IJAMML 6:1 (2017) 35-44 March 2017
ISSN: 2394-2258
Available at http://scientificadvances.co.in
DOI: http://dx.doi.org/10.18642/ijamml_7100121818

# TOTAL LEAST SQUARES ALGORITHMS FOR FITTING 3D STRAIGHT LINES

**Cuiping Guo[a], Junhuan Peng[b] and Chuantao Li[a]**

[a]School of Science, China University of Geosciences (Beijing), Beijing, 100083 P. R. China
[b]School of Land Science and Technology, China University of Geosciences (Beijing), Beijing, 100083, P. R. China

## Abstract

To address the problem of fitting a 3D straight line, the TLS method based on the Lagrange function is used to solve it. The number of parameter to be estimated is decreased from six to four by changing the standard equation of the straight line into the projective equation of it. The problem of fitting a 3D straight line is converted to the problem of fitting two 2D straight lines with errors in both coordinates. And then the total least square (TLS) and least square (LS) method are employed to fit the two 2D straight line. A simulated example is carried out to demonstrate the effectiveness and applicability of proposed algorithms.

*Keywords*: 3D straight line fitting, 2D straight line fitting, least square, total least squares.

*Corresponding author.
*E-mail address*: gcp@cugb.edu.cn (Cuiping Guo).


2010 Mathematics Subject Classification: 62F10.
Submitted by Jianqiang Gao.
Received February 10, 2017

## 1. Introduction

3D line fitting is a common problem in many metrological and measurement systems. Nowadays, as most of the instruments provide 3D coordinates, engineers and scientists have to work within a 3D coordinates frame. For all the coordinates contain errors, the 3D line fitting problem can be discussed in total least square (TLS) framework. Since Pearson [9] solved the problem of fitting 2D lines to data with errors in both coordinates, quite a number of total least squares (TLS) methods were developed to deal with the 2D line fitting. Now, there are many researches about the total least squares in algorithms such as the singular value decomposition (SVD) algorithm (Golub and Van Loan [2]) and the algorithm based on the Lagrange function (Schaffrin and Wieser [12]). For more information about the methodology of TLS, one can refer to Huffel et al. [3, 4]. York [15] gave a detailed discussion of the calculation of the "best straight line" by the method of least squares (LS). Reed [10] reiterated York's solution and indicated an easier way to solve for the slope of the best-fit line. Neri et al. [8] solved the line regression problem with a straightforward analytical approach that uses the minimization of the shortest distance between each experimental point and the theoretical line. Wong [14] discussed and compared the likelihood-based methods for obtaining approximate confidence intervals for the slope in a simple linear regression when both variables were measured with errors. Schaffrin et al. [11] obtained TLS solution by solving non-linear normal equations via a newly developed iterative approximation algorithms. Fitting a straight line to data with uncertainties in both coordinates was discussed by reducing to a one-dimensional search for a minimum in Krystek and Anton [6] and generalized to the case when there are correlations in Krystek and Anton [7]. Amiri-Simkooei et al. [1] presented a simple and reliable formulation for the linear regression fit using the weighted total least squares (WTLS) problem, when both variables are subjected to different and possibly correlated noise. However, these methods cannot be readily extended to solve the problem of fitting spatial line to points with noisy coordinates in three dimensions.

Up to now, there are little literatures focused on 3D line fitting. Kahn [5] presents a simple non-iterative linear procedures for finding the least squares line in three dimensions by minimizing the sum of the squared perpendicular distances between the data points and the fitted line. Snow and Schaffrin [13] solved the problem of fitting lines in 3D space using a new algorithm for the TLS solution under a nonlinear Gauss-Helmert model. In Snow's paper, only four parameters were estimated, thereby avoiding over-parameterization.

In this paper, by projecting the 3D straight line onto two coordinate plane, only four parameters are to be estimated. 3D line fitting is investigated with the objective of minimizing the total sum of all squared random errors in the 3D variables. The structure of this paper is as follows. Section 2 introduces 3D line fitting with errors in $y$ coordinate and $z$ coordinate. Section 3 shows 3D line fitting when all the three coordinates contain errors. A simulation study is included in Section 4, and it is concluded in Section 5.

## 2. 3D Line Fitting under LS Criterion

Suppose that a 3D straight line B goes through the point $(x_0, y_0, z_0)$ and has a direction vector $(p_x, p_y, p_z)$, and standard equation of line B can be expressed as

$$\frac{x - x_0}{p_x} = \frac{y - y_0}{p_y} = \frac{z - z_0}{p_z}.$$

The projective equations of the straight line is derived as

$$\begin{cases} x = az + b, \\ y = cz + d, \end{cases} \tag{1}$$

where

$$a = \frac{p_x}{p_z}, \quad b = x_0 - \frac{p_x}{p_z} z_0, \quad c = \frac{p_y}{p_z}, \quad d = y_0 - \frac{p_y}{p_z} z_0. \tag{2}$$

The straight line can be regarded as the intersection of the two planes represented by the two equations, and only four parameters need to be estimated. We can fit the $x$ and $y$ coordinates of the data to Equation (1) to get the estimated value of $a$ and $b$, and fit the $y$ and $z$ coordinates of the data to Equation (2) to get the estimated value of $c$ and $d$. We assume the variables $x$ and $y$ have equal variances $\sigma^2$. Using $n$ observations $(x_i, y_i)(i = 1, \cdots, n)$ and the accompanying fixed values $z_i\,(i = 1, \cdots, n)$, we can get the LS estimate of $a$, $b$, $c$, and $d$.

$$\hat{\beta}_1 = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \left(A^T A\right)^{-1} A^T X, \quad \hat{\beta}_2 = \begin{bmatrix} \hat{c} \\ \hat{d} \end{bmatrix} = \left(A^T A\right)^{-1} A^T Y, \tag{3}$$

where

$$A = \begin{bmatrix} z_1 & 1 \\ \vdots & \vdots \\ z_n & 1 \end{bmatrix}, \quad X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

For the normal vectors of the plane expressed by (1) and (2), we can formulate, respectively, as

$$\vec{n}_1 = (1, 0, -\hat{a}), \quad \vec{n}_2 = (0, 1, -\hat{c}),$$

and get a direction vector of the line B

$$\vec{l} = \vec{n}_1 \times \vec{n}_2 = (\hat{a}, \hat{c}, 1).$$

Let

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \quad \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}, \quad \bar{z}\ \frac{\sum_{i=1}^{n} z_i}{n}.$$

Because under the LS criterion the straight line fitted by the data goes through the center of the data, the place expressed by (1) goes through $(\bar{x}, y, \bar{z})(\forall y)$, and the place expressed by (2) goes through $(x, \bar{y}, \bar{z})(\forall x)$.

Therefore, the estimated line B goes through the center $(\bar{x}, \bar{y}, \bar{z})$ of the data. The estimated line B can be expressed as

$$\frac{x - \bar{x}}{\hat{a}} = \frac{y - \bar{y}}{\hat{c}} = \frac{z - \bar{z}}{1}. \tag{4}$$

## 3. 3D Line Fitting under TLS Criterion

Assuming the $z$ variable is contaminated by gross errors besides $x$ variable and the $y$ variable, and have variance $\sigma^2$. Using $n$ observations $(x_i, y_i, z_i)(i = 1, \cdots, n)$, we can form the following equations according to (1) and (2):

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} z_1 - e_{z_1} & 1 \\ \vdots & \vdots \\ z_n - e_{z_n} & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} e_{x_1} \\ \vdots \\ e_{x_n} \end{bmatrix}, \tag{5}$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} z_1 - e_{z_1} & 1 \\ \vdots & \vdots \\ z_n - e_{z_n} & 1 \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} + \begin{bmatrix} e_{y_1} \\ \vdots \\ e_{y_n} \end{bmatrix}. \tag{6}$$

Letting

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, Z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}, e_x = \begin{bmatrix} e_{x_1} \\ \vdots \\ e_{x_n} \end{bmatrix}, e_y = \begin{bmatrix} e_{y_1} \\ \vdots \\ e_{y_n} \end{bmatrix}, 1_n = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix},$$

(5) and (6) can be expressed in matrix notion

$$X = a(Z - e_z) + b \cdot 1_n + e_x, \tag{7}$$

$$Y = c(Z - e_z) + d \cdot 1_n + e_y. \tag{8}$$

Therefore, the problem of 3D line fitting is transformed into two 2D line fitting problems with errors in both coordinates. (7) and (8) have the same form, and the same coefficient matrix, so the solution of (7) and (8) have similar representation. For this reason, only the solution of (7) is discussed below.

The total least squares principle is to minimize the objective function

$$S = e_x{}^T Q_x^{-1} e_x + e_z{}^T Q_z^{-1} e_z.$$

By employing the equivalent target function in accordance with Lagrange method, we have

$$\Phi(e_x, e_z, \lambda, a, b) = e_x{}^T Q_x^{-1} e_x + e_z{}^T Q_z^{-1} e_z + 2\lambda^T (X - a(Z - e_z) - b \cdot 1_n - e_x).$$

And the necessary Euler-Lagrange conditions are derived, namely,

$$\frac{1}{2} \frac{\partial \Phi}{\partial e_x}\bigg|_{\tilde{e}_x, \tilde{e}_z, \hat{\lambda}, \hat{a}, \hat{b}} = Q_x^{-1} \tilde{e}_x - \hat{\lambda} = 0, \tag{9}$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial e_z}\bigg|_{\tilde{e}_x, \tilde{e}_z, \hat{\lambda}, \hat{a}, \hat{b}} = Q_z^{-1} \tilde{e}_z + a \cdot \hat{\lambda} = 0, \tag{10}$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial \lambda}\bigg|_{\tilde{e}_x, \tilde{e}_z, \hat{\lambda}, \hat{a}, \hat{b}} = X - aZ - b \cdot 1_n + e_z \cdot a - e_x = 0, \tag{11}$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial a}\bigg|_{\tilde{e}_x, \tilde{e}_z, \hat{\lambda}, \hat{a}, \hat{b}} = -(Z - e_z)^T \hat{\lambda} = 0, \tag{12}$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial b}\bigg|_{\tilde{e}_x, \tilde{e}_z, \hat{\lambda}, \hat{a}, \hat{b}} = -1_n^T \cdot \hat{\lambda} = 0, \tag{13}$$

where tildas indicate “predicted” vectors, and hats indicate “estimated” ones.

Now, $\tilde{e}_x$ and $\tilde{e}_z$ can be expressed in terms of $\hat{\lambda}$ by using (9) and (10). This leads to that

$$\tilde{e}_x = Q_x \hat{\lambda}, \tag{14}$$

$$\tilde{e}_z = -aQ_z \hat{\lambda}, \tag{15}$$

and after inserting this into (11), we have that

$$\hat{\lambda} = \left[Q_x + a^2 Q_z\right]^{-1}(X - aZ - b \cdot 1_n). \tag{16}$$

Let

$$Q_1 = Q_x + a^2 Q_z, \tag{17}$$

then $Q_1$ is invertible. We readily obtain

$$\hat{\lambda} = Q_1^{-1}(X - aZ - b \cdot 1_n). \tag{18}$$

Inserting (18) into (13), we get

$$\hat{b} = \left(1_n^T Q_1^{-1} 1_n\right)^{-1} 1_n^T Q_1^{-1}(X - aZ). \tag{19}$$

Let

$$Q_2 = \left(1_n^T Q_1^{-1} 1_n\right)^{-1} 1_n^T Q_1^{-1},\ Q_3 = Q_1^{-1}(I_n - 1_n Q_2),\ Q_4 = (Z^T Q_3 Z)^{-1},\ Q_5 = Z Q_4 Z^T,$$

$$\hat{b} = Q_2(X - aZ). \tag{20}$$

Using (20), we can get

$$\hat{a} = Q_4\left(Z^T Q_3 X - \hat{\lambda}^T \tilde{e}_z\right). \tag{21}$$

Inserting (21) into (19), the closed-form expression of the estimated parameter vector $a$ and $b$ can be derived as

$$\begin{cases} \hat{a} = Q_4\left(Z^T Q_3 X - \hat{\lambda}^T \tilde{e}_z\right), \\ \hat{b} = Q_2\left((I_n - Q_5 Q_3)X + (\hat{\lambda}^T \tilde{e}_z) \cdot Z\right). \end{cases} \tag{22}$$

## 4. Case Study

In this section, the proposed TLS approach will be applied to a simulated example, compared with the LS approach. A 3D straight line is given as

$$\frac{x-1}{4} = \frac{y+2}{-2} = \frac{z-3}{1}. \tag{23}$$

The projective equation of the straight line is given as

$$\begin{cases} x = 4z - 11, \\ y = 2z + 4. \end{cases} \tag{24}$$

Taking $z$ as 21 numbers distributed uniformly in [–10, 10], and calculating the values of $y$ and $z$ according to Equation (24), 21 pairs of points are formed. The normal random error is generated and added to the coordinates of each point. The design is as follows: the LS method, the proposed TLS method are implemented for comparison purposes. The LS solution are given as the initial values for iterations of the TLS method, and $\varepsilon = 10^{-10}$ is chosen as the convergence tolerance. The means and the root mean square errors (RMSE) of the parameters $\beta_1$ and $\beta_2$, and the maximum deviation between the calculated value and the true value are computed for 1000 experiments for the two methods. The results are shown in Table 1.

**Table 1.** Comparisons of the LS method and the proposed TLS method

| | | LS | TLS |
|---|---|---|---|
| True value(a) 4 | Mean(a) | 3.91414061 | 3.94550270 |
| | RMSE(a) | 0.22405128 | 0.08445361 |
| True value(b) – 11 | Mean(b) | – 11.16109969 | – 10.96450923 |
| | RMSE(b) | 0.10040905 | 0.06405110 |
| True value(c) – 2 | Mean(c) | – 1.95131821 | – 1.96537423 |
| | RMSE(c) | 0.74550929 | 0.05493878 |
| True value(d) 4 | Mean(d) | 4.70268519 | 3.95538625 |
| | RMSE(d) | 0.08971675 | 0.04103927 |

As we can be seen from Table 1, for all the parameters to be estimated, the parameter estimates obtained by TLS are closer to the true values than LS, and have smaller RMSE.

## 5. Conclusion

By changing the standard equation of the 3D straight line into the projective equation of it, the number of parameter to be estimated is decreased from six to four. Using a minimum parameterization, we have solved the 3D straight line fitting problem by converting it to the problem of fitting two 2D straight lines with errors in both coordinates. And then the TLS method is adopted to fit the 2D straight line by minimizing the sum of the squared error. Moreover, a simulated example is carried out to demonstrate the effectiveness and applicability of the proposed algorithm.

## Acknowledgement

The research is supported by the Fundamental Research Funds for Central Universities (No.2652015193), Undergraduate Teaching-Reform Project (No.JGYB201529), Graduate Teaching-Reform Project of China of University of Geosciences (Beijing), NSFC (41374016, 41104025 and 41330634), and SKLGED2013-4-8-E.

## References

[1] A. R. Amiri-Simkooei, F. Zangeneh-Nejad, J. Asgari and S. Jazaeri, Estimation of straight line parameters with fully correlated coordinates, Measurement 48(1) (2014), 378-386.

[2] G. H. Golub and C. F. Van Loan, An analysis of the total least squares problem, SIAM Journal on Numerical Analysis 17(6) (1980), 883-893.

[3] S. V. Huffel and J. Vandewalle, The Total Least Squares Problem: Computational Aspects and Analysis, Society for Industrial and Applied Mathematics, Philadelphia, 1997.

[4] S. V. Huffel, C. L. Cheng, N. Mastronardi, C. Paige and A. Kukush, Total Least Squares and Errors-in-Variables Modeling: Analysis, Algorithms and Applications, Springer Science and Business Media, Dordrecht, 2013.

[5] P. C. Kahn, Simple methods for computing the least squares line in three dimensions, Computers and Chemistry 13(13) (1989), 191-195.

[6] M. Krystek and M. Anton, A weighted total least-squares algorithm for fitting a straight line, Measurement Science and Technology 18(11) (2007), 3438.

[7] M. Krystek and M. Anton, A least-squares algorithm for fitting data points with mutually correlated coordinates to a straight line, Measurement Science & Technology 22(3) (2011), 035101.

[8] F. Neri, G. Saitta and S. Chiofalo, An accurate and straightforward approach to line regression analysis of error-affected experimental data, Journal of Physics E Scientific Instruments 22(22) (1989), 215-217.

[9] K. Pearson, On lines and planes of closest fit to systems of point in space, Philosophical Magazine 2(11) (1901), 559-572.

[10] B. C. Reed, Linear least-squares fits with errors in both coordinates, American Journal of Physics 57(57) (1989), 642-646.

[11] B. Schaffrin, I. Lee, Y. Felus and Y. Choi, Total least-squares (TLS) for geodetic straight-line and plane adjustment, Bollettino Di Geodesia E Scienze Affini 65(3) (2006), 141-168.

[12] B. Schaffrin and A. Wieser, On weighted total least-squares adjustment for linear regression, Journal of Geodesy 82(7) (2008), 415-421.

[13] K. Snow and B. Schaffrin, Line fitting in Euclidean 3D space, Studia Geophysica et Geodaetica 60(2) (2016), 210-227.

[14] M. Y. Wong, Likelihood estimation of a simple linear regression model when both variables have error, Biometrika 76(1) (1989), 141-148.

[15] D. York, Least-squares fitting of a straight line, Canadian Journal of Physics 44(5) (1966), 1079-1086.

■