Weighted total least squares formulated by standard least squares theory

Research article

A. Amiri-Simkooei^{1,2*}, S. Jazaeri³

1 Department of Surveying Engineering, Faculty of Engineering, University of Isfahan, 81746-73441 Isfahan, Iran

2 Acoustic Remote Sensing Group, Faculty of Aerospace Engineering, Delft University of Technology, Delft, the Netherlands

3 Department of Surveying and Geomatics Engineering, College of Engineering, University of Tehran, Tehran, Iran

Abstract:

This contribution presents a simple, attractive, and flexible formulation for the weighted total least squares (WTLS) problem. It is simple because it is based on the well-known standard least squares theory; it is attractive because it allows one to directly use the existing body of knowledge of the least squares theory; and it is flexible because it can be used to a broad field of applications in the error-in-variable (EIV) models. Two empirical examples using real and simulated data are presented. The first example, a linear regression model, takes the covariance matrix of the coefficient matrix as $Q_A = Q_n \otimes Q_m$, while the second example, a 2-D affine transformation, takes a general structure of the covariance matrix Q_A . The estimates for the unknown parameters along with their standard deviations of the estimates are obtained for the two examples. The results are shown to be identical to those obtained based on the *nonlinear* Gauss-Helmert model (GHM). We aim to have an impartial evaluation of WTLS and GHM. We further explore the high potential capability of the presented formulation. One can simply obtain the covariance matrix of the WTLS estimates. In addition, one can generalize the orthogonal projectors of the standard least squares from which estimates for the residuals and observations (along with their covariance matrix), and the variance of the unit weight can directly be derived. Also, the constrained WTLS, variance component estimation for an EIV model, and the theory of reliability and data snooping can easily be established, which are in progress for future publications.

Keywords:

standard least squares • errors-in-variables model • weighted total least squares • singular value decomposition

© Versita sp. z o.o.

Received 2012-03-16; accepted 2012-04-30

1. Introduction

A significant part of literature on the estimation theory distinguishes between the standard least squares (SLS) and the total least squares (TLS). The latter originates from the work of Golub and van Loan (1980) in mathematical literature in which they introduced the error-in-variable (EIV) models. An EIV model differs from the standard Gauss-Markov model (GMM) in the sense that the coefficient matrix connecting the parameters to the random observation vector is also affected by random errors. The simplest TLS model includes the case when both coordinate components of

a linear regression model are observed.

In geodetic literature, Burkhard Schaffrin has a significant contribution in developing different theories on the TLS technique. We may refer to Schaffrin and Wieser (2008) in which they developed weighted TLS for an EIV model using the traditional Lagrange function. Their algorithm is restricted to the class of $Q_A = Q_n \otimes Q_m$ for the covariance matrix of the coefficient matrix. Schaffrin and his colleagues developed new algorithms to solve the TLS adjustment on the model of condition equations and the TLS problem with linear and quadratic constraints. The reader is referred to Schaffrin and Felus (2009); Schaffrin and Wieser (2009) and Schaffrin and Wieser (2011). Shen et al. (2011) proposed an algorithm that resembles the standard least-squares formulation. Mahboub (2012) developed an algorithm to use a general covariance matrix and

*E-mail: A.AmiriSimkooei@tudelft.nl

provided some guidelines to make such a matrix. Xu et al. (2012) formulated WTLS as a nonlinear adjustment model without constraints and further extended it to a partial EIV model.

It is widely known that the EIV-model can also be considered as a nonlinear Gauss–Helmert model (GHM). It is because in a TLS adjustment with an EIV model the same objective function as in the adjustment by a nonlinear GHM is minimized. TLS adjustment may not be regarded as a new adjustment method, but rather as an additional possibility to formulate a new algorithm in the frame of the general method of least-squares (Neitzel, 2010). We aim to have an impartial evaluation of both methods. The solution of the nonlinear GHM is somewhat critical to handle due to the many pitfalls described by Pope (1972). In doing so, special attention should be paid to appropriate linearization and iteration of the model. In contrast, the elegance of the TLS algorithm lies on its simplicity in the sense that it is formulated in the standard GMM and avoids immediate linearization. Therefore, instead of solving a nonlinear GHM in an iterative manner, we use the WTLS adjustment within the EIV-model and solve it efficiently using a linearly-structured iterative algorithm. This efficiency might also include the number of iterations involved, as demonstrated in the present contribution. We should however note that the GHM can be used as a reference to cross-check the WTLS results. Particular attention is paid to the standard deviations of the estimates obtained using WTLS and GHM.

The objective of this contribution is three-fold. First, we aim to formulate the WTLS problem using the standard least squares theory. An alternative expression gives a different appearance of Tong et al. (2011) and Mahboub (2012) algorithm (but it is practically identical), while an alternative derivation gives a similar expression to that of Shen et al. (2011). This latter derivation is of interest because it is straightforward and does not treat the problem to be nonlinear—a prerequisite for Shen et al. (2011) derivation. The algorithm takes into consideration the complete structure for the covariance matrix of the coefficient matrix. Second, we further explore the high potential capability of the formulation. Having this formulation available, one can apply the existing body of the knowledge of the least squares to the WTLS problem. We can at least highlight that (1) one can simply obtain the covariance matrix of the WTLS estimated parameters, (2) one can simply generalize the orthogonal projectors of the standard least squares from which estimates for the residuals, observations, and variance of the unit weight can directly be derived, (3) one can define measures for the reliability and hence the data snooping procedures can simply be implemented, and (4) one can easily establish the variance component estimation for an EIV model. Third, because the WTLS problem can, in principle, be formulated equivalently in terms of a nonlinear GHM, to assure of the proper formulation of the estimated parameters along with their covariance matrices, we use the Gauss–Helmert results as a reference. It is shown that the results of these two methods are identical.

This paper is organized as follows. In Sect. 2, a general solution

is given to the WTLS problem which is formulated in the standard least squares framework. Section 3 shows the applicability of the algorithm to two numerical examples: (i) a linear regression model of which both x and y coordinates have been observed, and (ii) a planar linear affine transformation in which the coordinates are observed in both the start and target systems. For both examples, the results—estimates along with their standard deviations—are shown to be identical to those obtained using GHM. We draw some conclusions in Sect. 4.

2. Weighted total least squares (WTLS)

Before we continue with the weighted total least squares (WTLS), we briefly explain the theory of the standard least squares. Consider the linear model of observation equations, called the Gauss–Markov model, for a given geodetic observations set

$$y = Ax + e \quad (1)$$

where y is the m -vector of observations, A is the $m \times n$ design matrix, and x is the n -vector of unknown parameters. The least-squares estimate of the unknown parameters is

$$\hat{x} = (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} y \quad (2)$$

where Q_y is the $m \times m$ covariance matrix of the observations. The estimated vectors of the observations and of the residuals follow respectively from

$$\begin{cases} \hat{y} = P_A y \\ \hat{e} = P_A^\perp y \end{cases} \quad (3)$$

where $P_A = A(A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1}$ and $P_A^\perp = I - A(A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1}$ are two $m \times m$ orthogonal projectors. The covariance matrices of the least-squares estimates $\hat{x}$, $\hat{y}$, and $\hat{e}$ are $Q_{\hat{x}} = (A^T Q_y^{-1} A)^{-1}$, $Q_{\hat{y}} = P_A Q_y$, and $Q_{\hat{e}} = P_A^\perp Q_y$, respectively (Teunissen, 2000).

In many geodetic surveying applications, one usually assumes that only the observations y are corrupted by random noise of which the preceding formulation can be used. There are, however, cases that the model itself is also corrupted by random noise. In this case, the Gauss–Markov model is replaced by an ‘errors-in-variables’ (EIV) model expressed as

$$y = (A - E_A)x + e_y \quad (4)$$

with its stochastic properties characterized by

$$\begin{bmatrix} e_y \\ e_A \end{bmatrix} := \begin{bmatrix} e_y \\ \text{vec}(E_A) \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma_0^2 \begin{bmatrix} Q_y & 0 \\ 0 & Q_A \end{bmatrix} \right) \quad (5)$$

where e_y is the m -vector of observational errors, A is the $m \times n$ design matrix, E_A is the corresponding $m \times n$ matrix of random errors, x is the n -vector of unknown parameters, $D(e_y) = \sigma_0^2 Q_y$ and $D(e_A) = \sigma_0^2 Q_A$ are the corresponding symmetric and non-negative dispersion matrices of size $m \times m$ and $mn \times mn$, respectively. In both expressions, σ_0^2 is the unknown variance factor of the unit weight (for now we assume $\sigma_0^2 = 1$).

In the homoscedastic case, one assumes that $Q_y = I_m$ and $Q_A = I_{mn}$, i.e. identities matrices of sizes m and mn , respectively. This introduces the standard TLS which was originally introduced in the mathematical literature by Golub and van Loan (1980). They solved this TLS problem using the solution of the following eigenvalue problem:

$$\begin{bmatrix} A^T A & A^T y \\ y^T A & y^T y \end{bmatrix} \begin{bmatrix} \hat{x} \\ -1 \end{bmatrix} = \begin{bmatrix} \hat{x} \\ -1 \end{bmatrix} \hat{\nu}_{\min}, \quad \hat{\nu}_{\min} \geq 0 \quad (6)$$

where $\hat{\nu}_{\min}$, the smallest eigenvalue, along with $\hat{x}$ are obtained in an iterative manner. An introduction into the TLS methods can be found in van Huffel and Vandewalle (1991), where the solution is based on a SVD approach. Alternative iteration schemes and a comprehensive literature review are given by Schaffrin et al. (2006) as well. Schaffrin and Wieser (2008) argued that the so-called generalized TLS, introduced by van Huffel and Vandewalle (1991), does not really solve the weighted TLS problem because the “weights” introduced in TLS (GTLS) do not refer to the covariance matrix of the observations. They introduced the “weighted TLS” that follows the geodetic tradition based on the inverse of the covariance matrix.

Similar to Schaffrin and Wieser (2008), we use the target (Lagrange) function as

$$\Phi := e_y^T Q_y^{-1} e_y + e_A^T Q_A^{-1} e_A + 2\lambda^T \times (y - Ax - e_y + (x^T \otimes I_m) e_A) \quad (7)$$

with λ a m -vector of unknown Lagrange multipliers and I_m an identity matrix of size m . The first partial derivatives of Eq. (7) with respect to the vectors e_y , e_A , λ , and x follow

$$\frac{1}{2} \frac{\partial \Phi}{\partial e_y^T} = Q_y^{-1} \tilde{e}_y - \hat{\lambda} = 0 \quad (8)$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial e_A^T} = Q_A^{-1} \tilde{e}_A + (\hat{x} \otimes I_m) \hat{\lambda} = 0 \quad (9)$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial \lambda^T} = y - A\hat{x} - \tilde{e}_y + (\hat{x}^T \otimes I_m) \tilde{e}_A = 0 \quad (10)$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial x^T} = -(A^T \hat{\lambda} - \tilde{E}_A^T \hat{\lambda}) = 0 \quad (11)$$

where $(\sim)$ and $(\hat{\cdot})$ represent the “predicted” and “estimated” quantities, respectively. The predicted residual vectors follow from Eq. (8) and Eq. (9) as

$$\tilde{e}_y = Q_y \hat{\lambda} \quad (12)$$

and

$$\tilde{e}_A = \text{vec } \tilde{E}_A = -Q_A (\hat{x} \otimes I_m) \hat{\lambda} \quad (13)$$

Having substituted these equations into Eq. (10) yields

$$\hat{\lambda} = (Q_y + (\hat{x}^T \otimes I_m) Q_A (\hat{x} \otimes I_m))^{-1} (y - A\hat{x}) \quad (14)$$

or

$$\hat{\lambda} = Q_{\tilde{y}}^{-1} (y - A\hat{x}) \quad (15)$$

where

$$Q_{\tilde{y}} = Q_y + (\hat{x}^T \otimes I_m) Q_A (\hat{x} \otimes I_m) \quad (16)$$

is the covariance matrix of the predicted observations $\tilde{y} = y - \tilde{E}_A x$. Substitution of Eq. (15) into Eq. (11) yields

$$A^T Q_{\tilde{y}}^{-1} A \hat{x} - \tilde{E}_A^T Q_{\tilde{y}}^{-1} A \hat{x} = A^T Q_{\tilde{y}}^{-1} y - \tilde{E}_A^T Q_{\tilde{y}}^{-1} y \quad (17)$$

After a few simple mathematical operations, the estimated unknown vector follows directly from the preceding equation as

$$\hat{x} = \left((A - \tilde{E}_A)^T Q_{\tilde{y}}^{-1} A \right)^{-1} (A - \tilde{E}_A)^T Q_{\tilde{y}}^{-1} y \quad (18)$$

This formulation is indeed a different appearance (but equivalent) of the Tong et al. (2011) and Mahboub (2012) formulations. The formulation looks similar to the standard least squares formulation. But, the so-called normal matrix is not in general symmetric and positive-definite, and hence it is not possible to directly derive the covariance matrix of estimated parameters $Q_{\hat{x}}$. To make a symmetric positive-definite normal matrix after substituting for A from $A = \tilde{A} + \tilde{E}_A$ into Eq. (18) and applying a few simple mathematical operations, one obtains the least squares estimate $\hat{x}$ as

$$(A - \tilde{E}_A)^T Q_{\tilde{y}}^{-1} (A - \tilde{E}_A) \hat{x} = (A - \tilde{E}_A)^T Q_{\tilde{y}}^{-1} (y - \tilde{E}_A \hat{x}) \quad (19)$$

or

$$\hat{x} = \left((A - \tilde{E}_A)^T Q_{\tilde{y}}^{-1} (A - \tilde{E}_A) \right)^{-1} \times (A - \tilde{E}_A)^T Q_{\tilde{y}}^{-1} (y - \tilde{E}_A \hat{x}) \quad (20)$$

This is indeed the weighted total least-squares formulation which resembles the standard least squares method in which $\tilde{A} = A - \tilde{E}_A$ plays the role of the design matrix A , $Q_{\tilde{y}} = Q_y + (\hat{x}^T \otimes I_m) Q_A (\hat{x} \otimes I_m)$ plays the role of the covariance matrix Q_y , and $\tilde{y} = y - \tilde{E}_A \hat{x}$ plays the role of the observation vector y . Therefore, we deal with the predicted design matrix $\tilde{A}$, the predicted observation vector $\tilde{y}$, and the covariance matrix $Q_{\tilde{y}}$ of the predicted observations. Equation (20) is then written as

$$\hat{x} = (\tilde{A}^T Q_{\tilde{y}}^{-1} \tilde{A})^{-1} \tilde{A}^T Q_{\tilde{y}}^{-1} \tilde{y} \quad (21)$$

This formulation is similar to that of Shen et al. (2011). Our derivation uses the Lagrange multiplier structure in a linear form rather than the nonlinear Gauss-Newton formulation. The above formulation looks in fact as if to rewrite Eq. (4) in the form of $y - E_A x = (A - E_A) x + e_y - E_A x$, which is equivalent to $\tilde{y} = \tilde{A}x + \tilde{e}$.

The preceding formulation of the WTLS allows one to directly apply the existing body of knowledge of standard least squares theory. For example, without any derivation one obtains the covariance matrix of the estimated parameters $\hat{x}$ as

$$Q_{\hat{x}} = (\tilde{A}^T Q_{\tilde{y}}^{-1} \tilde{A})^{-1} \quad (22)$$

from which the variances and covariances among the estimates can be derived. We note that the standard deviations of the TLS estimates using Schaffrin and Wieser (2008) algorithm can be approximated by linear variance propagation and numerical computation of the required partial derivatives. This however gives a rigorous formulation. We also note that due to the nonlinear nature of the problem, in general, the estimate $\hat{x}$ along with its covariance matrix $Q_{\hat{x}}$ is biased (see Teunissen 1990). The results presented in this contribution however do not show that this bias is significant (see later on Section 3).

In a similar manner, we may further explore the potential capability of the preceding formulation. For example, the estimated vectors of the observations and of the residuals in the model $\tilde{y} = \tilde{A}x + \tilde{e}$, read respectively

$$\begin{cases} \tilde{y} = P_{\tilde{A}} \tilde{y} = \tilde{A} \hat{x} = A \hat{x} - \tilde{E}_A \hat{x} \\ \hat{e} = P_{\tilde{A}}^\perp \tilde{y} = \tilde{y} - \tilde{y} = y - A \hat{x} = Q_{\tilde{y}} \hat{\lambda} \end{cases} \quad (23)$$

where $P_{\tilde{A}} = \tilde{A} (\tilde{A}^T Q_{\tilde{y}}^{-1} \tilde{A})^{-1} \tilde{A}^T Q_{\tilde{y}}^{-1}$ and $P_{\tilde{A}}^\perp = I - \tilde{A} (\tilde{A}^T Q_{\tilde{y}}^{-1} \tilde{A})^{-1} \tilde{A}^T Q_{\tilde{y}}^{-1}$ are two orthogonal projectors. We highlight that $\hat{e}$ is the so-called 'total residuals' of the model, and hence different from $\tilde{e}_y$ in Eq. (12). In addition, the covariance matrices of the least-squares estimates $\hat{y}$ and $\hat{e}$ are $Q_{\hat{y}} = P_{\tilde{A}} Q_{\tilde{y}} = \tilde{A} Q_{\hat{x}} \tilde{A}^T$ and $Q_{\hat{e}} = P_{\tilde{A}}^\perp Q_{\tilde{y}} = Q_{\tilde{y}} - Q_{\hat{y}}$, respectively. Also, note that $P_{\tilde{A}}^\perp = R$ is the reliability matrix that contains the redundancy numbers on its main diagonal, i.e. $r_i = 1 - \left(\tilde{A} (\tilde{A}^T Q_{\tilde{y}}^{-1} \tilde{A})^{-1} \tilde{A}^T Q_{\tilde{y}}^{-1} \right)_{ii}$. The internal and external reliability along with the data snooping procedures can accordingly be established.

In the analogy with the standard least squares, the variance component estimator of the unit weight is given as

$$\hat{\sigma}_0^2 = \frac{\hat{e}^T Q_{\tilde{y}}^{-1} \hat{e}}{m - n} \quad (24)$$

The following expression (25) has already been rigorously proved by Schaffrin and Wieser (2008) and Mahboub (2012) as the estimate of the variance factor of the unit weight

$$\hat{\sigma}_0^2 = \frac{\hat{\lambda}^T Q_{\tilde{y}} \hat{\lambda}}{m - n} \quad (25)$$

Substitution for $\hat{e}$ from Eq. (23) into Eq. (24) provides the estimate of the variance factor of the unit weight of Eq. (25) in a straightforward manner.

To calculate $\tilde{E}_A$ three strategies are recommended.

1. The first strategy is based on Eq. (13) in which one obtains $\tilde{e}_A = \text{vec } \tilde{E}_A = -Q_A (\hat{x} \otimes I_m) Q_{\tilde{y}}^{-1} \hat{e}$. Having $\tilde{e}_A$ available as an nm -vector, one can simply reshape it as an $m \times n$ matrix, i.e. an 'inverse' vec operator. In other words, $\tilde{E}_A = -\text{vec}^{-1}(Q_A (\hat{x} \otimes I_m) Q_{\tilde{y}}^{-1} \hat{e})$, where the operator vec^{-1} reconstructs an mn -vector to an $m \times n$ matrix.
2. The second algorithm is based on considering the following block structure of the Q_A matrix:

$$Q_A = \begin{bmatrix} Q_{11} & \cdots & Q_{1n} \\ \vdots & \ddots & \vdots \\ Q_{n1} & \cdots & Q_{nn} \end{bmatrix} \quad (26)$$

where Q_{ij} , $i = 1, \dots, n$, $j = 1, \dots, n$, are submatrices of size $m \times m$. Q_A is then written as $Q_A = \sum_i^n \sum_j^n c_i c_j^T \otimes Q_{ij}$, where c_i is the canonical unit vector having one at i th position and zeros elsewhere. Substitution in $\text{vec } \tilde{E}_A = -Q_A (\hat{x} \otimes I_m) \hat{\lambda}$ yields

$$\begin{aligned} \text{vec } \tilde{E}_A &= -\sum_i^n \sum_j^n (c_i c_j^T \otimes Q_{ij}) (\hat{x} \otimes I_m) \hat{\lambda} \\ &= -\sum_i^n \sum_j^n (c_i c_j^T \hat{x} \otimes Q_{ij}) \text{vec}(\hat{\lambda}) \\ &= -\text{vec} \left(\sum_i^n \sum_j^n Q_{ij} \hat{\lambda} \hat{x}^T c_j c_i^T \right) \end{aligned} \quad (27)$$

which gives

$$\tilde{E}_A = - \sum_i^n \sum_j^n Q_{ij} \hat{\lambda} \hat{\lambda}^T c_j c_i^T = - \sum_i^n \sum_j^n Q_{ij} Q_y^{-1} \hat{e} \hat{\lambda}^T c_j c_i^T \quad (28)$$

3. The third strategy takes the possible Kronecker structure of the $Q_A = Q_n \otimes Q_m$ (as a special case). In this case one can simply show that

$$\tilde{E}_A = -Q_m Q_y^{-1} \hat{e} \hat{\lambda}^T Q_n \quad (29)$$

When the proper structure of the covariance matrix Q_A is introduced in the EIV model, the full potential capability of the WTLS can be exploited. Schaffrin and Wieser (2008) introduce the proper structure of the covariance matrix within the class of $Q_A = Q_n \otimes Q_m$ in order to consider fairly general covariance matrices where Q_n has size $n \times n$ and Q_m has size $m \times m$. Q_n and therefore Q_A could be singular but Q_m is nonsingular. An example of these cofactor matrices are given in the next section. Mahboub (2012) explains how a more general description of the dispersion matrix of the coefficient matrix A can be implemented in Q_A . For example, if an element of A is repeated (with positive or negative sign), one may use identical variances for both elements, and the 100% or -100% correlation between the two elements. Or, if an element of A is fixed, one may use zero for its corresponding variance. We thus highlight the proper use of the error propagation law in constructing Q_A . This is in conjunction with the classical weighted least squares in which the proper weight matrix based on the correct covariance matrix leads to the best (minimum variance) estimators.

Because all matrices involved along with the observation vector are functions of the unknown vector x , the final estimate can be sought in an iterative procedure. The iterative algorithm for estimating unknown parameters is given as follows:

Step 1: (Initialize x)

$$\hat{x}^0 = (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} y \quad (30)$$

Step 2: (Set $k = 0$ and repeat)

$$\hat{e}^{(k)} = y - A \hat{x}^{(k)} \quad (31)$$

$$Q_y^{-1(k)} = [Q_y + (\hat{x}^{T(k)} \otimes I_m) Q_A (\hat{x}^{(k)} \otimes I_m)]^{-1} \quad (32)$$

$$\left\{ \begin{array}{l} \tilde{E}_A^{(k)} = - \sum_i^n \sum_j^n Q_{ij} Q_y^{-1(k)} \hat{e}^{(k)} \hat{\lambda}^{T(k)} c_j c_i^T \\ \text{or, reshape } \tilde{e}_A^{(k)} \text{ to } \tilde{E}_A^{(k)} \\ \tilde{e}_A^{(k)} = -Q_A (\hat{x}^{(k)} \otimes I_m) Q_y^{-1(k)} \hat{e}^{(k)} \\ \text{or, as special case} \\ \tilde{E}_A^{(k)} = -Q_m Q_y^{-1(k)} \hat{e}^{(k)} \hat{\lambda}^{T(k)} Q_n \end{array} \right. \quad (33)$$

$$\tilde{A}^{(k)} = A - \tilde{E}_A^{(k)}, \quad \tilde{y}^{(k)} = y - \tilde{E}_A^{(k)} \hat{x}^{(k)} \quad (34)$$

$$Q_{\tilde{x}}^{(k+1)} = (\tilde{A}^{T(k)} Q_y^{-1(k)} \tilde{A}^{(k)})^{-1} \quad (35)$$

$$\hat{x}^{(k+1)} = Q_{\tilde{x}}^{(k+1)} \tilde{A}^{T(k)} Q_y^{-1(k)} \tilde{y}^{(k)} \quad (36)$$

$$k := k + 1$$

Step 3: (Check for convergence)

Repeat step 2 until one has

$$\|\hat{x}^{(k+1)} - \hat{x}^{(k)}\| < \in \quad (37)$$

where $\in$ is a chosen threshold value for the convergence.

3. Numerical results and discussions

To verify the efficacy of the presented algorithm, two case studies are provided. Both examples have been widely used in many TLS research papers and are particularly of interest in engineering Surveying and Geomatic applications. The first example is a linear regression model in which real and simulated data sets are used. The second example is a 2-D affine transformation for which simulated weighted datasets have been used. In both examples, the results are compared to the existing WTLS methods along with the results obtained using the nonlinear Gauss-Helmert model.

3.1. Linear regression model

Real data. The first example considers the problem of linear regression model in which variables u and v have been observed: $v_i - e_{v_i} = a(u_i - e_{u_i}) + b$. Therefore errors in both variables are involved. We use the data presented in Neri et al. (1989) and later used by Schaffrin and Wieser (2008) (see Table 1).

We aim to estimate the slope a and intercept b of the regression line using the presented WTLS algorithm. The precision of the estimates along with the correlation coefficient i.e. $\rho = \frac{\sigma_{ab}}{\sigma_a \sigma_b}$, will also be provided. If we define the parameter vector as $x = [a, b]^T$, only the first column of the coefficient matrix in Eq. (4) has random errors, while the values in the second column are fixed. We compose the cofactor matrices as follows: $Q_y = (\text{diag}([W_{v_1}, \dots, W_{v_{10}}]))^{-1}$, and $Q_A = Q_2 \otimes Q_{10}$, where $Q_2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $Q_{10} = (\text{diag}([W_{u_1}, \dots, W_{u_{10}}]))^{-1}$, with 'diag' an operator that converts a vector into a diagonal matrix of which the diagonal entries of the matrix are the vector's elements.

Table 1. Observed points and corresponding weights according to Neri et al. (1989)

Point no.	Observed data		Weights	
	v	u	W_v	W_u
1	5.9	0.0	1.0	1,000.0
2	5.4	0.9	1.8	1,000.0
3	4.4	1.8	4.0	500.0
4	4.6	2.6	8.0	800.0
5	3.5	3.3	20.0	200.0
6	3.7	4.4	20.0	80.0
7	2.8	5.2	70.0	60.0
8	2.8	6.1	70.0	20.0
9	2.4	6.5	100.0	1.8
10	1.5	7.4	500.0	1.0

The threshold $\epsilon = 10^{-12}$ is chosen such that it allows a comparison of our WTLS to the exact solution of Neri et al. (1989) and the nonlinear Gauss-Helmert model. The algorithm used to solve a nonlinear Gauss-Helmert method is described by Vanicek and Krakiwsky (1986), pages 202-207. The method, originally proposed by Pope (1974), approximates the nonlinear model by its linear version and performs iterations. The algorithm presented in this contribution converges after eight iterations. But, starting by the same initial unknown parameters, 15 iterations are required to meet the above threshold for the nonlinear Gauss-Helmert model.

Estimated line parameters using different algorithms are presented in Table 2. As can be seen, both algorithms provide the 'exact solution', given by Neri et al. (1989). The same results have already been reported by Schaffrin and Wieser (2008) on the same data set. The rigorous standard deviations of the presented TLS algorithm along with the correlation coefficient are presented in Table 2 (use is made of $Q_{\hat{x}} = \hat{\sigma}_0^2 \left(\tilde{A}^T Q_{\tilde{y}}^{-1} \tilde{A} \right)^{-1}$). The results of the standard deviations are identical to those obtained by the nonlinear Gauss-Helmert method.

The estimated vectors of the observations and of the (total) residuals, i.e. $\hat{y}$ and $\hat{e}$ in Eq. (23), the value of the objective function $\hat{e}^T Q_{\hat{y}}^{-1} \hat{e} = \hat{e}_y^T Q_y^{-1} \hat{e}_y + \hat{e}_A^T Q_A^{-1} \hat{e}_A$, and the estimated variance factor of the unit weight (Eq. 24) are provided in Table 3. In addition, this table includes the redundancy numbers of the 'equations'. We note that the redundancy numbers are added up to the total redundancy of the linear regression model, namely $\sum_{i=1}^{10} r_i = 8$.

Simulated data. 50 points are simulated in the linear regression model. The covariance matrices of the observations and of the coefficient matrix are, respectively, $Q_y = 0.25 I_{50}$ and $Q_A = Q_2 \otimes Q_{50}$, where $Q_{50} = 0.25 I_{50}$. The line parameters are set to $a = 1$ and $b = 10$ in this example. The u_i 's components are assumed to be $u_i = i$, $i = 1, \dots, 50$, and the v_i 's are calculated based on the line parameters given above. Both components are corrupted by white Gaussian noise using the preceding covariance

matrices. We will present the results of 100,000 independent runs.

The results of line parameters are given based on the algorithm presented in this contribution. The results are identical to those based on algorithms presented in Schaffrin and Wieser (2008), Mahboub (2012), and the Gauss-Helmert method for all of the 100,000 runs. The same threshold $\epsilon = 10^{-12}$ was used for all compared methods and the final results were identical. The algorithm presented in this contribution converges after 5.1 iterations on average, while, starting by the same initial unknown parameters, on average, 9.0 iterations are required for the nonlinear Gauss-Helmert model to meet the above threshold which verify faster convergence of our algorithm than GHM.

The histograms of the estimated parameters along with their standard deviations of the estimates are presented in Fig. 1. The standard deviations are directly obtained from $Q_{\hat{x}}$ of Eq. (22), which are also identical to those provided from the nonlinear Gauss-Helmert method. The mean values of the estimated parameters are $\hat{b} = 9.9979948465$ and $\hat{a} = 1.0000627604$. The standard deviations of the 100,000 estimated b's and a's are $\sigma_{\hat{b}} = 0.20344$ and $\sigma_{\hat{a}} = 0.006916$. These standard deviations closely follow the average (over 100,000 runs) standard deviations obtained from $Q_{\hat{x}}$, i.e. $\sigma_{\hat{b}} = 0.20301$ and $\sigma_{\hat{a}} = 0.006928$.

3.2. Two-dimensional affine transformation

When a set of points are observed in two coordinate systems, the transformation parameters can be estimated in an EIV model using WTLS. The data for the planar linear affine transformation (six-parameter transformation) are simulated. The model is expressed as

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} = \begin{bmatrix} u_s & v_s & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & u_s & v_s & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \\ c_1 \\ a_2 \\ b_2 \\ c_2 \end{bmatrix} \quad (38)$$

where the parameters c_1 and c_2 are the shifts along the u and v axes, respectively. The other parameters a_1, a_2, b_1 and b_2 are related to the four physical parameters of a 2-D linear transformation, which include two scales along the u and v axes, one rotation, and one non-perpendicularity (or affinity) parameter.

The coordinates of a series of points (i.e. $i = 1, \dots, k$ points) are observed in both the start and the target systems. Equation (38) makes in total $m = 2k$ number of equations and six number of unknown parameters to be estimated. The observation vector y and the design matrix A are

Table 2. Estimated straight-line parameters along with their standard deviations and correlation coefficient using data of Table 1; Exact solution by Neri et al. (1989), nonlinear Gauss-Helmert, and present contribution.

Parameter/ standard deviation	Exact solution (Neri et al.)	Nonlinear Gauss- Helmert model	WTLS (This paper)
$\hat{a}$	-0.480533407	-0.4805334074	-0.4805334074
$\hat{b}$	5.47991022	5.4799102240	5.4799102240
$\sigma_{\hat{a}}$	—	0.0706202698	0.0706202698
$\sigma_{\hat{b}}$	—	0.3592465226	0.3592465226
$\rho_{\hat{a}\hat{b}}$	—	-0.9630881337	-0.9630881337

Table 3. Estimated least squares vectors $\hat{y}$ and $\hat{e}$ along with value of objective function, and variance factor of the unit weight. Indicated in the table are also redundancy numbers r_i of equations.

Point no.	$\hat{e} [m]$	$\hat{y} [m]$	r_i
1	0.4200897760	5.4800072056	0.9130057041
2	0.3525698427	5.0475766394	0.8918851255
3	-0.2149500906	4.6145537457	0.8459521198
4	0.3694766353	4.2313745663	0.8075900023
5	-0.3941499795	3.8852539888	0.7125177500
6	0.3344367687	3.3838159332	0.8624392888
7	-0.1811365053	2.9426948374	0.6254534330
8	0.2513435614	2.6609974005	0.6342229529
9	0.0435569244	2.3968501980	0.8984788418
10	-0.4239630089	1.5036405369	0.8084547817
$OF = 11.86635319$		$\hat{\sigma}_0^2 = 1.4832941492$	Sum = 8

$$y = \begin{bmatrix} u_{t_1} \\ v_{t_1} \\ \vdots \\ u_{t_k} \\ v_{t_k} \end{bmatrix}, \quad A = \begin{bmatrix} u_{s_1} & v_{s_1} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & u_{s_1} & v_{s_1} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_{s_k} & v_{s_k} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & u_{s_k} & v_{s_k} & 1 \end{bmatrix} \quad (39)$$

The vector e_y of observational noise and the matrix E_A of the coefficient matrix random noise can be defined with an identical structure to the preceding equations. We further assume that the noise measurements in both the start and target systems is the independent Gaussian white noise with variances σ_s^2 and σ_t^2 , respectively. The dispersion matrix of the coefficient matrix A reads

$$Q_A = \begin{bmatrix} Q_{11} & 0 & 0 & Q_{14} & 0 & 0 \\ 0 & Q_{22} & 0 & 0 & Q_{25} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ Q_{41} & 0 & 0 & Q_{44} & 0 & 0 \\ 0 & Q_{52} & 0 & 0 & Q_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (40)$$

where

$$Q_{11} = Q_{22} = \sigma_s^2 I_k \otimes \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

$$Q_{44} = Q_{55} = \sigma_s^2 I_k \otimes \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

$$Q_{14} = Q_{25} = \sigma_s^2 I_k \otimes \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},$$

$$Q_{41} = Q_{52} = \sigma_s^2 I_k \otimes \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix},$$

Figure 1. Histograms of estimated intercept (top-left) and estimated slope (top-right) of linear regression model from 100,000 independent simulated data sets. The frames at bottom show standard deviations of intercept (bottom-left) and slope (bottom-right).

We note that in the second application of Tong et al. (2011) in which they applied a WTLS algorithm to an affine transformation, the above-mentioned structure of the covariance matrix is not used.

One method that can be used to solve this problem is the multivariate WTLS approach that was proposed by Schaffrin and Wieser (2009). This is accomplished by changing Eq. (39) into a multivariate model. The algorithm is restricted to the class of covariance matrices $Q_A = Q_n \otimes Q_m$. The alternative is to use the algorithm proposed by Mahboub (2012) in which a complete structure of the dispersion matrix can be used. Our results are identical to the results of these two algorithms and hence we will not repeat them in this contribution. We however make a comparison between our results and those obtained by the Gauss-Helmert method. Particular attention is paid to the standard deviation of the estimates using these two methods.

Suppose that the coordinates of $k = 20$ points in the start system

are transformed by the parameters $a_1 = 2$, $b_1 = -1$, $c_1 = 0$, $a_2 = -1$, $b_2 = 2$ and $c_2 = 0$ into the coordinates in the target system. The errorless coordinates in the start and target systems are shown in Fig. 2. The coordinates of the points in the start and target systems are corrupted by white Gaussian noise with variances $\sigma_s^2 = 0.01$ and $\sigma_t^2 = 0.02$, respectively. This process has been repeated over 100,000 independent runs. For each of the simulated data set, the transformation parameters are estimated using the WTLS method proposed in this paper and the nonlinear Gauss-Helmert method. The same threshold $\epsilon = 10^{-12}$ was used for both methods and the final results were identical. The algorithm presented in this contribution converges after 5.8 iterations on average, while, starting by the same initial unknown parameters, on average, 6.6 iterations are required for the nonlinear GHM to meet the above threshold.

The histogram of the estimated parameters is given in Fig. 3. The average (over 100,000 runs) values of the estimated param-

Figure 2. Coordinates of 20 points in start and target systems.

eters are $a_1 = 2.0000099697$, $b_1 = -1.0000109115$, $c_1 = 0.0000934864$, $a_2 = -0.9999816046$, $b_2 = 2.0000040795$, and $c_2 = -0.0012505003$. The histograms of the standard deviations of the estimates are presented in Fig. 4. The mean values (over 100,000 runs) are $\sigma_{\hat{a}_1} = 0.0041833081$, $\sigma_{\hat{b}_1} = 0.0041833307$, $\sigma_{\hat{c}_1} = 0.3016628024$, $\sigma_{\hat{a}_2} = 0.0041832793$, $\sigma_{\hat{b}_2} = 0.0041833108$, and $\sigma_{\hat{c}_2} = 0.3016610071$. The standard deviations are directly obtained from $Q_{\hat{x}}$ of Eq. (22), which are also identical to those provided from the nonlinear Gauss-Helmert method. These standard deviations closely follow the standard deviations of the 100,000 estimated a_1 , b_1 , c_1 , a_2 , b_2 , and c_2 , which are $\sigma_{\hat{a}_1} = 0.0041879403$, $\sigma_{\hat{b}_1} = 0.0041813018$, $\sigma_{\hat{c}_1} = 0.3017729384$, $\sigma_{\hat{a}_2} = 0.0041902274$, $\sigma_{\hat{b}_2} = 0.0041889952$, and $\sigma_{\hat{c}_2} = 0.3021218549$, respectively. These results indicate that the bias explained by Shen et al. (2011) is not significant for the two applications considered in the present contribution.

In the sequel, we consider the performance of the presented method in the case of fully populated covariance matrices of an affine transformation. The fully populated covariance matrices of coordinates of points in the start and target systems i.e. $Q_{u_s v_s}$ and $Q_{u_t v_t}$ are constructed using MATLAB built-in function `randn` as $(\text{randn}(100, 2 * k))^T \times \text{randn}(100, 2 * k)$. The noise is constructed accordingly. In this case, the dispersion matrix of the coefficient matrix A reads

$$Q_A = \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \\ Q_5 \\ Q_6 \end{bmatrix} Q_{u_s v_s} \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \\ Q_5 \\ Q_6 \end{bmatrix}^T \quad (41)$$

where

$$Q_1 = I_k \otimes \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

$$Q_2 = I_k \otimes \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},$$

$$Q_3 = I_k \otimes \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

$$Q_4 = I_k \otimes \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix},$$

$$Q_5 = I_k \otimes \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

$$Q_6 = I_k \otimes \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

The same threshold $= 10^{-12}$ was used for both methods and the coordinates of 20 points in the start system are transformed by the parameters $a_1 = 2$, $b_1 = -1$, $c_1 = 0$, $a_2 = -1$, $b_2 = 2$ and $c_2 = 0$ into the coordinates in the target system (as explained before). Over 10000 independent runs, on average, our algorithm converges after 30.4 iterations, while, starting by the same initial unknown parameters, 40.6 iterations are required for the nonlinear GHM to meet the above threshold. Both algorithms provide identical results for all independent runs which show the performance of the proposed method in the case of fully populated covariance matrices.

4. Conclusions and outlook

In this contribution we showed that the WTLS problem is an extension of the WLS problem. A new WTLS algorithm was formulated which is based on the well-known theory of the standard least squares problem. The WTLS problem uses the complete description of the covariance matrices of the observations, both of the observation vector and of the coefficient matrix in which a proper propagation law of the errors is applied. The efficacy of the proposed WTLS algorithm was demonstrated by solving two commonly used WTLS problems in geodetic literature, namely a linear regression model and 2-D affine transformation using real and simulated data. The results were shown to be identical to those provided by the nonlinear GHM.

Numerical results on all real and simulated data, including those with fully populated covariance matrices, showed faster convergence rate of our algorithm than the GHM. We also note that the

Figure 3. Histograms of estimated parameters of two-dimensional affine transformation from 100,000 independent simulated data sets.

solution of the nonlinear GHM is somewhat critical to handle because special attention needs to be paid to appropriate linearization and iteration of the model. The elegance of the WTLS algorithm lies on its simplicity in the sense that it is formulated in the standard GMM and avoids immediate linearization because the estimates are obtained based on a linearly-structured iterative algorithm. We should however note that the WTLS algorithms require the covariance matrix Q_A , which is of (large) size of $mn \times mn$. But the GHM directly uses the covariance matrix of the functionally in-

dependent observations (usually smaller than $mn \times mn$). We also note that making the covariance matrix Q_A is rather a tricky experience in some cases.

The proposed algorithm was shown to be simple in the concept, easy in the implementation, and attractive and flexible in comparison with the standard least squares theory. The presented method can be used as an alternative WTLS method to the existing WTLS methods for computing the exact solution. Using the standard least squares theory, the covariance matrix of the estimates can di-

Figure 4. Histograms of standard deviations of estimated parameters of two-dimensional affine transformation from 100,000 independent simulated data sets.

rectly be obtained. This formulation simply led us to generalize the orthogonal projectors of the standard least squares from which estimates for the total residuals and observations along with their covariance matrices were obtained. The variance of the unit weight was accordingly estimated. This formulation allows one to obtain the internal and external reliability and to apply the data snooping procedures for identification of outlying measurements. Also, further research is in progress for other kinds of WTLS problems such as the constrained WTLS and the variance component estimation

for an EIV model.

Acknowledgements

The editorial board of Journal of Geodetic Science and three anonymous reviewers are kindly acknowledged for their helpful comments.

References

- Golub G. and Van Loan C., 1980, An analysis of the total least squares problem, *SIAM J. Num. Anal.* 17, 883-893
- Mahboub V., 2012, On structured weighted total least-squares for geodetic transformations, *J Geod.*, 5, 359-367 DOI: 10.1007/s00190-011-0524-5.
- Neitzel F., 2010, Generalization of total least-squares on example of unweighted and weighted 2D similarity transformation, *J Geod.*, 84, 751-762.
- Neri F., Saitta G. and Chiofalo S., 1989, An accurate and straightforward approach to line regression analysis of error-affected experimental data, *J Phys. Ser. E: Sci. Instr.* 22, 215-217
- Pope A. J., 1972, Some pitfalls to be avoided in the iterative adjustment of nonlinear problems. In: *Proceedings of the 38th Annual Meeting of the American Society of Photogrammetry* (Washington, DC) 449-477.
- Pope A. J., 1974, Two approaches to nonlinear least squares adjustments, *Can. Surv.*, 28, 5, 663-669.
- Schaffrin B., 2006, A note on constrained total least-squares estimation, *Linear Alg. Appl.* 417, 245-258.
- Schaffrin B. and Felus Y., 2009, An algorithmic approach to the total least-squares problem with linear and quadratic constraints, *Stud. Geophys. Geod.*, 53, 1-16.
- Schaffrin B., Lee I., Felus Y. and Choi Y., 2006, Total least-squares for geodetic straight-line and plane adjustment, *Boll. Geod. Sci. Aff.*, 65, 141-168.
- Schaffrin B. and Wieser A., 2008, On weighted total least-squares adjustment for linear regression, *J Geod.*, 82, 7, 415-421.
- Schaffrin B. and Wieser A., 2009, Empirical affine reference frame transformations by weighted multivariate TLS adjustment. In: Drewes H. (ed) *International Association of geodesy symposia*, vol 134, *Geodetic reference frames* (Springer, Berlin) 213-218.
- Schaffrin B. and Wieser A., 2011, Total least-squares adjustment of condition equations, *Stud. Geophys. Geod.*, 55, 529-536
- Shen Y., Li B. and Chen Y., 2011, An iterative solution of weighted total least-squares adjustment, *J Geod.*, 85, 229-238.
- Teunissen P.J.G., 1990, Nonlinear least-squares, *Manus. Geod.*, 153, 3, 137-150.
- Teunissen P.J.G., 2000, *Adjustment theory: an introduction*. Delft University Press, Delft University of Technology, Series on Mathematical Geodesy and Positioning
- Tong, X., Jin Y. and Li L., 2011, An improved weighted total least squares method with applications in linear fitting and coordinate transformation. *J Surv. Eng.*, 137, 4, 120-128.
- Van Huffel S. and Vandewalle J., 1991, *The total least-squares problem. Computational aspects and analysis* (SIAM, Philadelphia) Vanicek P., Krakiwsky E., 1986, *Geodesy: The concepts* (North-Holland, New York, NY).
- Xu P., Liu J. and Shi C., 2012, Total least squares adjustment in partial errors-in-variables models: algorithm and statistical analysis, *J Geod.*, DOI: 10.1007/s00190-012-0552-9.