of piecewise curve fitting is used to deal with some experimental studies with large amounts of data. It can not only be used for plane space points, but also for three-dimensional space points. Firstly, this article introduces the principles of curve fitting and least squares. Secondly, it clarifies the steps based on least square fitting and gives examples of improvements made based on the principle of least square fitting functions by some scientists. Finally, the piecewise curve fitting in the three-dimensional space is obtained using the least squares piecewise curve fitting. It lays a certain theoretical foundation for drawing the laws between data in three-dimensional space.

Keywords: Least Squares; Curve Fitting; Piecewise Curve Fitting; Three-dimensional Space

1. Introduction

Least square method is proposed to optimize data, which compares the square of the deviation between the estimated value (function) and the actual value (function) to determine a data function that best matches the data. There are two scientists with outstanding contributions in the creation of the least square method: Legendre first introduced the least square method in the idea of solving linear equations; Gauss made a detailed elaboration of the theory of least square method using the normal distribution. Curve fitting based on least squares is common in data processing, and the segmented fitting method is more reflective of the actual data relationship than the traditional direct fitting method. So far, there are many applications in the experiment, and the better segmentation methods are: (1) Xie Youbao's method of using the angle between the lines to be less than a given threshold for data segmentation. (2) Liu Zongtian et al. proposed to use the deviation between the sum of squares of two errors for segmentation. However, these are all data processing that stay on the plane point. However, in practical applications, some data processing of three spatial points often exist (for example, data in the fields of physical experiments, astronomical observations, and cultural relics protection). Therefore, it requires to discuss the fitting of piecewise curves in points in three-dimensional space based on the fitting of piecewise curves in two-dimensional space to facilitate the optimization of data processing.

2. The principle of curve fitting and least square method

The experimental data obtained through experimental research or investigation assumes a functional relationship, and then determines the coefficients of each function in the function through methods such as formula deformation, that is, to obtain an empirical formula representing the quantitative relationship through the relevant data. From a geometric perspective, the relationship between these data can be expressed as a curve. Therefore, the curve fitting process is the process of using relevant data to find the empirical formula.

The least square method is most commonly used in curve fitting, whose central idea is to make the deviation between the experimental observation data and the estimated data (geometrically expressed as the vertical distance

between the experimental observation data point and the estimated data point). The sum of squares is minimized, which is to avoid the occurrence of too small errors or too large errors, so that the estimated values of the parameters play a practical role. The principle is as follows:

Given a hypothetical set of data on independent and dependent variables (where $i = 1, 2, 3, \dots, n$). Use this curve model to fit the given data. Given the independent variable, the deviation of the selected (estimated) from the actual dependent variable can be obtained. To minimize the deviation of each data, then the sum of all deviations should be minimal, because it may be positive or negative. In order to avoid canceling each other when adding and affecting the reflection of the real situation of the system, the value of their sum of squared deviations can be minimized to get the corresponding parameters. The fitting curve at this time is optimal.

3. The basic steps of curve fitting

3.1 Seek empirical formula

Seek an empirical formula so as to determine the function model of the fitted curve. When applying curve fitting in practice, it is difficult to determine what relationship exists between the various data. It may be a relatively regular linear relationship, which can be expressed by a straight line equation; however, if it is a nonlinear relationship, it is divided into an exponential relationship, a logarithmic relationship, and a composite relationship. Therefore, seeking empirical formulas is a major difficulty in fitting curves.

In general, the observation method will be adopted when seeking empirical formulas, that is, use mathematical expertise and related software to roughly determine which functional relationship exists between the data, and find the optimal and most likely functional relationship from it. This method is relatively simple, and more intuitively shows the relationship between the data.

Next, look at an example of using the observation method to determine the functional relationship. For example: give a set of physical experiments to measure the elasticity of the spring. The spring force is $F(N)$, the spring coefficient is K , and the length of the spring is $X(cm)$. Observation of data scatter plot 1 shows that $F = KX$ is satisfied within a certain elastic limit. At this time, the spring extension distance X and the spring force F have a linear relationship, but once it exceeds a certain value, it no longer conforms to the previous law.

X(cm)	2	3	4	5	6	7	7.5	7.5
F(N)	0	0.5	1	1.5	2	2.5	3	4

Table 1. Experimental data

Figure 1. Scatter plot of experimental data of elastic limit.

3.2 Calculate the minimum value of the sum of squared errors

From the above principle of least squares, it can be known that when the optimal curve equation is determined, it requires to solve the minimum sum of square errors. If (x_i, y_i) falsewhere $i = 1, 2, 3, \dots, n$, Φ is the function type whose

degree is less than n. Now find $f_k(x_i) = \sum_{k=0}^m a_k x_i^k \in \Phi$, and make the deviation equal to the minimum Q.

$$Q = \sum_{i=0}^n [f_k(x_i) - y_i]^2 = \sum_{i=0}^n \left(\sum_{k=0}^m (a_k x_i^k - y_i) \right)^2$$

When the partial derivative of Q is equal to zero, Q is the smallest.

Therefore:

$$\frac{\partial Q}{\partial a_j} = 2 \sum_{i=0}^n \left(\sum_{k=0}^m (a_k x_i^k - y_i) x_i^j \right) = 0, \quad j = 1, 2, 3, \dots, m$$

Thus $\sum_{k=0}^m \left(\sum_{i=0}^n (x_i^{k+j}) \right) a_k = \sum_{i=0}^n x_i^j y_i, \quad j = 1, 2, 3, \dots, m$, The value of a_k thus obtained

determines the fitting function.

For $f_k(x_i) = \sum_{k=0}^m a_k x_i^k$, When the exponent k is 1, the function is a linear function. When $k > 1$, it is a nonlinear fit. If other types of nonlinear functions are encountered during data processing, they can be converted into linear functions and then calculated based on the above ideas. Of course, some data processing tools also can be used, like Matlab's own function polyfit automatic calculation, EXCEL data analysis and SPSS software.

3.3 Comparison of curve fitting performance

When the empirical formula is difficult to determine, or it is difficult to observe the relationship between the data, the number of function models can be estimated through actual calculation and comparison, and the optimal function model can be obtained according to the principle of least squares. Generally when the square sum of deviation is the smallest, it is the optimal function model, and the fitting effect is the best. Of course, some researchers will use the value of the correlation coefficient R when processing data to determine whether it is the optimal function model.

4. Development of the least squares piecewise curve fitting

4.1 Principle of piecewise curve fitting

In actual research, when analyzing the scatter plots of the experimental data, there are some large quantities and the order of complex data is too high. Directly fit may lead inaccurate effect. Therefore, in order to improve the accuracy and effect of complex data processing, the method of piecewise curve fitting is often used to process the data. The basic principle of segment fitting is to determine the number of segments and the function type corresponding to each segment according to the distribution characteristics of the scatter plot of the experimental data, and then use least squares method for curve fitting and solving the coefficients of the fitting equation.

4.2 Segmentation based on least square method: proposal of line fitting scheme

Based on the basic principle of least squares, many scientists have improved it, and the method of segmentation to select and fit is also incorporated into it. E.g:

(1) Give data about independent variable x_i and dependent variable $y_i(x_i, y_i)$ where $(i=1, 2, 3, \dots, n)$, and divide them into groups of data, respectively $N_1, N_2, N_3, \dots, N_k$:

$$\begin{aligned} &(x_{11}, y_{11}), (x_{12}, y_{12}), \dots, (x_{1N_1}, y_{1N_1}) \\ &(x_{21}, y_{21}), (x_{22}, y_{22}), \dots, (x_{2N_2}, y_{2N_2}) \\ &\dots \\ &(x_{k1}, y_{k1}), (x_{k2}, y_{k2}), \dots, (x_{kN_k}, y_{kN_k}) \end{aligned}$$

Then fit the straight line $y=f(x_i)=a_i x+b$ to the above data to get the straight line $(i=1, 2, 3, \dots, k)$; then proceed to piecewise curve fitting. The above is the algorithm proposed by Zhang Donglin. Although his algorithm embodies the

idea of segmentation, it does not specify the values of k and $N_j(i=1,2,3,\dots,k)$, and when segmenting data, it is too subjective and lacks scientific basis.

(2) Set an initial point and end point as $A(x_a, y_a)$ and $B(x_b, y_b)$ respectively, and then take another new point $C(x_c, y_c)$. Use the least square method to fit the two straight line equations, and find the angle between the two straight lines AB and AC:

$$\angle BAC = \arccos\left(\frac{(L1)^2 + (L2)^2 - (L3)^2}{2 * L1 * L2}\right)$$

Among them, suppose the length of AB is $L1$, the length of AC is $L2$, and the length of BC is $L3$. After the value of this angle is obtained, it is compared with the given threshold. If it is greater than the given threshold, the straight line is directly fitted by the least square method. If it is less than the threshold, then point B and point C will be included in the subsequent data and recalculated until it is greater than the given threshold. The algorithm is more praised by Xie Youbao's piecewise curve fitting scheme. However, the limitation of this method is that he can only get the optimal function model in local data processing, and cannot represent the optimal function of the overall data model.

5. Plane straight line

Through least square method, the segmented curve is proposed in three-dimensional spatial transition.

5.1 The smallest plane straight line, square fitting

First, give the space standard straight line equation:

$$\frac{x - x_0}{m} = \frac{y - y_0}{n} = \frac{z - z_0}{p}$$

Transform the above equation to obtain the following two equations:

$$x = \frac{m}{p}(z - z_0) + x_0 = k_1 z + b_1$$

$$y = \frac{n}{p}(z - z_0) + y_0 = k_2 z + b_2$$

So $k_1 = \frac{m}{p}$, $b_1 = x_0 - \frac{m}{p}z_0$, $k_2 = \frac{n}{p}$, $b_2 = y_0 - \frac{n}{p}z_0$

The straight line can be regarded as the result of the intersection of two planes, then the straight line fitting problem can be transformed into the fitting problem of two intersecting planes. Use the least square method to fit the above two straight line equations:

First get the sum of squared residuals as:

$$Q_1 + Q_2 = \sum (x_i - k_1 z_i - b_1)^2 + \sum (y_i - k_2 z_i - b_2)^2$$

According to the residual and the minimum:

$$\sum 2(x_i - k_1 z_i - b_1) \times (-z_i) = 0$$

$$\sum 2(x_i - k_1 z_i - b_1) \times (-1) = 0$$

$$\sum 2(y_i - k_2 z_i - b_2) \times (-z_i) = 0$$

$$\sum 2(y_i - k_2 z_i - b_2) \times (-1) = 0$$

Finally, the parameters in the formula are:

$$k_1 = \frac{2 \sum_{i=1}^n x_i z_i - \sum_{i=1}^n x_i \times \sum_{i=1}^n z_i}{2 \sum_{i=1}^n z_i^2 - \sum_{i=1}^n z_i \times \sum_{i=1}^n z_i}$$

$$b_1 = \frac{\sum_{i=1}^n x_i - k_1 \sum_{i=1}^n z_i}{2}$$

$$k_2 = \frac{2 \sum_{i=1}^n y_i z_i - \sum_{i=1}^n y_i \times \sum_{i=1}^n z_i}{2 \sum_{i=1}^n z_i^2 - \sum_{i=1}^n z_i \times \sum_{i=1}^n z_i}$$

$$b_2 = \frac{\sum_{i=1}^n y_i - k_2 \sum_{i=1}^n z_i}{2}$$

5.2 Three-dimensional empty

Piecewise curve fitting refers to Xie Youbao's scheme of piecewise curve fitting, and substitutes the three-

dimensional space points into it to get the rules between the data in three-dimensional space. First, take the two points $S(x_1, y_1, z_1)$ and $N(x_2, y_2, z_2)$ in the data to be processed, use the least square method to fit the mathematical model of the line $E(x_3, y_3, z_3)$, and then take a point S to calculate the cosine of the angle between SN and NE, namely

$$\cos\theta = \frac{|a(x_3 - x_2) + c(y_3 - y_2) + (z_3 - z_2)|}{\sqrt{a^2 + c^2 + 1} \times \sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2 + (z_3 - z_2)^2}}$$

Second, compare the cosine of the included angle with a given threshold. If the cosine of the included angle is greater than the given threshold, the least square method is used to obtain the fitting curve of the first three-point space. Conversely, if the cosine of the included angle is less than the given threshold, then point N and point E are put into the previous data and recalculated together. Finally, repeat the above steps for point C and the next point until all points are involved. Finally, a piecewise fitting curve of a three-dimensional space point is obtained.

5. Conclusion

In summary, the method of piecewise curve fitting is better than the method of single curve fitting. In other words, in order to obtain the optimal fitting curve, it is necessary to master the optimal function segmentation method. Among the existing segmentation fitting methods, some are too subjective and do not reflect the data relationship well, while some others can only show advantages in local data processing, and cannot represent the overall situation. Therefore, it is possible to improve the curve segmentation fitting method for these problems and scientifically segment the data, that is, the data at the segments are connected in order on the graph. Of course, in order to increase the utilization range of curve fitting in actual production and life, the above ideas and methods can be transitioned into three-dimensional space to discover the rules between the data corresponding to the three-dimensional space points.

Acknowledgement

Project source: Higher educational scientific research projects of Inner Mongolia Autonomous Region.

Project name: Research on piecewise curve fitting method in the three dimensional space (No. NJZY17373)

References

1. Xu H, Xue W, Cheng X. A new method of fitting straight lines in 3D space (in Chinese). *Bulletin of Surveying and Mapping* 2015; 462(9): 32-35. doi: CNKI:SUN:CHTB.0.2015-09-008.
2. Xue L. Piecewise straight line fitting method based on least square method in three-dimensional space points (in Chinese). *Journal of Qiqihar University (Natural Science Edition)* 2015; (4): 87-88+92. doi: CNKI:SUN:QQHE.0.2015-04-021.
3. Chen L, Zheng Y. Study on curve fitting based on least square method. *Journal of Wuxi Vocational and Technical College* 2012; 11(5): 52-55. doi: 10.3969/j.issn.1671-7880.2012.05.017.
4. Liu X, Wang Y. Research of automatically piecewise polynomial curve-fitting method based on least-square principle. *Science and Technology, Engineering* 2014; 14(3): 55-58. doi: CNKI:SUN:KXJS.0.2014-03-013.