Research on Piecewise Linear Fitting Method Based on Least Square Method in 3D Space Points

Lihong Xue[*] and Dongping Li

Department of mathematics, JiNing Normal University, Wulanchabu City, 012000, Inner Mongolia Autonomous Region, China

Abstract: Because the known data cannot be completely found to determine the function of the corresponding relationship, it is needed to describe the data in tabular form. As the research goes on, people want to use some simple mathematical functions to express the relationship between the data and the data. In the process of data processing, the least square method is one of the most common methods. Based on the least square method, the piecewise linear fitting algorithm is proposed to replace the polynomial curve fitting. This paper firstly discusses the plane segmentation method to be extended to three dimensional discrete point; followed by the discrete points in the Euclidean space square piecewise linear fitting method to analyze the inherent law between the 3D data, through the experiment shows the method is effective.

Keywords: Data processing, least square method, linear fitting, spatial discretization.

1. INTRODUCTION

The application of fitting is very extensive, such as in the field of econometrics can use the economic statistical data over the years to predict the tendency of the economic development of the next stage, through the discovery of quantitative relationship between some toxic dose and mortality of animals to direct the medical work, in sensing technology can be used for multi sensor measurement calibration in the field of medical statistics [1, 2]. The common curve fitting method is the least square method. Least square method is a mathematical optimization technique, which is based on the square of the error and the best matching of data. Using the least square method, the unknown data can be easily obtained, and the square of the error between the data and the actual data is obtained [3]. When using the method of least squares curve fitting, although can well fitting in an orderly manner, and a good mathematical theory and very high accuracy, but need to calculate a large number of complex, and the obtained fitting function is often higher, which further increased the complexity. In order to simplify the establishment of the mathematical model of the algorithm and reduce the amount of computation, this paper proposes a new use least square method implement piecewise linear fitting method, with the observed ordered pairs of modeling [4-7].

The least square method is a mathematical optimization technique, which is used to infer the best function matching relationship between the discrete data by the square of the difference between the measured value and the actual value [8]. In engineering technology and scientific experiment, the data set is often obtained [9]. How to determine their function curve according to the data, this is the problem of curve fitting in experimental data processing. Curve fitting method is the most commonly used method, and the fitting of the piecewise curve is better than the general curve fitting effect [10-12].

So far, a lot of practical curve fitting methods have been put forward. In 1966 York D proposed the use of least squares fitting line [13]. In 1987 Arun K S fitted the two dimensional point set with the least square method [14]. In 1999 Fitzgibbon A, Pilu M, Fisher R B realized the ellipse direct least squares fitting [15].

2. SPACE LINEAR LEAST SQUARES FITTING PRINCIPLE

Data fitting is an important method of data processing, and polynomial curve fitting is a more common method of data fitting. When the data point is too large, the polynomial order is too low, the fitting precision and the effect is not very ideal. In order to improve the fitting precision and effect, we need to increase the number of curves, but the complexity of the order is too high and the complexity of the calculation and other aspects of the disadvantages. Therefore, it is difficult to obtain good fitting accuracy and effect by using a polynomial curve fitting of the data points. In order to solve the above problems, the piecewise curve fitting is generally used. In the past, the fitting method of piecewise curve fitting method is mainly aimed at the measurement of data in the field of natural science. Therefore, in fitting of these data, the traditional segmentation method is generally based on the subjective experience of the data segmentation, and then fitting [16]. But for some practical problems of data, such as social and economic life of a large amount of statistical data, the mechanism of these changes in data is very complicated in general, often like the laws of physics as a strict rules, so changes in uncertainty is very strong. Therefore, the

traditional segmentation curve fitting based on the subjective experience of the data is segmented into a clearly insufficient. For this problem, this paper presents an objective to segment the three curve fitting method, which can eliminate the defects of the data from the subjective experience [17]. Finally, the test results show that the fitting effect is good.

Polynomial curve fitting is a more common method of data fitting. When the data point is too large, the polynomial order is too low, the fitting precision and the effect is not very ideal. In order to improve the fitting precision and effect, we need to increase the number of curves, but the complexity of the order is too high and the complexity of the calculation and other aspects of the disadvantages. Therefore, it is difficult to obtain good fitting accuracy and effect by using a polynomial curve fitting of the data points. In order to solve the above problems, the piecewise curve fitting is generally used, and local least squares fitting is performed on each interval [18]. The traditional piecewise curve fitting is based on the subjective experience and the drawing data to determine the empirical function and the segmentation point. A fitting method is proposed to segment the piecewise interval. The three data points are determined by each of the 4 data points, but the interval is too dense and not suitable for dense data fitting. Later, it is proposed that the global continuous fitting method of the polynomial basis function is restricted to 2 piecewise interval. In the next few years, the curve fitting method of the global continuous curve is put forward, but the basis function is restricted to a polynomial. The method proposed by the researcher is based on the subjective experience to determine the empirical function and the segmentation interval, and then the fitting [19, 20]. This method has the following points: 1 in the practical application of the method of automatic segmentation polynomial curve fitting. Provide several different empirical function, according to the different experience function was fitted to the data and the measured data, the error variance, to automatically determine the better experience function; 2. From the data fitting method many, such as logarithmic curve fitting, inverse function curve fitting, quadratic curve fitting, cubic curve fitting, exponential curve fitting, exponential curve fitting. First commonly observed scatter diagram to determine the curve type, but scattered point diagram is correlation between the coarse representations, and sometimes scattered point graph and several curves are very close, then establish corresponding empirical function may be reasonable, but due to the choice of different curves, a problem with a number of different empirical function, how to select the optimal from these experiences function in a. With several function fitting calculation of the historical data value and the fitting values of the error square and minimum for optimal function of the methods may exist such a problem: minimizing the sum of squared errors, but error fluctuation larger, i.e. some point error is very small, some error is relatively large. In view of this situation, this paper proposes a piecewise linear fitting method, with several different function fitting and the selection of optimal experience function and optimal function to determine the conditions are as follows: for historical data point error and error is negative number difference less than adaptation parameters [21].

The least squares method is a mathematical optimization technique It matches the optimal function of the data by minimizing the squared error of the squared error. Using the least square method, the unknown data can be easily obtained, and the square of the error between the data and the actual data is minimum. The least square method can be used in curve fitting. Other optimization problems can be expressed by the least square method for minimizing energy or maximum entropy. In the limited measurement series, the accuracy of the measurement is to seek the true value, which makes the square and the minimum of the error. The least square method is a basic method to measure the adjustment of different kinds of observation data [22]. If in different precision multiple observations of one or more unknown quantity, in order to determine the unknown quantities are the most reliable value, the concept of measurement must be with corrections to the modified the sum of weights positive multiplied by the square of the observation value is minimum. So called least squares method. The so-called "weight " is a measure of the relative reliability of the quality of observation results [23].

From the point of view, the least square method is similar to the interpolation method, which is the algorithm of processing data. But from the creation of the thought, the two are essentially different. The former sought a curve, and the observation data is "closest" to represent the observation data of the trend; the latter is the curve strictly by the given observational data, its purpose is through data from function model to approximately describe the function [24]. In the observational data with measurement error will make these observations deviate from the curve function, as a result and observational data to stay consistent interpolation method is better than the method of least squares curve more in line with the objective reality.

The basic idea of the least square method is: given a set of data $(x_i, y_i)(i = 1, 2, ..., n)$. The initial equation is $F(x)$ and the equation contains some undetermined coefficients a_n. Put (x_i, y_i) into the equation and calculate $y_i - F(x_i)$. Considering the error fluctuations, to prevent the emergence of error of plus or minus cancellation, will need equation square:

$$e = \sum (y_i - F(x_i))^2 \tag{1}$$

The minimum value of e can be obtained by a_n. In order to find the best fitting function of the group data, the function makes the error square and the minimum. Extended to a certain point in space. Standard equation for space line:

$$\frac{x - x_0}{k_1} = \frac{y - y_0}{k_2} = \frac{z - z_0}{k_3} \tag{2}$$

Can be simplified to get:

$$x = \frac{k_1}{k_3}(z - z_0) + x_0 = az + b \tag{3}$$

$$y = \frac{k_2}{k_3}(z - z_0) + y_0 = xz + d \tag{4}$$

Among them: $a = \dfrac{k_1}{k_3}$, $b = x_0 - \dfrac{k_1}{k_3}z_0$, $c = \dfrac{k_2}{k_3}$,

$d = y_0 - \dfrac{k_2}{k_3}z_0$.

For linear space, space straight line is equivalent to two plane equation of intersection, so on linear fitting can be respectively of the plane equation fitting. The fitting equation are obtained for the approximate value and the actual value of the difference between the square and recorded as:

$$\Delta x = \sum \left[x_i - (a z_i \quad b) \right]^2 \tag{5}$$

$$\Delta y = \sum \left[y_i - (a z_i \quad d) \right]^2 \tag{6}$$

According to the principle of least square method, the partial derivative of a, b, c, d in the formula is zero.

$$\begin{cases} \sum 2[x_i - (a z_i + b)] \times (-z_i) = 0 \\ \sum 2[x_i - (a z_i + b)] \times (-1) = 0 \\ \sum 2[y_i - (c z_i + d)] \times (-z_i) = 0 \\ \sum 2[y_i - (c z_i + d)] \times (-1) = 0 \end{cases} \tag{7}$$

On the formula (2.7), And the numerical value of a, b, c, d is obtained.

$$a = \frac{2 \times \sum x_i z_i - \sum x_i \times \sum z_i}{2 \times \sum z_i^2 - \sum z_i \times \sum z_i} \tag{8}$$

$$b = \frac{\sum x_i - a \times \sum z_i}{2} \tag{9}$$

$$c = \frac{2 \times \sum y_i z_i - \sum y_i \times \sum z_i}{2 \times \sum z_i^2 - \sum z_i \times \sum z_i} \tag{10}$$

$$d = \frac{\sum y_i - c \times \sum z_i}{2} \tag{11}$$

3. PRINCIPLE OF FITTING STRAIGHT LINE SEGMENT

The function relationship between the coordinates of the representation of the discrete point set in a continuous curve or a relatively flat surface. More broadly speaking, the corresponding problem in space or high dimensional space is also a category. In the numerical analysis, the curve fitting is used to approximate the discrete data, that is, the formula of the discrete data. In practice, group or discrete point data is often various physical problems and statistical problem was

related to the amount of times of observation value or experimental value, they are scattered, not only it is not easy to deal with and cannot generally be exact and fully reflect the inherent law. This defect is being made up by an appropriate analytical expression.

The TLS method has been widely used in the line and the plane fitting, the SVD method or iterative method to establish the coefficient matrix contains error model to solve the parameters, however, for the space of linear fitting of the TLS method does not see more. At present, there are commonly used spatial straight line fitting method: on the Newton gradient optimization algorithm for representing iteration algorithm, the algorithm is with two initial value, reach the accuracy requirements of the given by an iterative method, it involves a large number of equations are solved, solving process is relatively cumbersome, practicability is poor; non iterative algorithm, the method is essentially obtained with measuring point distance and the minimum plane, then three-dimensional problem into a two-dimensional problem, in plane using least squares method of fitting a straight line. The method takes into account the error of y_i and x_i directions, and the error of the z_i direction is neglected, and the error is not tight in theory [25].

The method of judging the fitting line is that the distance between the data coordinates and the corresponding fitting line should be less than the fixed value of ε. In the implementation of the algorithm, with minimal space discrete data points of straight line fitting, meet the conditions for a data point increases when the distance between all of the selected data points and fitting straight line less than ε. If the maximum distance between the data points and the maximum distance of them exceeds the fixed value of ε, the fitting process of the segment is stopped. When the data points are not added to the fixed value of ε, continue to perform an increase in the operation of the data points until the maximum number of points is added. When all of the above processes are performed, a spatial line is intended to be synthesized with the most data points [6-8].

For example, the space of any point $P(x_0, y_0, z_0)$ to the space of a two linear $A_1 x + B_1 y + C_1 z + D_1 = 0$ and $A_2 x + B_2 y + C_2 z + D_2 = 0$ distance is

$$d = \frac{\left| (A_1 x + B_1 y + C_1 z + D_1) \times \overrightarrow{n_2} - (A_2 x + B_2 y + C_2 z + D_2) \times \overrightarrow{n_1} \right|}{\left| \overrightarrow{n_1} \times \overrightarrow{n_2} \right|} \tag{12}$$

Among them: $\overrightarrow{n_1} = \{A_1, B_1, C_1\}, \overrightarrow{n_2} = \{A_2, B_2, C_2\}$. By the above data to fit the parameters of the line should be: $A_1 = -1, B_1 = 0, C_1 = a, D_1 = b, A_2 = 0, B_2 = -1, C_2 = c, D_2 = d$. The resulting system can be brought into the distance formula, the distance between the data points and the fitting line is obtained.

$$d_i = \frac{\sqrt{(-y + c z_i + d)^2 + (x_i - a z_i - b)^2 + (-c x_i + a y_i + b c - a d)^2}}{\sqrt{a^2 + c^2 + 1}} \tag{13}$$

The obtained d_i is the distance between the data points to the fitting line, and then the data can be fitted by a piecewise linear fitting by using this value.

4. COMPUTATIONAL PROCEDURE

Through the analysis of the above fitting principle, we can get the main steps: the first step is to take the experimental data two points ($A(x_1, y_1, z_1)$, $B(x_2, y_2, z_2)$). Mathematical model of fitting line l_1 by using least square method. Then a new point $C(x_3, y_3, z_3)$. The cosine of the angle between the line BC and the line l_1 is calculated:

$$\cos\theta = \frac{|a(x_3-x_2)+a(y_3-y_2)+(z_3-z_2)|}{\sqrt{a^2+c^2+1} \times \sqrt{(x_3-x_2)^2+(y_3-y_2)^2+(z_3-z_2)^2}} \quad (14)$$

If their cosine value is smaller than a given threshold, the point into the front, re calculated by least square method to fit the linear equation until greater than a given value. And we got the first fitting a linear equation. The second step is to repeat the last point of the straight line to the next point, until all points are fitted, so that all the equations of the piecewise linear equation are obtained. All processes as shown in Fig. (**1**).

5. SIMULATED DATA EXPERIMENT

Ten sets of data were obtained from statistics.

Using the above data to draw the space line Fig. (**2**).

CONCLUDING REMARKS

The above example shows that the improved least square method for piecewise linear fitting, in the processing of the experimental data, can significantly simplify the mathematical model establishment and reduce the computation, can better fitting of the sequence point, which proves that the method is feasible. On the other hand, the method can automatically determine the number of branching, automatic search for break point position, the overall optimal fitting, it is proved that the method is efficient. But this method also has certain limitations, the data of random changing fitting often fail to get the ideal results, but to change relatively slow data can get better effect, especially a monotone or monotone approximation of data. In the actual work, the data obtained from the test are often nonrandom, so the application of this method is quite wide. The classical least squares polynomial fitting as the foundation, through expansion of the original data, number of domain switching, data expansion, a series of data reconstruction, to ensure the piecewise linear fitting reliability and stability, to reconstruct the data with different number of polynomial repeatedly iterative fitting, after each iteration of the data points of large error correction and then continue the iterative, until the fitting stability. Finally, through iteration error, number of iterations modified the size of integrated and adaptive to get the best fitting results. In this paper, different theoretical models are used to verify
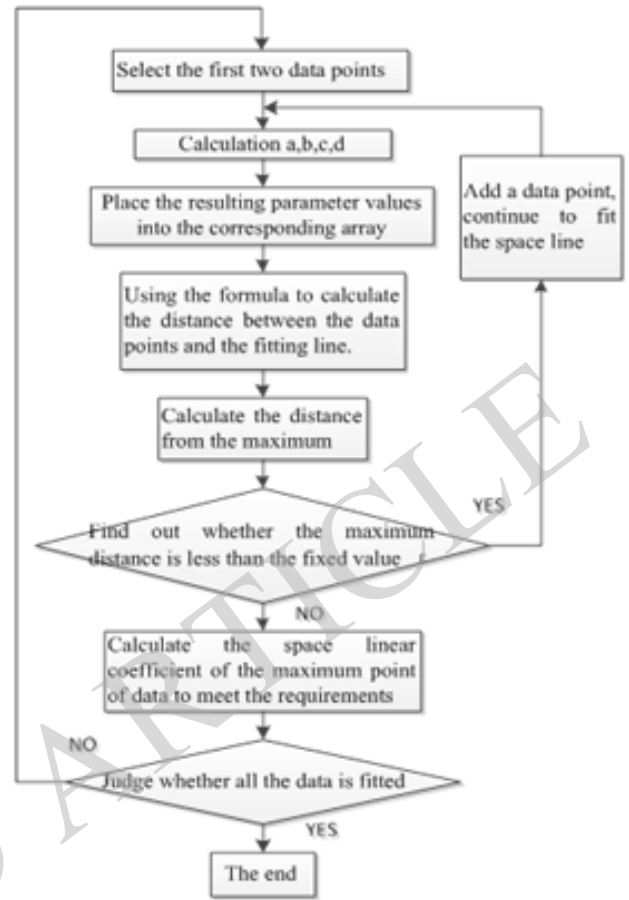

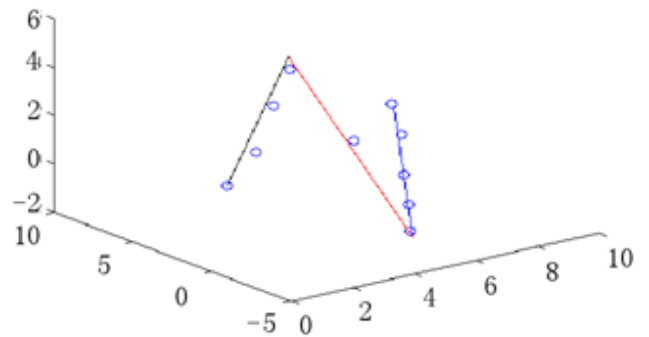**Fig. (1).** Fitting subsection linear program block diagram.

Fig. (2). Spatial piecewise linear fitting.

the optimization algorithm, and the results of different models are analyzed, and the algorithm is correct and feasible. The treatment effect of the actual experimental data to verify the analysis, and draws the conclusion that the optimization algorithm can be used for any type of random error of the test data fitted better than today's method of conclusions. The disadvantage is that the iteration time is longer, and the evaluation of the optimal fitting results needs to find a better judgment method.

According to the principle of least square method is proposed for spatial discrete points by the method of least squares piecewise linear fitting, calculate the space discrete data fitting for all coefficients of linear space algorithm, and

Table 1.   Data sheets for any 10 groups.

x	y	z	x	y	z
1	1.0	1.0	6	-0.8	-1.8
2	1.4	2.1	7	1.2	1.4
3	2.0	3.5	8	3.5	-1.1
4	2.9	4.5	9	5.6	-0.2
5	0.8	-1.8	10	8.2	0.1

Table 2.   Equation coefficients for running programs.

k	a	b	C	d
1	0.8332	0.1836	0.5492	0.2383
2	-0.3139	5.4719	0.5785	0.1003
3	1.9039	9.7173	4.2626	7.3789

puts forward the piecewise linear fitting method to distinguish, the basic principle and then come to the least squares method space segments of straight line fitting. The validity of this method is verified by the results of the program.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

REFERENCES

[1]   A. Chkifa, A. Cohen, G. Migliorati, F. Nobile and R. Tempone, "Discrete least squares polynomial approximation with random evaluations− application to parametric and stochastic elliptic PDEs", *ESAIM: Mathematical Modelling and Numerical Analysis*, vol. 49, no. 3, pp. 815-837, 2015.

[2]   H.C. Goicoechea, A.C. Olivieri and R. Tauler, "Application of the correlation constrained multivariate curve resolution alternating least-squares method for analyte quantitation in the presence of unexpected interferences using first-order instrumental data", *Analyst*, vol. 135, no. 3, pp. 636-642, 2010.

[3]   O. Götz, K. Liehr-Gobbers and M. Krafft, *Evaluation of structural equation models using the partial least squares (PLS) approach, Handbook of partial least squares,* Springer, Berlin Heidelberg, 2010, pp. 691-711.

[4]   W.R. Schwartz, A. Kembhavi, D. Harwood and L.S. Davis, "Human detection using partial least squares analysis", *IEEE 12ᵗʰ International Conference on Computer Vision,* pp. 24-31, 2009.

[5]   H. Gavin, "*The Levenberg-Marquardt Method for Nonlinear Least Squares Curve-Fitting Problems*", Department of Civil and Environmental Engineering, Duke University, pp. 1-15, 2011.

[6]   X. Tong, Y. Jin, and L. Li, "An improved weighted total least squares method with applications in linear fitting and coordinate transformation", *Journal of Surveying Engineering*, vol. 137, no. 4, pp. 120-128, 2011.

[7]   F. Ding, P.X. Liu and G. Liu, "Multiinnovation least-squares identification for system modelling", *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, no. 3, pp. 767-778, 2010.

[8]   H. Han, L. Xie, F. Ding and X. Liu, "Hierarchical least-squares based iterative identification for multivariable systems with moving average noises", *Mathematical and Computer Modelling*, vol. 51, no. 9, pp. 1213-1220, 2010.

[9]   J. Hesch and S. Roumeliotis, A direct least-squares (DLS) method for PnP, In: *IEEE International Conference on Computer Vision (ICCV),* pp. 383-390, 2011.

[10]   T. Quan, X. Liu, and Q. Liu, "Weighted least squares support vector machine local region method for nonlinear time series prediction", *Applied Soft Computing*, vol. 10, no. 2, pp. 562-566, 2010.

[11]   S. Van Huffel, *Total Least Squares and Errors-in-Variables Modeling: Analysis, Algorithms and Applications*, Springer Science & Business Media, Berlin, 2013.

[12]   J.B. Lohmöller, "*Latent Variable Path Modeling with Partial Least Squares*", Springer Science & Business Media, Berlin, 2013.

[13]   D. York, "Least-squares fitting of a straight line", *Canadian Journal of Physics,* vol. 44, no. 5, pp. 1079-1086, 1966.

[14]   K.S. Arun, T.S. Huang and S.D. Blostein, "Least-squares fitting of two 3-D point sets", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 5, pp. 698-700, 1987.

[15]   A. Fitzgibbon, M. Pilu and R.B. Fisher, "Direct least square fitting of ellipses", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 476-480, 1999.

[16]   S.J. Kim, K. Koh, M. Lustig, S. Boyd and D. Gorinevsky "An interior-point method for large-scale l 1-regularized least squares", *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606-617, 2007.

[17]   K.M. Brownf and J.E. Dennis, "A new algorithm for nonlinear least-squares curve fitting", *Mathematical Software,* vol. 5, no. 23, p. 391, 2014.

[18]   J. Shults, "Quasi‐least squares fitting", *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 7, no. 3, pp. 194-204, 2015.

[19]   X. Zhuang, H. Zhu and C. Augarde, "An improved meshless Shepard and least squares method possessing the delta property and requiring no singular weight function", *Computational Mechanics,* vol. 53, no. 2, pp. 343-357, 2014.

[20]   T. Kuosmanen and A.L. Johnson, "Data envelopment analysis as nonparametric least-squares regression", *Operations Research*, vol. 58, no. 1, pp. 149-160, 2010.

[21]   Y. Liu, D. Wang and F. Ding, "Least squares based iterative algorithms for identifying Box–Jenkins models with finite measurement data", *Digital Signal Processing*, vol. 20, no. 5, pp. 1458-1467, 2010.

[22]   I. Daubechies, R. DeVore, M. Fornasier and C.S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery", *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1-38, 2010.

[23]   Z.-M. Zhang, S. Chen and Y.-Z. Liang, "Baseline correction using adaptive iteratively reweighted penalized least squares", *Analyst*, vol. 135, no. 5, pp. 1138-1146, 2010.

[24]   K.C. Toh and S. Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems", *Pacific Journal of Optimization*, vol. 6, no. 15, pp. 615-640, 2010.

[25]   R. Chudamani, K. Vasudevan and C.S. Ramalingam, "Real-time estimation of power system frequency using nonlinear least squares", *IEEE Transactions on Power Delivery*, vol. 24, no. 3, pp. 1021-1028, 2009.