# TOTAL LEAST SQUARES?

Keijo Inkilä
Helsinki University of Technology
Institute of Photogrammetry and Remote Sensing

## ABSTRACT

*The paper reviews the Total Least Squares Method, which has a long history in statistics and in numerical analysis, but is less-known in photogrammetry. In contrast to the ordinary Least Squares Problem, where errors are confined in the observation vector only, it is assumed in the TLS Problem that the coefficient matrix of parameters is also contaminated by errors. The solution of the TLS Problem using the Singular Value Decomposition is shown. Particular attention is paid to the geometry of TLS and differences between LS and TLS.*

## 0. INTRODUCTION

The (ordinary) *Least Squares Problem* (LS-Problem) is a totally familiar problem to photogrammetrists. But, what is the *Total Least Squares Problem* (TLS-Problem)? We start the answering by defining the LS-problem in a way that is perhaps less familiar, but that is appropriate in this context:

Given an overdetermined system $\mathbf{Ax} \cong \mathbf{b}$ ( $\mathbf{A}$ mxn-matrix, m>n, $\mathbf{b}$ m-vector), find $\mathbf{x}$ that

$$\text{minimizes} \quad \|\mathbf{v}\|_2^2 = \|\mathbf{Ax}-\mathbf{b}\|_2^2 \tag{1}$$

$$\text{subject to} \quad \mathbf{b}+\mathbf{v} \in \mathbf{R}(\mathbf{A}) \tag{2}$$

In (2) R(**A**) denotes the column space (range) of **A**, i.e., the subspace of $R^m$ spanned by the columns of **A** $(R(\mathbf{A}) \subset R^m)$.

The underlying assumption in the ordinary LS-problem is that only the vector **b** (the observation vector) contains errors and consequently corrections are applied only on its components. The matrix **A** (the design matrix) is assumed to be error-free. However, this assumption is frequently unrealistic: the matrix **A** is also contaminated by errors and corrections should be thus applied on the elements of it also. This is done in the Total Least Squares Problem (TLS Problem).

The name "Total Least Squares" appeared only in the early 80's in the literature, but the method has a long history in statistics, where it is known as *orthogonal regression* or *errors-in-variables regression* (it is useful to take a look at the example in Sec. 5 at this point). In numerical analysis, TLS Problem was first studied by Golub and Van

Loan, who have published several papers on the subject. Their work is strongly based on the use of the *Singular Value Decomposition* (SVD). Later, several authors (e.g. Van Huffel and Vandewalle) have analyzed the method and essentially extended it.

The paper seeks to provide a short introduction to TLS based mainly on references (Golub & Van Loan, 1989) and (Van Huffel and Vandewalle, 1991). First, the TLS-Problem is formally stated in Chap. 1. Chap. 2 shows how to solve the TLS-Problem using the SVD (fundamentals of the SVD are given in Appendix). In Chap. 3 the LS- and TLS-Problems are compared from a geometric point of view. Some extensions of the (basic) TLS-Problem are enumerated in Chap. 4. A simple line fitting example is given in Chap. 5. Finally, in Chap. 6 some remarks are made on topics, whose more extensive treatment is beyond the scope of this paper.

## 1. DEFINITION OF THE TLS-PROBLEM

The TLS-problem is defined as follows: ($\|.\|_F$ denotes the Frobenius norm, see the appendix for the definition)

Given an overdetermined system $\mathbf{Ax}\cong\mathbf{b}$ ($\mathbf{A}$ mxn-matrix, m>n, rank($\mathbf{A}$)=n (full)), the TLS-problem seeks to

$$\text{minimize} \quad \|[\mathbf{E}\ \mathbf{v}]\|_F \tag{3}$$

$$\text{subject to} \quad \mathbf{b}+\mathbf{v} \in \mathbf{R}(\mathbf{A}+\mathbf{E}) \tag{4}$$

If $[\hat{\mathbf{E}}\ \hat{\mathbf{v}}]$ is the solution of (3) and $\hat{\mathbf{A}}=\mathbf{A}+\hat{\mathbf{E}}$ and $\hat{\mathbf{b}}=\mathbf{b}+\hat{\mathbf{v}}$, then any $\hat{\mathbf{x}}$ satisfying

$$\hat{\mathbf{A}}\hat{\mathbf{x}} = \hat{\mathbf{b}} \tag{5}$$

is a TLS-solution.

## 2. SOLUTION OF THE TLS-PROBLEM

Rewrite $\mathbf{Ax}\cong\mathbf{b}$ in the following form

$$[\mathbf{A}\ \ \mathbf{b}]\begin{bmatrix}\mathbf{x}\\ \mathbf{-1}\end{bmatrix} \cong \mathbf{0} \tag{6}$$

If rank($[\mathbf{A}\ \ \mathbf{b}]$)=n+1 (this amounts to $\mathbf{b}\notin$R($\mathbf{A}$), which is generally the case), then dim(R($[\mathbf{A}\ \mathbf{b}]^T$)=n+1 and dim(N($[\mathbf{A}\ \mathbf{b}]$))=0. Thus (6) obviously has no solution. To obtain a solution the following steps have to be taken:

1. *Approximate* $[\mathbf{A}\ \mathbf{b}]$ by the rank n matrix $[\hat{\mathbf{A}}\ \hat{\mathbf{b}}]$, which is the best approximation with respect to the Frobenius norm.

2. Of all vectors $\mathbf{y}$ that lie in the orthogonal complement of R($[\hat{\mathbf{A}}\ \hat{\mathbf{b}}]^T$),that is, in the null space N($[\hat{\mathbf{A}}\ \ \hat{\mathbf{b}}]$) and thus satisfy $[\hat{\mathbf{A}}\ \ \hat{\mathbf{b}}]\mathbf{y}=\mathbf{0}$, *choose* the one, whose last component is -1. Since N($[\hat{\mathbf{A}}\ \hat{\mathbf{b}}]$) has dimension one, the solution must be unique.

The solution of the TLS Problem is rather straightforward using the SVD. Let

$$[\mathbf{A}\ \mathbf{b}] = \mathbf{U}\mathbf{S}\mathbf{V}^T \tag{7}$$

be the SVD of $[\mathbf{A}\ \mathbf{b}]$. If $s_{n+1} \neq 0$, then rank($[\mathbf{A}\ \mathbf{b}]$)=n+1 and $R([\mathbf{A}\ \mathbf{b}]^T)=R^{n+1}$. Consequently, $N([\mathbf{A}\ \mathbf{b}])=\{0\}$ and there is no solution to (6). According to the Eckart-Young-Mirsky Theorem the best rank n approximation is given by

$$[\hat{\mathbf{A}}\ \hat{\mathbf{b}}] = \mathbf{U}\hat{\mathbf{S}}\mathbf{V}^T \quad \text{with } \hat{\mathbf{S}} = \mathrm{diag}(s_1,s_2,\ldots,s_n,0) \tag{8}$$

This completes the step 1. Since the last right singular vector $\mathbf{v}_{n+1}$ is a basis of $N([\hat{\mathbf{A}}\ \hat{\mathbf{b}}])$, the TLS-solution is obtained simply by scaling $\mathbf{v}_{n+1}$ until its last component is -1,

$$\begin{bmatrix}\hat{\mathbf{x}}\\ -1\end{bmatrix} = \frac{-1}{v_{n+1,n+1}}\mathbf{v}_{n+1} \tag{9}$$

It can be shown that the TLS solution exists and is unique, if

$$s_n > s_{n+1} \quad \text{and} \quad v_{n+1,n+1} \neq 0 \tag{10}$$

For the (elementary) quality evaluation two results are obtained easily. First, the correction matrix is

$$[\hat{\mathbf{E}}\ \hat{\mathbf{v}}] = s_{n+1}\mathbf{u}_{n+1}\mathbf{v}_{n+1}^T \tag{11}$$

Note that this correction matrix has rank one: all columns are parallel to $\mathbf{u}_{n+1}$. Second, the minimum sum of the squared corrections is

$$\| [\hat{\mathbf{E}}\ \hat{\mathbf{v}}] \|_F = s_{n+1} \tag{12}$$

Finally, we give two useful characterizations of the TLS-solution $\hat{x}$ and minimum correction $s_{n+1}$:

$$\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A} - s_{n+1}^2\mathbf{I})^{-1}\mathbf{A}^T\mathbf{b} \tag{13}$$

and

$$s_{n+1}^2(1+c) = \|\mathbf{A}\hat{\mathbf{x}}_{LS} - \mathbf{b}\|_2^2 \quad \textit{with} \quad c > 0 \tag{14}$$

The comparison of (13) with the solution of the LS-Problem ($x_{LS}=(A^TA)^{-1}A^Tb$) suggests that TLS is a "deregularizing" procedure. Indeed, the TLS-Problem may be very ill-conditioned, if **A** is nearly rank-deficient or the equation system is highly incompatible.

## 3. GEOMETRICAL COMPARISON OF THE LS- AND TLS-PROBLEMS

Geometrical considerations are very useful in revealing the differences between LS and TLS. The most complete understanding can be obtained by performing such considerations in both $R^m$ (column space approach) and $R^{n+1}$ (row space approach).

It is well-known that in $R^m$ LS amounts to projecting **b** orthogonally to the subspace R(**A**), where **A** is a given (known) matrix (Fig. 1). On the other hand in TLS both the columns of **A** and the vector **b** are projected orthogonally to R($\hat{\mathbf{A}}$), where $\hat{\mathbf{A}}$ is a matrix whose determination is a part of the solution.

In $R^{n+1}$ (Fig. 2) step 1 amounts to finding the best (in the sense of (3)) fitting subspace (hyperplane) of dimension n to the row vectors of **[A b]**. In LS corrections are applied to the last components of the rows only (correction vectors are parallel to the $x_3$ axis in Fig. 2), while in TLS they are applied to all components (correction vectors are orthogonal to the fitted hyperplane and thus parallel to the normal of the hyperplane).In step 2 the solution is found at the intersection of N([$\hat{\mathbf{A}}$ $\hat{\mathbf{b}}$]) (normal of the hyperplane) with the hyperplane $x_{n+1}$=-1.

## 4. EXTENSIONS TO THE BASIC TLS-PROBLEM

The basic TLS-Problem can be extended e.g. in the following ways:

1. The right-hand side vector **b** can be replaced by a matrix **B** (multidimensional TLS Problem).
2. If required (small singular values coincide), the uniqueness of the solution can be preserved by choosing the minimum-norm solution.
3. The weighted TLS solution is obtained by pre- and postmultiplication of **[A b]** by non-singular matrices.
4. Some columns of **A** can be regarded as error free (mixed LS-TLS Problem).

The mixed LS-TLS Problem is considered briefly in the following.

Let **A** =[$\mathbf{A}_1$ $\mathbf{A}_2$] be mxn-matrix, whose first p columns have no error and have full rank. The basic idea is to eliminate the first p parameters by orthogonal transformations (inconsistent system!), solve the reduced TLS-Problem and finally solve the reduced LS-Problem. The following algorithm performs these steps.

1. Compute the partial QR-factorization of [$\mathbf{A}_1\mathbf{A}_2\mathbf{b}$] so that

$$\mathbf{Q}^T[\mathbf{A}_1 \ \ \mathbf{A}_2 \ \ \mathbf{b}] = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \mathbf{R}_{13} \\ \mathbf{0} & \mathbf{R}_{22} & \mathbf{R}_{23} \end{bmatrix} \qquad (15)$$

   where $\mathbf{R}_{11}$ is a pxp upper triangular matrix.

2. Compute the TLS-solution of the system

$$\mathbf{R}_{22}\mathbf{x}_2 \cong \mathbf{R}_{23} \qquad (16)$$

   (This yields the last n-p components of the solution vector.)

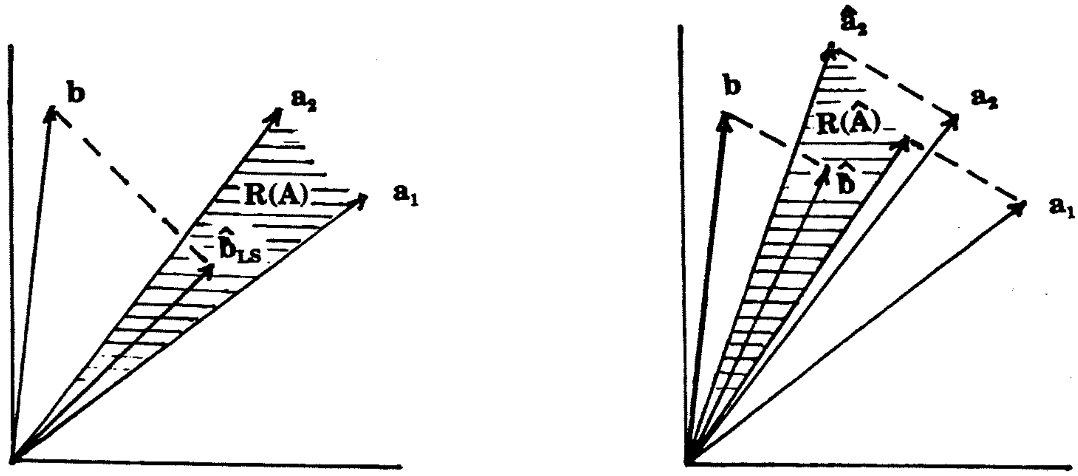


**FIGURE 1.** Geometrical illustration of LS (left) and TLS (right) in $R^m$ (m=3, n=2). Note that correction vector is orthogonal to R(**A**) in LS and that the correction vectors are orthogonal to R(**Â**) in TLS.

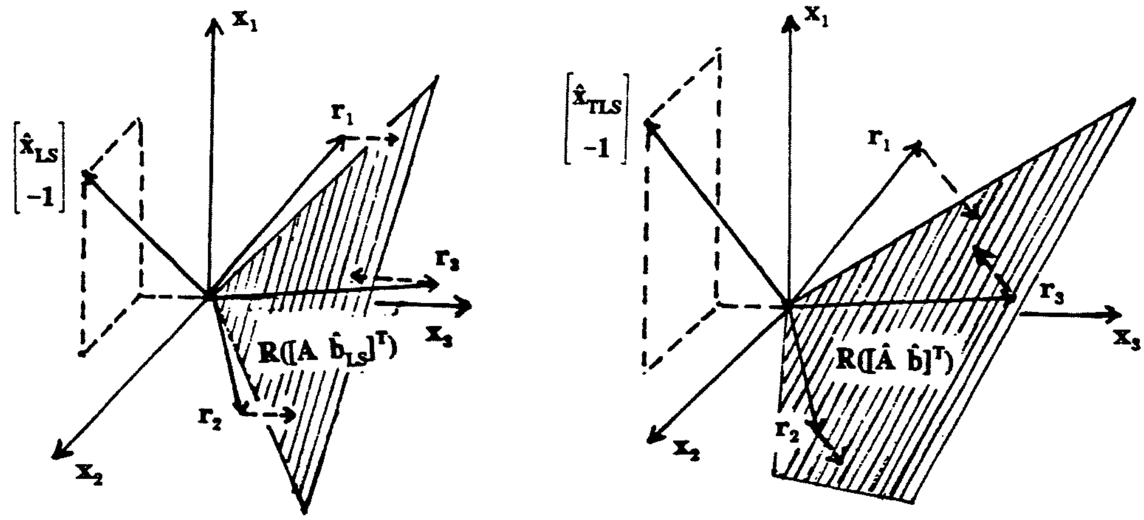


**FIGURE 2.** Geometrical illustration of LS (left) and TLS (right) in $R^{n+1}$ (row space). Vectors $\mathbf{r}_i$ are the rows of **[A b]**. Note that the solution vectors are orthogonal to fitted hyperplanes (shaded planes) in both case. In TLS the corrections are also orthogonal to the same plane, but in LS the corrections are parallel to the last coordinate axis ( only the last components of row vectors can vary).

3. Solve

$$\mathbf{R}_{11}\mathbf{x}_1 = \mathbf{R}_{13} - \mathbf{R}_{12}\hat{\mathbf{x}}_2 \qquad (15)$$

(This yields the first p components of the solution vector by simple back-substitution.)

Although more complicated algorithmically, the computation of the SVD (the heaviest part of the routine) is, on the other hand, in mixed LS-TLS smaller task than in TLS.

## 5. AN EXAMPLE

Consider the fitting of a straight line y=b+ax into the three points, whose measured coordinates are (1,2), (2,6) and (6,1). This leads to the overdetermined system

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 6 \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} \approx \begin{bmatrix} 2 \\ 6 \\ 1 \end{bmatrix}$$

The solution of the LS-Problem (x coordinates are assumed error free) yields a=-0.5 and b=4.5. The mixed LS-TLS solution (the first column of **A** is regarded error free) is a=-1 and b=6. The resulting lines and residuals have been presented in Fig. 3. Note in particular that the residual vectors are orthogonal to the line in TLS-fitting.

It is interesting to note that the mixed LS-TLS gives the same solution as the so-called *eigenvector line fitting* (Duda & Hart 1973). On the other hand, TLS should not be mixed up with the so-called General Least Squares method. The latter method, although giving orthogonal residual vectors, gives the same line as ordinary LS.

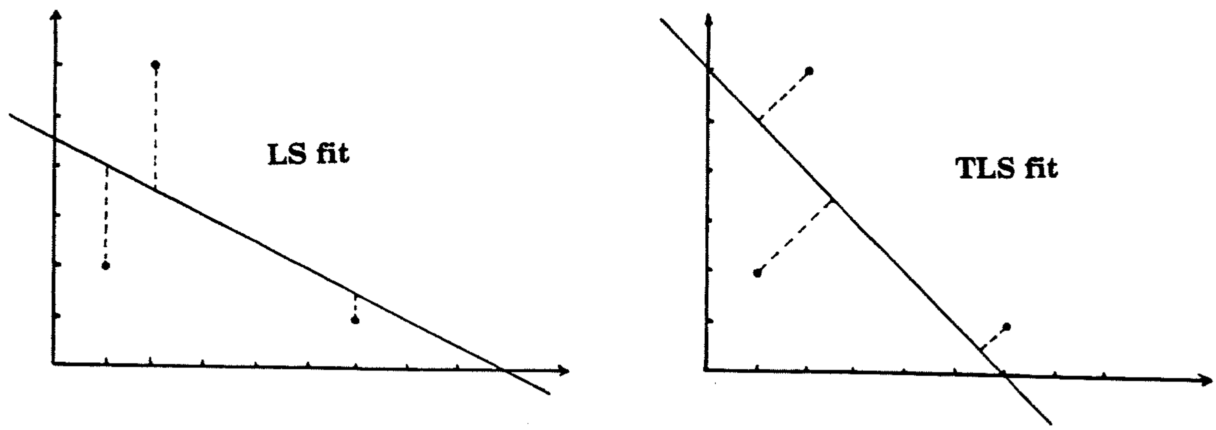


**FIGURE 3.** The fitting of a line to three points using LS (left) and TLS (right).

## 6. CONCLUDING REMARKS

To conclude, we make some remarks about the issues, whose more extensive treatment was beyond the scope of this paper.

*LS or TLS?*

Generally speaking, consider TLS, if the primary interest is the *determination* of the parameters of a model rather than the *prediction* of values of one variable based on other variables. Of course, other aspects (numerical, statistical) may dictate the decision.

*Applications of TLS in photogrammetry?*

Referring to the previous item, it can be expected that there are applications to TLS in photogrammetry. The well-known DLT-model may serve as an example from analytical photogrammetry (although this model fails to fulfil the conditions for favourable statistical estimates). Most data fitting problems in digital photogrammetry are also potential applications of TLS.

*Statistical properties of TLS?*

By making assumptions on errors in observations, statistical properties of the TLS solution can be derived. These are of great importance to photogrammetry, but their consideration is beyond the scope of this paper. Suffice it to say here that the subject is well covered in literature, since TLS has its roots in statistics.

## REFERENCES

Duda, R., Hart, P.: Pattern Classification and Scene Analysis. John Wiley & Sons, New York, 1973.

Golub, G.H., van Loan, C.F.: Matrix Computations. The Johns Hopkins University Press, Baltimore, 1989.

Nievergelt, Y.: Total Least Squares: State-of-the-art Regression in Numerical Analysis. SIAM Review 36, June 1994, pp. 258-264.

Van Huffel, S., Vandewalle, J.: The Total Least Squares Problem: Computational Aspects and Analysis. SIAM, Philadelphia, 1991.

## APPENDIX: SINGULAR VALUE DECOMPOSITION

Let **A** be a real mxn matrix. Then there exist *orthogonal* matrices **U** (mxm) and **V** (nxn) such that

$$\mathbf{U}^T\mathbf{A}\mathbf{V} = \mathbf{S} \quad \text{or} \quad \mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T \tag{A1}$$

where **S** is a *diagonal* mxn matrix with elements $s_1 \geq s_2 \geq \ldots \geq s_n$.

The orthogonal decomposition in (A1) is called the *Singular Value Decomposition* (SVD). The diagonal elements $s_i$ are the *singular values*, the columns $u_i$ are the *left singular vectors* and the columns $v_i$ are the *right singular vectors*.

The proof of the SVD is given e.g. in [Golub & van Loan]. There is also given a stable and efficient algorithm for computing the SVD. The computation essentially consists of two phases: A matrix is first reduced to the bidiagonal form and then to the diagonal form in an *iterative* process. This so-called QR-algorithm is a special adaptation of the famous algorithm of the same name for computing the eigenvalues and the eigenvectors of a symmetric matrix. This is a reflection of the fact that, if $\mathbf{A}=\mathbf{U}\mathbf{S}\mathbf{V}^T$ is the SVD of **A**, then $\mathbf{A}^T\mathbf{A}=\mathbf{V}\mathbf{S}^2\mathbf{V}^T$ is the Eigenvalue Decomposition of $\mathbf{A}^T\mathbf{A}$.

The SVD reveals a great deal about the structure of a matrix. If

$$s_1 \geq \ldots \geq s_r > s_{r+1} = \ldots = s_n = 0 \tag{A2}$$

then

$\text{rank}(\mathbf{A}) = r$
$R(\mathbf{A}) = R(\mathbf{U}_1)$ ($\mathbf{U}_1$ contains the first r columns of **U**)
$N(\mathbf{A}) = R(\mathbf{V}_2)$ ($\mathbf{V}_2$ contains the last n-r columns of **V**)
$R(\mathbf{A}^T) = R(\mathbf{V}_1)$ ($\mathbf{V}_1$ contains the first r columns of **V**)
$N(\mathbf{A}^T) = R(\mathbf{U}_2)$ ($\mathbf{U}_2$ contains the last m-r columns of **U**)

Thus, the SVD gives orthogonal bases of the four interesting subspaces: *range space* and *null space (kernel)* of a matrix and its transpose. This is of great importance for the solution of both LS and TLS problems.

The SVD also plays important role in matrix approximation problems. First, the Frobenius norm and the 2-norm are neatly characterized in terms of the SVD:

$$\|\mathbf{A}\|_F^2 \doteq \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 = s_1^2 + \ldots + s_n^2 \tag{A3}$$

$$\|\mathbf{A}\|_2 \doteq \sup \frac{\|\mathbf{A}y\|_2}{\|y\|_2} = s_1 \tag{A4}$$

Second, a matrix can be written as the sum of the rank one matrices (the SVD expansion):

$$\mathbf{A} = \sum_{i=1}^{n} s_i \mathbf{u}_i \mathbf{v}_i^T \quad \text{(A5)}$$

Third, concerning the approximation of a matrix by another of *lower* rank there is the following Eckart-Young-Mirsky Theorem:

Let

$$\mathbf{A} = \sum_{i=1}^{n} s_i \mathbf{u}_i \mathbf{v}_i^T \quad \text{(A6)}$$

be the SVD expansion of **A**. If k<n and

$$\mathbf{A}_k = \sum_{i=1}^{k} s_i \mathbf{u}_i \mathbf{v}_i^T \quad \text{(A7)}$$

then

$$\min_{rank(D)=k} \|\mathbf{A}-\mathbf{D}\|_F = \|\mathbf{A}-\mathbf{A}_k\|_F = \sqrt{\sum_{i=k+1}^{n} s_i^2} \quad \text{(A8)}$$

This theorem is of particular importance to TLS.