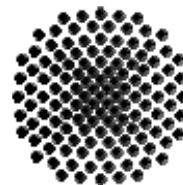
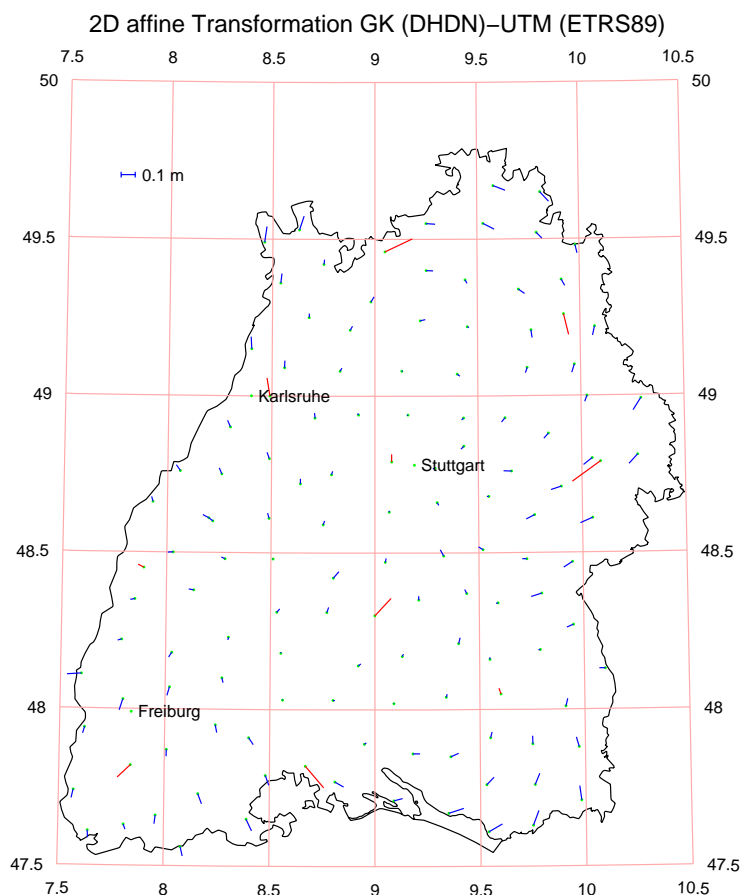


Universität Stuttgart

Geodätisches Institut



Systematical analysis of the transformation procedures in Baden-Württemberg with Least Squares and Total Least Squares methods



Studienarbeit im Studiengang
Geodäsie und Geoinformatik
an der Universität Stuttgart

Ronggang Guo

Stuttgart, 2007

Betreuer:

Dr.-Ing. Jianqing Cai
Prof. Dr.-Ing. Nico Sneeuw
Universität Stuttgart

Selbstständigkeitserklärung

Hiermit erkläre ich, Ronggang Guo, dass ich die von mir eingereichte Studienarbeit zum Thema

“Systematical analysis of the transformation procedures in Baden-Württemberg with Least Squares and Total Least Squares methods”

selbstständig verfasst und ausschließlich die angegebenen Hilfsmittel und Quellen verwendet habe.

Datum, Ort

Unterschrift

Acknowledgment

It is a pleasure to acknowledge Dr.-Ing. Jianqing Cai and Prof. Dr.-Ing. Nico Sneeuw (Geodetic Institute, University Stuttgart), for their guidance, discussion, advice and also encouragement, without their supervising I couldn't finish my thesis in time.

Especially thanks would be given to Dr.-Ing. Jianqing Cai for his supervising to remove the difficulty and spending time to discuss it with me. In addition, I want to say thanks to Prof. Dr.-Ing. Nico Sneeuw for his serious checking of this thesis.

It is also necessary to mention all the help from my friends and colleagues, from them I have gotten many advices and ideas. Finally, I would like to dedicate this thesis to my parents, my family. With their support, I could concentrate on my study here in Stuttgart.

Abstract

During the 88th and 96th Conferences of the Working Committee of the Surveying Authorities of the States of the Federal Republic of Germany (AdV) in 1991 and 1995 the resolution was passed to introduce the European Terrestrial Reference System 1989 (ETRS89) with the universal transversals Mercator projection (UTM) as map projection of GRS80 Ellipsoid in Germany and Europe uniformly. The task of transferring the existing two-dimensional Gauss-Krüger coordinates in German geodetic reference system (DHDN) into UTM coordinates in the ETRS89 datum arises from these resolutions.

For the concrete “Introduction of ETRS89 into Baden-Württemberg” the transformation with the two models of the 7-Parameter Helmert transformation and the 6-Parameter Helmert transformation using the 131 collocated points (131 BWREF points in Baden-Württemberg) are firstly tested and discussed. Because of the special characteristic of the main triangle net of Baden-Württemberg (countrywide variable net scales, inhomogeneous point accuracies and transformation residual in the decimeter level) an alternative transformation procedure with the Total Least-Squares method is also applied in the estimation of the 7-Parameter Helmert transformation and 6-Parameter Affine transformation based on the 131 collocated points. After the review of basis mathematic background of the TLS method, these methods are complemented with MATLAB. Furthermore, 10 selected points are as test points to study the influence on those points after using TLS transformation parameters. The results are analyzed and compared with these results from the conventional LS method, and the advantages and shortcomings of this TLS method are discussed.

Keywords: Similarity transformation, affine transformation, Helmert transformation, Total Least-Squares method, DHDN, ETRS89, SVD (Singular value decomposition).

Contents

1. Introduction.....	1
2. Transformation models and appropriate analysis with collocated points in Baden-Württemberg.....	2
2.1. Six parameter affine transformation models (2D).....	2
2.2. Seven parameter Helmert transformation models (3D)	3
2.3. Data preparation and the solution of transformation parameters	5
3. Transformation procedures based on Total Least Squares Estimation	7
3.1. Total Least Squares (TLS): Introduction.....	7
3.2. Solving the Total Least-Squares problem with singular value decomposition (SVD)	11
3.2.1. Notation and preliminaries.....	11
3.2.2. Definition of the total least squares problem.....	13
3.2.3. The singular value decomposition	14
3.2.4. Basic Solution of TLS problem	17
3.3. The Euler-Lagrange approach	20
3.4. Properties of the Total Least Squares problem.....	23
4. Case studies: Applications of the TLS method in Baden- Württemberg.....	25
4.1. Six parameter Affine transformation models (2D).....	25
4.2. Seven parameter Helmert transformation model (3D).....	30
5. Conclusions and further studies	37
REFERENCES.....	38
Appendix I	40
Appendix II.....	45

1. Introduction

With the development of space geodesy, the modern geodetic reference frames are substantially produced by different national and international organizations, such as the IERS (International Earth Rotation and Reference Systems Service) etc. This brings the tasks to transform the classical coordinates to the new space geodesy based reference frames (Cai, 2006). The IAG sub-commission for the European Reference Frame (EUREF), following its Resolution adopted in Florence meeting in 1990, recommends that the terrestrial reference system to be adopted by EUREF will be coincident with ITRS at the epoch 1989.0 and fixed to the stable part of the Eurasian Plate. It was named as European Terrestrial Reference System 89 (ETRS89).

And ETRS89 is realized through several ways, and specifically:

- 1) Using ITRS realizations: For each frame, labeled ITRF_y a corresponding frame in ETRS89 can be computed and labeled ETRF_y. The following ETRF solutions are presently available: ETRF89; ETRF90; ETRF91; ETRF92; ETRF93; ETRF94; ETRF96; ETRF97; ETRF2000.
- 2) Positioning with GPS measurements of a campaign or permanent stations: using recent ITRF_y station coordinates and IGS precise ephemerides.

In Germany, the Federal Agency for Cartography and Geodesy (BKG) is involved in the maintenance of this system through the German Reference Network (DREF91). It has been carried out in April 1991 by GPS campaign, which comprises 109 sites. The standard deviation of the final set of three-dimensional coordinates is 1 to 2 cm horizontal, 2 cm vertical (Cai, 2000).

The Primary Triangulation Net of the West Germany (DHDN) was emerged between 1870 and 1950 from the combination of several individual networks. The Bessel Ellipsoid (1841) served as a reference surface. The astronomic azimuth of the triangle side connecting Rauenberg and Berlin-Marienkirche oriented the origin points in Rauenberg near Berlin and the networks. In Baden-Württemberg, there are more than 10000 two-dimensional Gauss-Krüger coordinates in DHDN. They should be transformed into UTM coordinates in the ETRS89 datum with GRS80 Ellipsoid. The transformation possibilities with different transformation models and with different estimation methods have been widely discussed and developed. In most case the estimation of the transformation-parameters are performed with the Least Squares method. However, in some cases, design matrix elements are also stochastic. Usually, this is ignored in classical least squares and this ignorance propagates as an uncertainty in the solution results (Cai, 2000 and 2006).

Total Least Squares method (TLS) is a new method of parameter estimation in linear models that considers errors in some or all variables. By using TLS, coordinates of points in two coordinate systems are considered with their error component, which lead to the errors in the design matrix elements of the transformation equations. By this means, residuals of transformation can be reduced. Especially, as the old local

coordinates are with lower accuracy or with some distortions.

In this study the two transformations models (six-parameter affine transformation model and seven-parameter Helmert transformation mode) are firstly reviewed. They will be analyzed using the 131 collocated points (131 BWREF points in Baden-Württemberg), in which the height reference surface in Germany will be shortly reviewed. Then the mathematic foundations of TLS method are reviewed. Two different solutions (Van Huffel, 1991 and Schaffrin, 2005) are introduced. Then the transformation procedures with Least Squares (LS) and Total Least Squares (TLS) estimators are performed with the same data sets (131 collocated points). At the same time in order to study the influence of TLS method on the new points, 10 points are selected as test points. Finally, some conclusion and further studies will be given based on the comparison and analysis.

2. Transformation models and appropriate analysis with collocated points in Baden-Württemberg

2.1. Six parameter affine transformation models (2D)

In the case of map coordinates, which result from the projection of the reference ellipsoid into plane, a two-dimensional model is more useful. For example, when between the respective reference systems (DHDN, Bessel and ETRS89, GRS80) no direct mathematical relationship exists. Two-dimensional transformation models are used. As a result, the Gauss-Krüger coordinates of the net points in DHDN can be transformed only over collocated points into UTM coordinates related to ETR89. For the two dimensional transformation models, there are three, four, five or six transformation parameters, whose number depends on the respective requirements. Because the models with three or five parameter can make for some problems, e.g. non-linear equations problem, so they will not be considered here. In most applications of the plane-transformation, the 6-parameter affine transformation model is used and is recommended by the Surveying Authorities of the States of the Federal Republic of Germany (AdV). Therefore, the 6-parameter affine model will be reviewed and applied in estimating the parameters of the plane transformation parameters based on 131 collocated points in Baden Württemberg (Cai, 2006).

With the planar affine transformation, where six parameters are to be determined, both coordinate directions are rotated with two different angles α and β . So that not only the distances and the angles are distorted, but usually also the original orthogonality of the axes of coordinates is lost. An affine transformation preserves collinearity and ratios of distances. While an affine transformation preserves proportions on lines, it does not necessarily preserve angles or lengths (Cai, 2006).

The 6-parameter affine transformation model between any two plane coordinates systems, e.g. from Gauss-Krüger coordinate (H, R) in DHDN (G) directly to the UTM-Coordinate (N, E) in ETRS89 can be written as

$$\begin{bmatrix} N \\ E \end{bmatrix} = \begin{bmatrix} \lambda_H \cos \alpha & -\lambda_R \sin \beta \\ \lambda_H \sin \alpha & \lambda_R \cos \beta \end{bmatrix} \begin{bmatrix} H \\ R \end{bmatrix} + \begin{bmatrix} t_N \\ t_E \end{bmatrix}. \quad (2.1.1)$$

Where:

t_N, t_E	Translation parameters
α, β	Rotation parameters
λ_H, λ_R	Scale corrections

2.2. Seven parameter Helmert transformation models (3D)

Several mathematical models have been developed which describe the functional relationship between pairs of three-dimensional coordinates. Three mathematical models, namely Bursa-Wolf (Bursa, 1962, Wolf, 1963), which is also usually noted as the seven parameter Helmert model, Molodensky-Badekas (Molodensky et.al., 1960; Badekas, 1969) and Veis (Veis, 1960) are noted as standard models due to their extensive use around the world over many years. The models differ from each other in several ways including a prior condition, the type of coordinate used and the interpretation of results.

The three-dimensional conformal coordinate transformation is also known as the seven-parameter similarity transformation. It transfers points from one three-dimensional coordinate system to another (Wolf and Ghilani 1997). The parameters are three translations (shift of origin), three rotations, and one parameter modeling a possible scale difference. The transformation preserves the shape of objects (Acar, 2006).

- 7 parameter Helmert transformation model (Bursa-Wolf model)

$$\begin{bmatrix} X_G \\ Y_G \\ Z_G \end{bmatrix} = (1 + d\lambda) \begin{bmatrix} 1 & \gamma & -\beta \\ -\gamma & 1 & \alpha \\ \beta & -\alpha & 1 \end{bmatrix} \begin{bmatrix} X_L \\ Y_L \\ Z_L \end{bmatrix} + \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix} \quad (2.2.1)$$

- Molodensky-Badekas model

$$\begin{bmatrix} X_G \\ Y_G \\ Z_G \end{bmatrix} = \begin{bmatrix} X_L \\ Y_L \\ Z_L \end{bmatrix} + \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix} + \begin{bmatrix} 0 & \omega & -\psi \\ -\omega & 0 & \varepsilon \\ \psi & -\varepsilon & 0 \end{bmatrix} \begin{bmatrix} X_L - X_{L0} \\ Y_L - Y_{L0} \\ Z_L - Z_{L0} \end{bmatrix} + d\lambda \begin{bmatrix} X_L - X_{L0} \\ Y_L - Y_{L0} \\ Z_L - Z_{L0} \end{bmatrix} \quad (2.2.2)$$

- Veis model

$$\begin{pmatrix} X_G \\ Y_G \\ Z_G \end{pmatrix} = \begin{pmatrix} T_X \\ T_Y \\ T_Z \end{pmatrix} + \begin{pmatrix} X_L \\ Y_L \\ Z_L \end{pmatrix} + \begin{pmatrix} X_L - X_{L0} \\ Y_L - Y_{L0} \\ Z_L - Z_{L0} \end{pmatrix} d\lambda + \quad (2.2.3)$$

$$\begin{pmatrix} 0 & Z_{L0} - Z_L & Y_L - Y_{L0} \\ Z_L - Z_{L0} & 0 & X_L - X_{L0} \\ Y_{L0} - Y_L & X_L - X_{L0} & 0 \end{pmatrix} \begin{pmatrix} -\sin B_{L0} \cos L_{L0} & -\sin L_{L0} & \cos B_{L0} \cos L_{L0} \\ -\sin B_{L0} \sin L_{L0} & \cos L_{L0} & \cos B_{L0} \sin L_{L0} \\ \cos B_{L0} & 0 & \sin B_{L0} \end{pmatrix} \begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix}$$

The similarity of the transformation is particularly important since the conformal characteristics of the coordinates after the transformation are maintained. They are applied particular for the discrepancies between a local (e.g. DHDH related Bessel ellipsoid) and a global reference system (e.g. ETRS89 related GRS80 Ellipsoid) which are due to the differences in the geodetic datum. However, the three seven parameter models are expressed in different forms with different origin and parameters, their transformation results are completely equivalent. So 7-Parameter Helmert model is used most commonly (Cai, 2006).

The further reasons for the choice of 7-Parameter Helmert transformation model are:

- 1) It is the only known method which allows a direct interpretation of the origin shifts;
- 2) The rotations around the “Earth-Centered, Earth-Fixed (ECEF)” Cartesian axes can have physical interpretations in global reference frames (Cai, 2006).

It performs a conformal transformation, where the ratios of distances and the angles preserve invariantly. A “local” non-geocentric X_L, Y_L, Z_L -system can be transformed into a “global” geocentric X_G, Y_G, Z_G -system with the help of a 7-Parameter Helmert transformation model.

$$\begin{aligned} \begin{bmatrix} X_G \\ Y_G \\ Z_G \end{bmatrix} &= \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix} + (1 + d\lambda) \begin{bmatrix} 1 & \gamma & -\beta \\ -\gamma & 1 & \alpha \\ \beta & -\alpha & 1 \end{bmatrix} \begin{bmatrix} X_L \\ Y_L \\ Z_L \end{bmatrix} \\ &\cong \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix} + \begin{bmatrix} d\lambda & \gamma & -\beta \\ -\gamma & d\lambda & \alpha \\ \beta & -\alpha & d\lambda \end{bmatrix} \begin{bmatrix} X_L \\ Y_L \\ Z_L \end{bmatrix} + \begin{bmatrix} X_L \\ Y_L \\ Z_L \end{bmatrix} \end{aligned} \quad (2.2.4)$$

Where:

- T_X, T_Y, T_Z : Translation parameters
- α, β, γ : Differential rotation parameters
- $d\lambda$: Scale correction

Note that these approximations are prepared and these angles are small enough, as well as the scale correction, i.e., the term

$$d\lambda \begin{bmatrix} 0 & \gamma & \beta \\ -\gamma & 0 & \alpha \\ -\beta & -\alpha & 0 \end{bmatrix} \quad (2.2.5)$$

is so small that it can be eliminated. In general, these parameters are unknown. Therefore, we must determine them by analyzing of those collocated points, and to solving this problem at least three collocated points in both systems are required.

2.3. Data preparation and the solution of transformation parameters

In preparation of local coordinate of collocated points the Gauss-Krüger coordinates of DHDN will be transformed to Bessel ellipsoidal coordinates latitude (B_L) and longitude (L_L) through inverse conformal mapping formulas and the conversion of ellipsoidal coordinates (B_L, L_L, H_L) to geodetic Cartesian coordinates (X_L, Y_L, Z_L) and the reverse conversion are accomplished through the general formula. For the global coordinates can be also converted with the same algorithms on the GRS80 ellipsoid. Then we can construct the quasi-observations with the 131 collocated points (131 BWREF points in Baden-Württemberg) and perform the estimation of the transformation parameters of the 7-parameter Helmert transformation and the 6-parameter affine transformation. The transformation parameter solutions using above two models are listed in table 1 and the residuals are illustrated in figure 1 and figure 2.

From the distribution of horizontal residuals shown in figure 1 by 3-D 7-parameter Helmert transformation and figure 2 by 2-D 6-Parameter affine transformation we can find that there are two rotational trends of the direction of the horizontal residual vectors, which are clockwise in the northern part and counter-clockwise in the southern part. The cause for it lies despite the homogeneity of the network structure in the remaining distortions of the DHDN, which is highly correlated over larger areas. The horizontal position residuals of these sites bordering the boundary of the state of Baden-Württemberg are larger than these inner sites. The largest residual occur in site 130 by 0.43 m. similar results can also be found in figure 2 by 2-D 6-parameter affine transformation (Cai, 2006).

After the transformation of DHDN/Gauss Krüger coordinates into ETRS89/UTM coordinates, the inherent traditional network distortions of the DHDN in Baden-Württemberg (BWREF) can be visually shown through the residuals in 131 collocated points. Since the special characteristic of the main triangulation network in Baden-Württemberg (statewide variable net scales, inhomogeneous point accuracies and network distortions in the decimeter level) a statewide similarity or affine transformation parameter set cannot satisfy the requirement of the transformation accuracy in Baden-Württemberg (Cai, 2006).

Table 1 Transformation parameters and their standard deviation

	6-parameter affine transformation GK (DHDN) – UTM (ETRS89)								
	$t_N (m)$	$t_E (m)$	$\alpha (")$	$\beta (")$	$d\lambda_H (\times 10^{-4})$	$d\lambda_R (\times 10^{-4})$	$RMS^* (m)$	$\hat{\sigma} (m)$	
131 BWREF points	437.1946	119.7567	0.1654	-0.1965	-3.9968	-3.9884	0.1187	0.1199	
	7-parameter Helmert transformation GK (DHDN) – UTM (ETRS89)								
	$T_x (m)$	$T_y (m)$	$T_z (m)$	$\alpha (")$	$\beta (")$	$\gamma (")$	$d\lambda (\times 10^{-6})$	$RMS^* (m)$	$\hat{\sigma} (m)$
131 BWREF points	582.9017	112.1681	405.6031	-2.2550	-0.3350	2.0684	9.1172	0.1241	0.1026

(*RMS: quadratic means of the horizontal residuals)

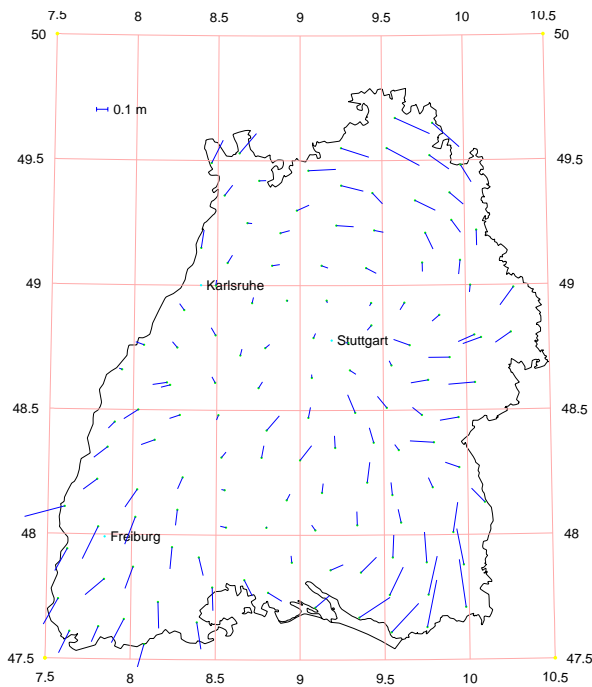


Figure 1 Horizontal residuals after 7-parameter Helmert transformation in Baden-Württemberg network

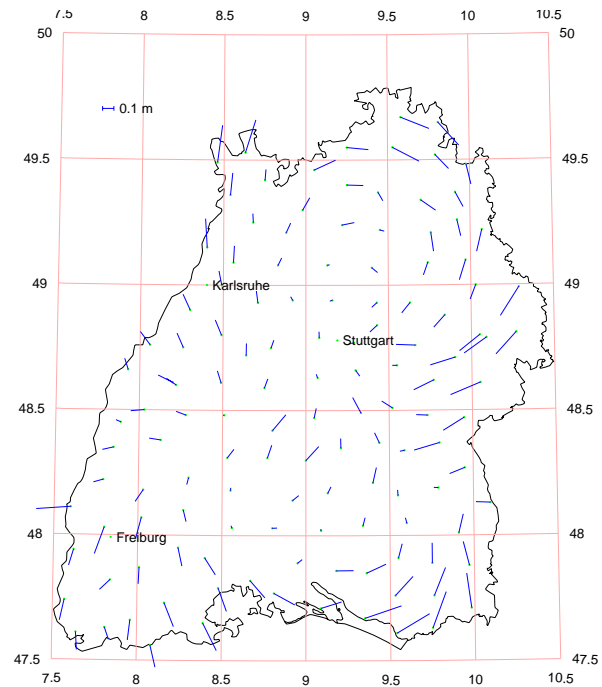


Figure 2 Horizontal residuals after 6-parameter affine transformation in Baden-Württemberg network

3. Transformation procedures based on Total Least Squares Estimation

3.1. Total Least Squares (TLS): Introduction

The total least squares method is one of several linear parameter estimation techniques that have been devised to compensate for data inconsistencies. The basic motivation for total least squares (TLS) is the following: Let a set of multidimensional data points (vectors) be given. How can one obtain a linear model that explains these data? The idea is to modify all data points in such a way that some norm of the modification is minimized subject to the constraint that the modified vectors satisfy a linear relation (Van Huffel, 1991).

The origin of this basic idea can be traced back to the beginning of this century. It was rediscovered many times, often independently, mainly in the statistical and psychometric literature. However, it is only in the last decade that the technique also found wide use in scientific and engineering applications. One of the main reasons for its popularity is the availability of efficient and numerical robust algorithms, in which the singular value decomposition plays a prominent role. Another reason is the fact that TLS is an application-oriented procedure. It is ideally suited for situations in which all data are corrupted by noise, which is usually the case in engineering applications. In this sense, it is a powerful extension of the classical least squares method, which corresponds only to a partial modification of the data (Van Huffel, 1991).

The problem of linear parameter estimation arises in a broad class of scientific disciplines such as signal processing, automatic control, system theory and in general engineering, statistics, physics, economics, biology, geodesy, medicine, etc.. It starts from a model described by a linear equation

$$y = a_1\xi_1 + \dots + a_m\xi_m, \quad (3.1.1)$$

Where a_1, \dots, a_m and y denote the variables and $\xi = [\xi_1, \dots, \xi_m]^T \in \mathfrak{R}^m$ plays role of a parameter vector that characterizes the specific system (\mathfrak{R} denotes the set of real numbers). The basic problem is then to determine an estimate of the true but unknown parameters from certain measurements of the variables. This gives rise to an overdetermined set of n linear equations ($n > m$):

$$\mathbf{y} = \mathbf{A}\xi + \mathbf{e} \quad (3.1.2)$$

Where the i -th row of the design matrix $\mathbf{A} \in \mathfrak{R}^m$ and the vector of observations $\mathbf{y} \in \mathfrak{R}^n$ contain the measurements of the variables a_1, \dots, a_m and y , respectively. In the classical least squares (LS) approach the measurements \mathbf{A} of the variables a_i (the right-hand side of (3.1.2) are assumed to be free of error and hence, all errors are confined to the observation vector \mathbf{y} (the left-hand side of (3.1.2)). However, this

assumption is frequently unrealistic: sampling errors, human errors, modeling errors and instrument errors may imply inaccuracies of the design matrix \mathbf{A} as well. TLS is one method of fitting that is appropriate when there are errors in both the observation vector \mathbf{y} and the design matrix \mathbf{A} . It amounts to fitting a “best” subspace to the measurement data (\mathbf{A}_i^T, y_i) , $i=1, \dots, n$, where \mathbf{A}_i^T is the i -th row of \mathbf{A} . To illustrate the effect of the use of TLS in comparison with LS, we consider here a simple example of parameter estimation, i.e., only one parameter ($m=1$) is to be estimated. Hence, equation (3.1.1) reduces to the following

$$y = a\xi. \quad (3.1.3)$$

An estimate for parameter ξ is to be found from n measurements of the variables a and y :

$$\begin{aligned} a_i &= a_i^0 + \Delta a_i \\ y_i &= y_i^0 + \Delta y_i, \quad i=1, \dots, n \end{aligned} \quad (3.1.4)$$

By solving the linear system (3.1.2) with $\mathbf{A} = [a_1, \dots, a_n]^T$ and $\mathbf{y} = [y_1, \dots, y_n]^T$.

Δa_i and Δy_i represent the random errors added to the true values a_i^0 and y_i^0 of the variables a and y . If a can be observed exactly, i.e., $\Delta a_i = 0$, errors only occur in the measurements of y contained in the right-hand side vector \mathbf{y} . Hence, the use of LS for solving (3.1.2) is appropriate. This method perturbs the observation vector \mathbf{y} by a minimum amount \mathbf{e} so that $(\mathbf{y} - \mathbf{e})$ can be predicted by $\mathbf{A}\xi$. This is done by minimizing the sum of squared differences

$$\sum_{i=1}^n (y_i - a_i \xi)^2. \quad (3.1.5)$$

The best estimate $\hat{\xi}_y$ of ξ follows then immediately:

$$\hat{\xi}_y = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y} = \frac{\sum_{i=1}^n a_i y_i}{\sum_{i=1}^n a_i^2} \quad (3.1.6)$$

This LS estimation has a nice geometric interpretation as is shown in Figure 3(a)

If y can be measured without error, i.e., $\Delta y_i = 0$, the use of LS is again appropriate.

Indeed we can rewrite as

$$\frac{y}{\xi} = a, \quad (3.1.7)$$

and confine all errors to the measurements of a contained in the right-hand side vector

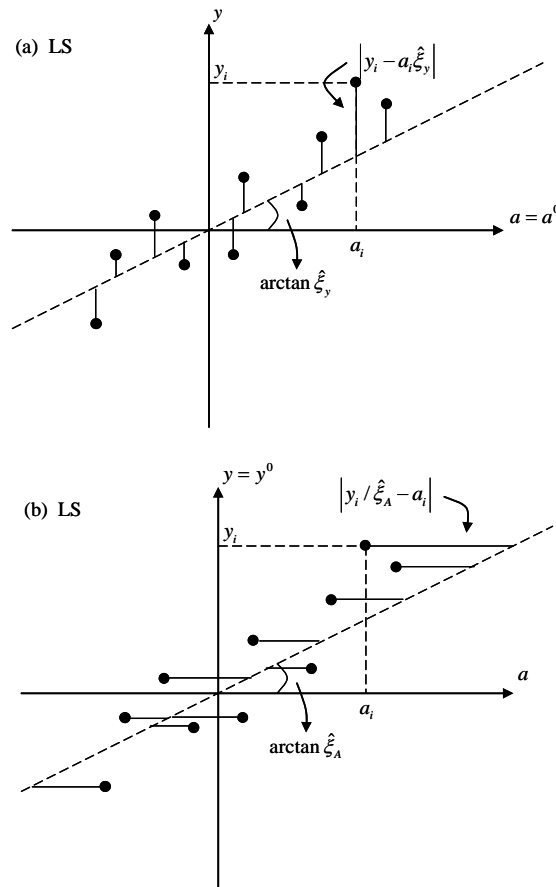
\mathbf{A} of the corresponding set of equations $\mathbf{A} \approx \mathbf{y}\xi^{-1}$. By minimizing the sum of squared differences between the measured values a_i and the predicted values y_i/ξ , the best estimate $\hat{\xi}_A$ of ξ is given by (see Figure 3(b))

$$\hat{\xi}_A = ((\mathbf{y}^T \mathbf{y})^{-1} \mathbf{y}^T \mathbf{A})^{-1} = \frac{\sum_{i=1}^m y_i^2}{\sum_{i=1}^m a_i y_i}. \quad (3.1.8)$$

In many applications, however, both variables are measured with errors, i.e., $\Delta a_i \neq 0$ and $\Delta y_i \neq 0$. If the errors are independently and identically distributed with zero mean and common variance σ_v^2 , the best estimate $\hat{\xi}$ is obtained by minimizing the sum of squared distances of the observed points from the fitted line,

$$\text{i.e., } \sum_{i=1}^n (y_i - a_i \xi)^2 / (1 + \xi^2). \quad (3.1.9)$$

This is in fact the solution $\tilde{\xi}_{TLS}$ we obtain by solving (3.1.2) with the TLS method for $m=1$. Figure 3(c) illustrates the estimation. The deviations are orthogonal to the fitted line: it is the sum of squares of their lengths that is minimized. Therefore, this estimation procedure is sometimes also known as orthogonal regression (Van Huffel, 1991).



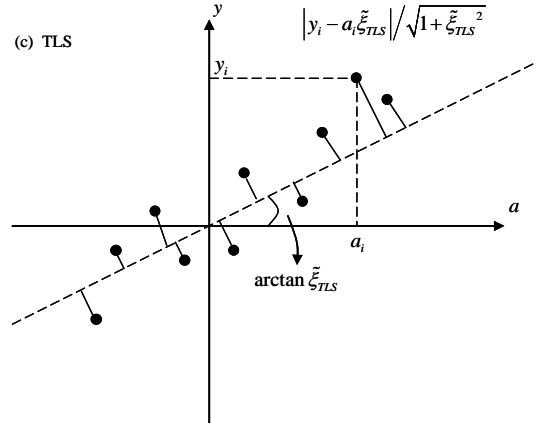


Figure 3 Geometric interpretation of one-parameter estimation $y = a\xi$ with errors in (a) the measurements y_i of y only (LS solution), (b) the measurements a_i of a Only (LS solution), and (c) both the measurements a_i of a and y_i of y (TLS solution) (Van Huffel 1991)

Although the name “total least squares” appeared only recently in the literature, this method of fitting is certainly not new and has a long history in the statistical literature, where the method is known as orthogonal regression or errors-in-variables regression. Indeed, the univariate line fitting problem ($m=1$) was already discussed in the previous century. Some well-known contributors are Adcock (1878), Pearson (1901), Koopmans (1937), Madansky (1959) and York (1966). The method of orthogonal regression has been rediscovered many times, often independently. About twenty years ago, the technique was extended to multivariate problems ($m>1$) and later to multidimensional problems which deal with more than one observation vector (Van Huffel, 1991).

More recently, the TLS approach to fitting has also stimulated interests outside statistics. In the field of numerical analysis, Golub (1973) and Van Loan (1980) first studied this problem. Their analysis, as well as their algorithm, is strongly based on the singular value decomposition (SVD). Geometrical insight into the properties of the SVD brought us independently to the same concept. The algorithm of Golub (1973) and Van Loan (1980) was generalized to all cases in which their algorithm fails to produce a solution, described the properties of these so-called nongeneric TLS problems and proved that the proposed generalization still satisfies the TLS criteria if additional constraints are imposed on the solution space. This seemingly different linear algebraic approach is actually equivalent to the method of multivariate errors-in-variables regression analysis, studied by Gleser (1981). Gleser’s method is based on an eigenvalue-eigenvector analysis, while the TLS method uses the SVD, which is numerically more robust in the sense of algorithmic implementation. Furthermore, the TLS algorithm computes the minimum norm solution whenever the TLS problem lacks a unique minimizer. Gleser (1981) does not consider these extensions (Van Huffel, 1991).

In addition, in the field of experimental modal analysis, the TLS technique was studied recently. Finally, in the field of system identification, Levin (1964) first studied the

same problem. His method, called the eigenvector method or Koopmans-Levin method, computes the same estimate as the TLS algorithm whenever the TLS problem has a unique solution.

In the field of Geodesy, the TLS technique, was also studied recently. Grafarend (2006) studied the principle of TLS and analyzed its availability in geodesy. Beside Schaffrin (2003) invented his method to solve TLS problem in coordinate transformation, Acar (2006) und Oezluedemir (2006) from turkey worked out this problem by using SVD method.

Remember that TLS is only one possible fitting technique for estimating the parameters of a linear multivariate problem. It gives the “best” estimates (in a statistical sense) when all variables are subject to independently and identically distributed errors with zero mean and common covariance matrix equal to the identity matrix, up to a scaling factor. Several other and more general approaches to this problem have led to as many other fitting techniques for the linear as well as for the nonlinear case.

3.2. Solving the Total Least-Squares problem with singular value decomposition (SVD)

3.2.1. Notation and preliminaries

Before starting, we introduce some notation, list the assumptions, and define the elementary statistical concepts used throughout this thesis.

- The superscript T denotes the transpose of a vector or matrix.
- $R(\mathbf{S})$ denotes the column space of matrix \mathbf{S} , and $N(\mathbf{S})$ denotes the null space or kernel of \mathbf{S} .
- Special notation \mathbf{B} is convenient for diagonal matrices. If \mathbf{B} is an $n \times m$ matrix and we write

$$\mathbf{B} = \text{diag}(b_1, \dots, b_p), \quad p = \min\{n, m\}.$$

Then $\mathbf{B}_{ij} = 0$ whenever $i \neq j$ and $\mathbf{B}_{ii} = b_i$ for $i = 1, \dots, p$.

- The $n \times n$ identity matrix is denoted by \mathbf{I}_n or, more simply, by \mathbf{I} .
- The set of n linear equations in m unknowns ξ is represented in matrix form by

$$\mathbf{y} \approx \mathbf{A}\xi. \tag{3.2.1}$$

\mathbf{A} is the $n \times m$ design matrix and \mathbf{y} is the n -dimensional vector of observations. Unless stated otherwise, we assume that the set of equations $\mathbf{y} \approx \mathbf{A}\xi$ is overdetermined, i.e., $n > m$, and that all preprocessing measures on the data (such as scaling, whitening, centering, standardizing, etc.) have been performed in advance.

- The Frobenius norm of an $n \times m$ matrix \mathbf{M} is defined by (“tr” denotes trace)

$$\|\mathbf{M}\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^m m_{ij}^2} = \sqrt{\text{tr}(\mathbf{M}^T \mathbf{M})}.$$

The 2-norm or Euclidean norm of an m -dimensional vector \mathbf{b} is defined by

$$\|\mathbf{b}\|_2 = \sqrt{\sum_{i=1}^m \mathbf{b}_i^2}.$$

The 2-norm of an $m \times n$ matrix \mathbf{M} is defined by

$$\|\mathbf{M}\|_2 = \sup_{\mathbf{b} \neq 0} \frac{\|\mathbf{M}\mathbf{b}\|_2}{\|\mathbf{b}\|_2}$$

and equals the largest singular value of \mathbf{M} .

- Denote the singular value decomposition (SVD) of the $n \times m$ design matrix \mathbf{A} , $n > m$, in (3.2.1) by

$$\mathbf{A} = \mathbf{U}' \boldsymbol{\Sigma}' \mathbf{V}'^T \tag{3.2.2}$$

with

$$\mathbf{U}' = [\mathbf{U}'_1; \mathbf{U}'_2], \quad \mathbf{U}'_1 = [u'_1, \dots, u'_m], \quad \mathbf{U}'_2 = [u'_{m+1}, \dots, u'_n], \quad u'_i \in \mathfrak{R}^n, \quad \mathbf{U}'^T \mathbf{U}' = \mathbf{I}_n,$$

$$\mathbf{V}' = [v'_1, \dots, v'_m], \quad v'_i \in \mathfrak{R}^m, \quad \mathbf{V}'^T \mathbf{V}' = \mathbf{I}_m,$$

$$\boldsymbol{\Sigma}' = \text{diag}(\sigma'_1, \dots, \sigma'_m) \in \mathfrak{R}^{n \times m}, \quad \sigma'_1 \geq \dots \geq \sigma'_m \geq 0,$$

and denote the SVD of the $n \times (m+1)$ matrix $[\mathbf{A}; \mathbf{y}]$, $n > m$, in (4.2.1) by

$$[\mathbf{A}; \mathbf{y}] = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^T$$

with

$$\mathbf{U} = [\mathbf{U}_1; \mathbf{U}_2], \quad \mathbf{U}_1 = [u_1, \dots, u_m], \quad \mathbf{U}_2 = [u_{m+1}, \dots, u_n], \quad u_i \in \mathfrak{R}^n, \quad \mathbf{U}^T \mathbf{U} = \mathbf{I}_n,$$

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \begin{matrix} m \\ 1 \end{matrix} = [v_1, \dots, v_{m+1}], \quad v_i \in \mathfrak{R}^{m+1}, \quad \mathbf{V}^T \mathbf{V} = \mathbf{I}_{m+1},$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & 0 \\ 0 & \boldsymbol{\Sigma}_2 \end{bmatrix} = \text{diag}(\sigma_1, \dots, \sigma_{m+1}) \in \mathfrak{R}^{n \times (m+1)},$$

Where:

$$\boldsymbol{\Sigma}_1 = \text{Diag}(\sigma_1, \dots, \sigma_m) \in \mathfrak{R}^{m \times m}, \quad \boldsymbol{\Sigma}_2 = \sigma_{m+1} \in \mathfrak{R}$$

and $\sigma_1 \geq \dots \geq \sigma_{m+1} \geq 0$.

For convenience of notation, we define $\sigma_i = 0$ if $n < i \leq m+1$. The σ'_i and σ_i

are the singular value of \mathbf{A} and $[\mathbf{A}; \mathbf{y}]$, respectively. The vectors u_i' and u_i are the i th left singular vector, and the vectors v_i' and v_i are the i th right singular vector of \mathbf{A} and $[\mathbf{A}; \mathbf{y}]$, respectively.

3.2.2. Definition of the total least squares problem

In the following, we formulate the main principle of the TLS problem. A good way to introduce and motivate the method is to recast the ordinary Least Squares (LS) problem.

Definition 1 (The least squares problem). Given an overdetermined set of n linear equations $\mathbf{y} \approx \mathbf{A}\xi$ in m unknowns ξ . The least squares problem is the following (Van Huffel, 1991):

$$\min_{\hat{\mathbf{y}} \in \mathbb{R}^n} \|\mathbf{y} - \hat{\mathbf{y}}\|_2 \quad \text{subject to } \hat{\mathbf{y}} \in \mathbf{R}(\mathbf{A}).$$

Once a minimizing $\hat{\mathbf{y}}$ is found, then any ξ satisfying:

$$\hat{\mathbf{y}} = \mathbf{A}\xi \tag{3.2.3}$$

is called a LS solution and $\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}}$ the corresponding LS correction. Above equations are satisfied if $\hat{\mathbf{y}}$ is the orthogonal projection of \mathbf{y} onto $\mathbf{R}(\mathbf{A})$. Thus, the LS problem amounts to perturbing the observation vector \mathbf{y} by a minimum amount $\hat{\mathbf{e}}$ so that $\hat{\mathbf{y}} = \mathbf{y} - \hat{\mathbf{e}}$ can be ‘predicted’ by the columns of \mathbf{A} .

The underlying assumption here is that errors only occur in the vector \mathbf{y} and that the matrix \mathbf{A} is exactly known. Often this assumption is not realistic because sampling or modeling or measurement errors also affect the matrix \mathbf{A} .

One way to take errors in \mathbf{A} into account is to introduce perturbations in \mathbf{A} also and to consider the following TLS problem.

Definition 2 (The total least squares problem). Given an overdetermined set of n linear equations $\mathbf{y} \approx \mathbf{A}\xi$ in m unknowns ξ . The total least squares problem is the following (Van Huffel, 1991):

$$\min_{[\hat{\mathbf{A}}; \hat{\mathbf{y}}] \in \mathbb{R}^{n \times (m+1)}} \left\| [\mathbf{A}; \mathbf{y}] - [\hat{\mathbf{A}}; \hat{\mathbf{y}}] \right\|_F \quad \text{subject to } \hat{\mathbf{y}} \in \mathbf{R}(\hat{\mathbf{A}}).$$

Once a minimizing $[\hat{\mathbf{A}}; \hat{\mathbf{y}}]$ is found, then any ξ satisfying:

$$\hat{\mathbf{y}} = \hat{\mathbf{A}}\xi \tag{3.2.4}$$

is called a TLS solution and $[\hat{\mathbf{E}}_A; \hat{\mathbf{e}}] = [\mathbf{A}; \mathbf{y}] - [\hat{\mathbf{A}}; \hat{\mathbf{y}}]$ the corresponding TLS correction. The TLS solution is denoted by $\hat{\xi}$. It is clear that the TLS problem is a generalization of the LS problem as is formulated in Definition 1. See figure 4.

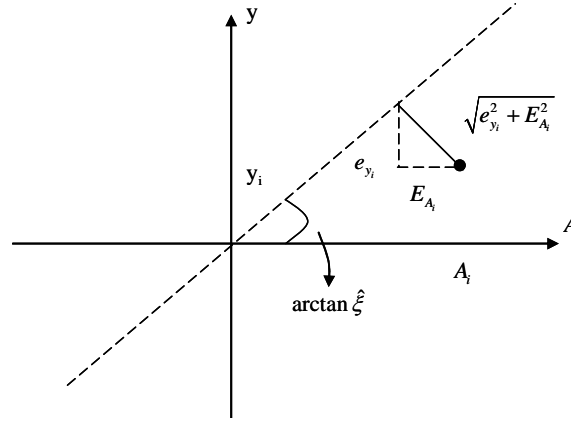


Figure 4 The straight line fit of total least squares ($y = \mathbf{A}\xi$, $\mathbf{y} + \mathbf{e}_y = (\mathbf{A} + \mathbf{E}_A)\xi$)

In most problems, the TLS problem has a unique solution, which can be obtained from a simple scaling of the right singular matrix of $[\mathbf{A}; \mathbf{y}]$ corresponding to its smallest singular value. In this section, we will present the basic algorithm for solving the TLS problem. We will start with a brief review of the singular value decomposition (SVD) and some of its properties (Van Huffel, 1991).

3.2.3. The singular value decomposition

The singular value decomposition (SVD) is of great theoretical and practical importance for the LS and TLS problems. Not only does it provide elegant geometrical and algebraic insights into many numerical linear algebra problems, but also at the same time, a numerically reliable algorithm can be devised.

Theorem 1 (Singular value decomposition (SVD)). If $\mathbf{C} \in \mathfrak{R}^{n \times m}$ then there exist orthonormal matrices $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_n] \in \mathfrak{R}^{n \times n}$ and $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_m] \in \mathfrak{R}^{m \times m}$ (Huffel, 1991) such that

$$\mathbf{U}^T \mathbf{C} \mathbf{V} = \mathbf{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_p), \sigma_1 \geq \dots \geq \sigma_p \geq 0 \text{ and } p = \min\{n, m\}. \quad (3.2.5)$$

The σ_i are the singular values of \mathbf{C} and they are collectively known as the singular value spectrum. The vectors \mathbf{u}_i and \mathbf{v}_i are the i -th left singular vector and the i -th right singular vector, respectively. The triplet $(\mathbf{u}_i, \sigma_i, \mathbf{v}_i)$ is called a singular triplet. It is easy to verify by comparing columns in the equations $\mathbf{C} \mathbf{V} = \mathbf{U} \mathbf{\Sigma}$ and $\mathbf{C}^T \mathbf{U} = \mathbf{\Sigma}^T \mathbf{V}$ that

$$\mathbf{C} \mathbf{v}_i = \sigma_i \mathbf{u}_i \text{ and } \mathbf{C}^T \mathbf{u}_i = \sigma_i \mathbf{v}_i \quad i = 1, \dots, p. \quad (3.2.6)$$

The SVD reveals many interesting structures of a matrix. If the SVD of \mathbf{C} is given by Theorem 1, and we define r by

$$\sigma_1 \geq \dots \geq \sigma_r > \sigma_{r+1} = \dots = \sigma_p = 0$$

the number of positive singular values, then

$$\begin{aligned}
\text{rank}(\mathbf{C}) &= r, \\
\mathbf{R}(\mathbf{C}) &= \mathbf{R}([\mathbf{u}_1, \dots, \mathbf{u}_r]), \\
\mathbf{N}(\mathbf{C}) &= \mathbf{R}([\mathbf{v}_{r+1}, \dots, \mathbf{v}_m]), \\
\mathbf{R}_r(\mathbf{C}) &= \mathbf{R}(\mathbf{C}^T) = \mathbf{R}([\mathbf{v}_1, \dots, \mathbf{v}_r]), \\
\mathbf{N}_r(\mathbf{C}) &= \mathbf{N}(\mathbf{C}^T) = \mathbf{R}([\mathbf{u}_{r+1}, \dots, \mathbf{u}_n]).
\end{aligned} \tag{3.2.7}$$

Moreover, if $\mathbf{U}_r = [\mathbf{u}_1, \dots, \mathbf{u}_r]$, $\mathbf{\Sigma}_r = \text{diag}(\sigma_1, \dots, \sigma_r)$, and $\mathbf{V}_r = [\mathbf{v}_1, \dots, \mathbf{v}_r]$, then we have the SVD expansion

$$\mathbf{C} = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^T = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T. \tag{3.2.8}$$

Above equation, which is also called the dyadic decomposition, decomposes the matrix \mathbf{C} of rank r in a sum of r matrices of rank one. Also, the 2-norm and the Frobenius norm are neatly characterized in terms of the SVD:

$$\|\mathbf{C}\|_F^2 = \sum_{i=1}^n \sum_{j=1}^m c_{ij}^2 = \sigma_1^2 + \dots + \sigma_p^2, \quad p = \min\{n, m\},$$

$$\|\mathbf{C}\|_2 = \sup_{\mathbf{y} \neq 0} \frac{\|\mathbf{C}\mathbf{y}\|_2}{\|\mathbf{y}\|_2} = \sigma_1.$$

From (3.2.6) it follows that

$$\mathbf{C}^T \mathbf{C} = \mathbf{V} \mathbf{\Sigma}^T \mathbf{\Sigma} \mathbf{V}^T \quad \text{and} \quad \mathbf{C} \mathbf{C}^T = \mathbf{U} \mathbf{\Sigma} \mathbf{\Sigma}^T \mathbf{U}^T, \tag{3.2.9}$$

Thus $\sigma_i^2, i=1, \dots, p$, are eigenvalues of the symmetric and nonnegative definite matrices $\mathbf{C}^T \mathbf{C}$ and $\mathbf{C} \mathbf{C}^T$, and v_i and u_i are the corresponding eigenvectors. Hence, in principle, the SVD can be reduced to the eigenvalue problem for symmetric matrices (Van Huffel, 1991).

The SVD plays an important role in a number of matrix approximation problems. For our purpose the following is the most important, where we consider the approximation of one matrix by another of lower rank.

Subspace problems are characterized by the following structure: A matrix \mathbf{A} is formed from an overdetermined set of measurements. If the measurements are error free, the constraints given by the specific problem are reflected in the rank deficiency of this matrix, having a rank k smaller than maximum rank r . But due to measurement errors, the measurement vectors will in general occupy a higher dimensional linear manifold. The task to be solved is the estimation of the underlying subspace. Subspace problems inherently lead to eigensystem problems; the solutions are directions or sets of directions in the given vector space, not single points. By this characteristic, they stand in sharp contrast to ordinary least squares problems, which are in general linear problems. One of several possibilities for identifying the

directions (subspaces) where the true unperturbed (but unknown) points most probably are located is to employ a SVD (Mühlich and Mester, 1999). This subspace identification is equivalent to a rank reduction of the given matrix \mathbf{A} . The key to correctly performing this rank reduction is given by the Eckart-Young-Mirsky theorem (1936):

Theorem 2 (Eckart-Young-Mirsky matrix approximation theorem). Let the SVD of $\mathbf{C} \in \mathfrak{R}^{n \times m}$ be given by $\mathbf{C} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ with $r = \text{Rank}(\mathbf{C})$. If $k < r$ and

$\mathbf{C}_k = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T$, then $\min_{\text{rank}(\mathbf{D})=k} \|\mathbf{C} - \mathbf{D}\|_2 = \|\mathbf{C} - \mathbf{C}_k\|_2 = \sigma_{k+1}$, And

$$\min_{\text{rank}(\mathbf{D})=k} \|\mathbf{C} - \mathbf{D}\|_F = \|\mathbf{C} - \mathbf{C}_k\|_F = \sqrt{\sum_{i=k+1}^p \sigma_i^2}, \quad p = \min\{n, m\}. \quad (3.2.10)$$

Eckart and Young originally proved the theorem for the Frobenius norm in 1936. In 1960, Mirsky proved the theorem for the 2-norm. Therefore, Theorem 2 is called the Eckart-Young-Mirsky Theorem (Van Huffel, 1991).

Example: Here is a SVD for \mathbf{C}

$$\mathbf{C} = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3] \begin{bmatrix} 40 & 0 & 0 \\ 0 & 40 & 0 \\ 0 & 0 & 0.01 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \mathbf{v}_3^T \end{bmatrix}$$

The three non-zero singular values tell you that the matrix has rank three. However, the value 0.01 is so small that \mathbf{C} is nearly a rank two matrix.

In fact, the matrix \mathbf{D} was created by setting that last singular value to zero.

$$\mathbf{D} = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3] \begin{bmatrix} 40 & 0 & 0 \\ 0 & 40 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \mathbf{v}_3^T \end{bmatrix}$$

Now the rank one decomposition of \mathbf{C} is

$$\mathbf{C} = 40 \cdot \mathbf{u}_1 \mathbf{v}_1^T + 40 \cdot \mathbf{u}_2 \mathbf{v}_2^T + 0.01 \cdot \mathbf{u}_3 \mathbf{v}_3^T.$$

and the rank one decomposition of \mathbf{D} is

$$\mathbf{D} = 40 \cdot \mathbf{u}_1 \mathbf{v}_1^T + 40 \cdot \mathbf{u}_2 \mathbf{v}_2^T + 0 \cdot \mathbf{u}_3 \mathbf{v}_3^T.$$

So $\mathbf{C} - \mathbf{D} = 0.01 \cdot \mathbf{u}_3 \mathbf{v}_3^T$ and $\|\mathbf{C} - \mathbf{D}\|_F = 0.01$

So if \mathbf{C} has a small singular value, and then you can get a lower rank matrix \mathbf{D}

close to \mathbf{C} by setting the small singular value to zero.

The SVD is a powerful computational tool for solving LS problems. The reason is that the orthogonal matrices that transform \mathbf{C} to diagonal form (3.2.6) do not change the 2-norm of vectors. We have the following fundamental result.

Theorem 3 (Minimum norm LS solution of $\mathbf{y} \approx \mathbf{A}\boldsymbol{\xi}$). Let the SVD of $\mathbf{A} \in \mathfrak{R}^{n \times m}$ be given by (4.2.2), i.e., $\mathbf{A} = \sum_{i=1}^m \sigma_i' \mathbf{u}_i' \mathbf{v}_i'^T$, and assume that $\text{rank}(\mathbf{A}) = r$. If $\mathbf{y} \in \mathfrak{R}^n$, then

$$\boldsymbol{\xi}_{LS} = \sum_{i=1}^r \sigma_i'^{-1} \mathbf{v}_i' \mathbf{u}_i'^T \mathbf{y} \quad (3.2.11)$$

minimizes $\|\mathbf{A}\boldsymbol{\xi} - \mathbf{y}\|_2$ and has the smallest 2-norm of all minimizers. Moreover,

$$\|\mathbf{A}\boldsymbol{\xi}_{LS} - \mathbf{y}\|_2^2 = \sum_{i=r+1}^n (\mathbf{u}_i'^T \mathbf{y})^2$$

3.2.4. Basic Solution of TLS problem

We now analyze the TLS problem by making substantial use of the SVD decomposition. We bring $\mathbf{y} \approx \mathbf{A}\boldsymbol{\xi}$ into the form (Van Huffel, 1991)

$$[\mathbf{A} \ \mathbf{y}] \begin{bmatrix} \boldsymbol{\xi} \\ -1 \end{bmatrix} \approx 0. \quad (3.2.12)$$

Let the SVD of $[\mathbf{A}; \mathbf{y}]$ be

$$[\mathbf{A}; \mathbf{y}] = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^T = \sum_{i=0}^{n+1} \sigma_i \mathbf{u}_i \mathbf{v}_i^T, \quad \sigma_1 \geq \dots \geq \sigma_{m+1} \geq 0 \quad (3.2.13)$$

If $\sigma_{m+1} \neq 0$, $[\mathbf{A}; \mathbf{y}]$ is of rank $m+1$ and the space S generated by the rows of $[\mathbf{A}; \mathbf{y}]$ coincides with \mathfrak{R}^{m+1} . There is no nonzero vector in the orthogonal complement of S . In order to obtain a solution, the rank of $[\mathbf{A}; \mathbf{y}]$ must be reduced to m . Using the Eckart-Young-Mirsky theorem, the best rank m TLS approximation $[\hat{\mathbf{A}}; \hat{\mathbf{y}}]$ of $[\mathbf{A}; \mathbf{y}]$, which minimizes the deviations in variance, is given by

$$[\hat{\mathbf{A}}; \hat{\mathbf{y}}] = \mathbf{U} \hat{\boldsymbol{\Sigma}} \mathbf{V}^T, \quad \text{with } \hat{\boldsymbol{\Sigma}} = \text{diag}(\sigma_1, \dots, \sigma_m, 0).$$

The minimal TLS correction is then

$$\sigma_{m+1} = \min_{\text{rank}([\hat{\mathbf{A}}; \hat{\mathbf{y}}])=m} \left\| [\mathbf{A}; \mathbf{y}] - [\hat{\mathbf{A}}; \hat{\mathbf{y}}] \right\|_2$$

and is attained by

$$[\mathbf{A}; \mathbf{y}] - [\hat{\mathbf{A}}; \hat{\mathbf{y}}] = [\hat{\mathbf{E}}_A; \hat{\mathbf{e}}] = \sigma_{m+1} \mathbf{u}_{m+1} \mathbf{v}_{m+1}^T.$$

Note that this TLS correction matrix has rank one. It is clear that the approximate set

$$[\hat{\mathbf{A}} \hat{\mathbf{y}}] \begin{bmatrix} \xi \\ -1 \end{bmatrix} = 0 \Rightarrow \mathbf{U} \hat{\Sigma} \mathbf{V}^T \begin{bmatrix} \xi \\ -1 \end{bmatrix} = 0 \Rightarrow \hat{\Sigma} \mathbf{V}^T \begin{bmatrix} \xi \\ -1 \end{bmatrix} = 0$$

is now compatible and its solution is given by the only vector \mathbf{v}_{m+1} (i.e., that last column of \mathbf{V}) which belongs to $\mathcal{N}([\hat{\mathbf{A}}; \hat{\mathbf{y}}])$. Thus

$$\begin{bmatrix} \xi \\ -1 \end{bmatrix} = \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \cdots + \lambda_m \mathbf{v}_m + \lambda_{m+1} \mathbf{v}_{m+1}.$$

If $\sigma_i \neq 0, \lambda_i = 0$, then

$$\begin{bmatrix} \xi \\ -1 \end{bmatrix} = \lambda_{m+1} \mathbf{v}_{m+1} \Rightarrow \lambda_{m+1} = -\frac{1}{v_{m+1,m+1}}.$$

The TLS solution is then obtained by scaling \mathbf{v}_{m+1} such that its last component is -1, i.e.,

$$\begin{bmatrix} \xi \\ -1 \end{bmatrix} = \frac{-1}{v_{m+1,m+1}} \mathbf{v}_{m+1}. \quad (3.2.14)$$

If $v_{m+1,m+1} \neq 0$, then $\hat{\mathbf{y}} = \hat{\mathbf{A}} \hat{\xi} = -1/(v_{m+1,m+1}) \hat{\mathbf{A}} [v_{1,m+1}, \dots, v_{m,m+1}]^T \in \mathcal{R}(\hat{\mathbf{A}})$. Hence $\hat{\xi}$ solves the basic TLS problem. Observe that, if σ_{m+1} is zero, $[\mathbf{A}; \mathbf{y}]$ is of rank m and hence $\mathbf{v}_{m+1} \in \mathcal{N}([\mathbf{A}; \mathbf{y}])$. The following theorem gives conditions for the uniqueness and existence of a TLS solution.

Theorem 4 (Solution of the basic TLS problem $\mathbf{y} \approx \mathbf{A}\xi$). Suppose σ'_m is the smallest singular value of \mathbf{A} . If $\sigma'_m > \sigma_{m+1}$, then

$$[\hat{\mathbf{A}}; \hat{\mathbf{y}}] = \mathbf{U} \hat{\Sigma} \mathbf{V}^T \quad \text{and} \quad \hat{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_m, 0) \quad (3.2.15)$$

with corresponding TLS correction matrix

$$[\hat{\mathbf{E}}_A; \hat{\mathbf{e}}] = [\mathbf{A}; \mathbf{y}] - [\hat{\mathbf{A}}; \hat{\mathbf{y}}] = \sigma_{m+1} \mathbf{u}_{m+1} \mathbf{v}_{m+1}^T \quad (3.2.16)$$

Solves the TLS problem and

$$\hat{\xi} = -\frac{1}{v_{m+1,m+1}} [v_{1,m+1}, \dots, v_{m,m+1}]^T. \quad (3.2.17)$$

Exists and is the unique solution to $\hat{\mathbf{y}} = \hat{\mathbf{A}} \hat{\xi}$. It is interesting to note here that the conditions

$$\sigma'_m > \sigma_{m+1} \Leftrightarrow \sigma_m > \sigma_{m+1} \quad \text{and} \quad v_{m+1,m+1} \neq 0$$

are equivalent (Van Huffel, 1991).

An illustration of the geometry of the TLS solution in the column space of \mathbf{A} is depicted in Figure 5 for $m = 2$. The TLS problem is tantamount to finding a ‘closest’

subspace $R([\hat{\mathbf{A}}; \hat{\mathbf{y}}])$ to the $m+1$ columns of $[\mathbf{A}; \mathbf{y}]$. Hereto, the sum of squared perpendicular distances from each column $A_i, \mathbf{y}_i \in \mathfrak{R}^n$ to $R([\hat{\mathbf{A}}; \hat{\mathbf{y}}])$ is minimized, and each column A_i, \mathbf{y}_i is approximated by its orthogonal projection $\hat{A}_i, \hat{\mathbf{y}}_i$ onto that subspace.

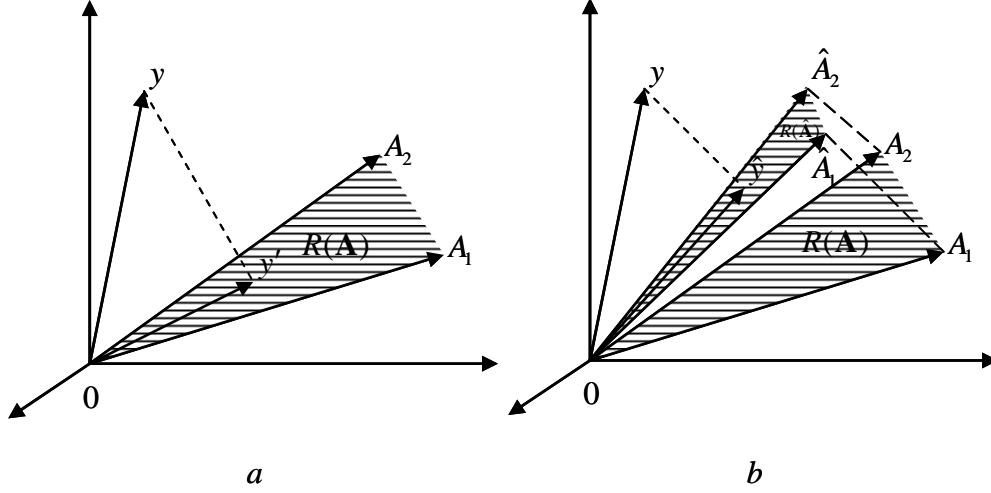


Figure 5 (a) The LS solution is obtained by projecting y orthogonally onto $R(A)$ and solving $\hat{y} = A\xi$.
(b) The TLS solution is obtained by approximating the columns A_i of A and y by \hat{A}_i and \hat{y} until \hat{y} is in the space $R(\hat{A})$, generated by the columns \hat{A}_i , and $\hat{y} = \hat{A}\xi$.

A well-known and useful characterization of the TLS solution $\hat{\xi}$ and the minimal TLS correction σ_{m+1} is proved in the next theorem.

Theorem 5 (Closed-form expression of the basic TLS solution). Let $[\mathbf{A}; \mathbf{y}] = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ be the SVD of $[\mathbf{A}; \mathbf{y}]$. If $\sigma'_m > \sigma_{m+1}$, then

$$\hat{\xi} = (\mathbf{A}^T \mathbf{A} - \sigma_{m+1}^2 \mathbf{I})^{-1} \mathbf{A}^T \mathbf{y}$$

and

$$\sigma_{m+1}^2 \left[1 + \sum_{i=1}^m \frac{(\mathbf{u}_i^T \mathbf{y})}{\sigma_i'^2 - \sigma_{m+1}^2} \right] = \min_{\xi \in \mathfrak{R}^m} \|\mathbf{A}\xi - \mathbf{y}\|_2^2.$$

The following algorithm computes a TLS solution $\hat{\xi}$ of $\mathbf{y} \approx \mathbf{A}\xi$ such that $\mathbf{y} - \mathbf{e} = (\mathbf{A} - \mathbf{E}_A)\hat{\xi}$ and $\|[\mathbf{E}_A; \mathbf{e}]\|_F$ is minimal (Van Huffel, 1991).

Algorithm 1 (Basic TLS solution of $\mathbf{A}\xi \approx \mathbf{y}$).

Given: $\mathbf{A} \in \mathfrak{R}^{n \times m}$ and $\mathbf{y} \in \mathfrak{R}^n$;

Step 1: Compute the SVD,

$$\text{i.e., } [\mathbf{A}; \mathbf{y}] = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T. \quad (3.2.18)$$

$$\text{Step 2: } v_{m+1,m+1} \neq 0 \Rightarrow \hat{\xi} = -\frac{1}{v_{m+1,m+1}} [v_{1,m+1}, \dots, v_{m,m+1}]^T \quad (3.2.19)$$

End

The conditions $\sigma'_m > \sigma_{m+1}$ or, equivalently, $\sigma_m > \sigma_{m+1}$ and $v_{m+1,m+1} \neq 0$, ensure that algorithm 1 computes the unique TLS solution of $\mathbf{y} \approx \mathbf{A}\xi$. These conditions are generally satisfied provided \mathbf{A} is of full rank. Hence, most TLS problems that arise in practice can be solved by means of algorithm 1 (Van Huffel, 1991).

3.3. The Euler-Lagrange approach (an alternative method)

Recently, Schaffrin and Felus (2003) have introduced a multivariate version of Total Least Squares (TLS) adjustment in order to determine the optimal parameters of an affine coordinate transformation empirically.

The following model, with full-rank matrix \mathbf{A} , is assumed (Schaffrin, 2003):

$$\begin{aligned} (\mathbf{A} - \mathbf{E}_A)\xi - (\mathbf{y} - \mathbf{e}) &= 0 \\ E\{\mathbf{E}_A, \mathbf{e}\} &= 0 \\ C\{\mathbf{E}_A, \mathbf{e}\} &= 0 \\ D\{\text{vec}[\mathbf{E}_A, \mathbf{e}]\} &= \Sigma_0 \otimes \mathbf{I}_n \end{aligned} \quad (3.3.1)$$

Where \mathbf{e} and \mathbf{E}_A denote a random error vector, resp. matrix. $\Sigma_0 = \sigma_0^2 \mathbf{I}_{m+1}$ is a $(m+1) \times (m+1)$ matrix with an unknown variance component σ_0^2 and given identity matrix \mathbf{I}_{m+1} . The symbol \otimes denotes the ‘‘Kronecker-Zehfuss product’’ of matrices, defined by:

$$\mathbf{M} \otimes \mathbf{N} := [m_{ij} \cdot \mathbf{N}]$$

For $\mathbf{M} = [m_{ij}]$ and \mathbf{N} arbitrary.

The ‘‘vec’’ operator stacks one column of a matrix under the other, moving from left to right. In contrast to the Least-Squares (LS) method, this is based on the minimization of

$$\mathbf{e}^T \mathbf{e} = (\mathbf{y} - \mathbf{A}\xi)^T (\mathbf{y} - \mathbf{A}\xi). \quad (3.3.2)$$

Under the condition $\mathbf{E}_A := 0$, the (equally weighted) TLS principle is based on minimizing the objective function (Schaffrin, 2003):

$$\mathbf{e}^T \mathbf{e} + (\text{vec} \mathbf{E}_A)^T (\text{vec} \mathbf{E}_A) = \min(\xi), \quad (3.3.3)$$

When performing an adjustment, it is sometimes necessary to fix some parameters to specific values. Here, a total least squares solution will be presented, along with an

iteration scheme (Schaffrin, 2003).

In order to solve the TLS problem as presented in (3.3.3) and minimize the respective objective function in view of the model (3.3.1), we define the Lagrange target function as follows where

$$\begin{aligned} \mathbf{e}_A &:= \text{vec}(\mathbf{E}_A) \sim (0, \sigma_0^2 \mathbf{I}_m \otimes \mathbf{I}_n) \\ \Phi(\mathbf{e}, \mathbf{e}_A, \boldsymbol{\lambda}, \boldsymbol{\xi}) &= \mathbf{e}^T \mathbf{e} + \mathbf{e}_A^T \mathbf{e}_A + 2\boldsymbol{\lambda}^T [\mathbf{y} - \mathbf{e} - \mathbf{A}\boldsymbol{\xi} + \mathbf{E}_A \boldsymbol{\xi}] \end{aligned} \quad (3.3.4)$$

Where $\boldsymbol{\lambda}$ denotes the $n \times 1$ vector of Lagrange multipliers; note that

$$\mathbf{E}_A \boldsymbol{\xi} = (\boldsymbol{\xi}^T \otimes \mathbf{I}_n) \mathbf{e}_A, \quad (3.3.5)$$

thus the Euler-Lagrange necessary conditions are (Schaffrin, 2003):

$$\begin{aligned} \frac{1}{2} \frac{\partial \Phi}{\partial \mathbf{e}} &= \hat{\mathbf{e}} - \hat{\boldsymbol{\lambda}} = 0, \\ \frac{1}{2} \frac{\partial \Phi}{\partial \mathbf{e}_A} &= \hat{\mathbf{e}}_A - (\hat{\boldsymbol{\xi}} \otimes \mathbf{I}_n) \hat{\boldsymbol{\lambda}} = 0, \\ \frac{1}{2} \frac{\partial \Phi}{\partial \boldsymbol{\lambda}} &= \mathbf{y} - \mathbf{A}\hat{\boldsymbol{\xi}} - \hat{\mathbf{e}} + \hat{\mathbf{E}}_A \hat{\boldsymbol{\xi}} = 0, \\ \frac{1}{2} \frac{\partial \Phi}{\partial \boldsymbol{\xi}} &= -\mathbf{A}^T \hat{\boldsymbol{\lambda}} + \hat{\mathbf{E}}_A^T \hat{\boldsymbol{\lambda}} = 0. \end{aligned}$$

This system is simplified into:

$$\begin{aligned} (\mathbf{A}^T \mathbf{A}) \hat{\boldsymbol{\xi}} &= \mathbf{A}^T \mathbf{y} + \hat{\boldsymbol{\xi}} \underbrace{(\hat{\boldsymbol{\lambda}}^T \hat{\boldsymbol{\lambda}})}_{\hat{\nu}} (1 + \hat{\boldsymbol{\xi}}^T \hat{\boldsymbol{\xi}}) \\ \hat{\boldsymbol{\lambda}} &= (\hat{\mathbf{e}} - \hat{\mathbf{E}}_A \hat{\boldsymbol{\xi}}) (1 + \hat{\boldsymbol{\xi}}^T \hat{\boldsymbol{\xi}})^{-1} = (\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\xi}}) (1 + \hat{\boldsymbol{\xi}}^T \hat{\boldsymbol{\xi}})^{-1} \end{aligned} \quad (3.3.6)$$

Note that

$$\hat{\nu} = \frac{(\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\xi}})^T (\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\xi}})}{(1 + \hat{\boldsymbol{\xi}}^T \hat{\boldsymbol{\xi}})} = (\hat{\boldsymbol{\lambda}}^T \hat{\boldsymbol{\lambda}}) (1 + \hat{\boldsymbol{\xi}}^T \hat{\boldsymbol{\xi}}) = \hat{\mathbf{e}}^T \hat{\mathbf{e}} + \hat{\mathbf{E}}_A^T \hat{\mathbf{E}}_A \hat{\boldsymbol{\xi}} = \min(\boldsymbol{\xi}) \quad (3.3.7)$$

is Rayleigh's quotient for the matrix

$$\begin{bmatrix} \mathbf{A}^T \mathbf{A} & \mathbf{A}^T \mathbf{y} \\ \mathbf{y}^T \mathbf{A} & \mathbf{y}^T \mathbf{y} \end{bmatrix} \quad (3.3.8)$$

with $\begin{bmatrix} \hat{\boldsymbol{\xi}}^T, -1 \end{bmatrix}^T$ as the vector argument. Rayleigh's quotient defines the minimum eigenvalue of the augmented matrix, based on the corresponding eigenvector (see, e.g.,

G. Strang, 1988) (Schaffrin, 2003).

Using these equations the following algorithm had been developed by Schaffrin (2003) to solve the TLS problem (Schaffrin, 2003):

1) Compute the LS solution:

$$\hat{\xi}^1 = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y} \quad (\text{for } \hat{v}^0 := 0)$$

2) Insert the solution of step (1) as the initial value for the following iterative process:

$$\hat{\xi}^{i+1} = (\mathbf{A}^T \mathbf{A})^{-1} \left[\mathbf{A}^T \mathbf{y} + \hat{\xi}^i \frac{(\mathbf{y} - \mathbf{A} \hat{\xi}^i)^T (\mathbf{y} - \mathbf{A} \hat{\xi}^i)}{(1 + (\hat{\xi}^i)^T \hat{\xi}^i)} \right] \quad (3.3.9)$$

3) End when $\|\hat{\xi}^{i+1} - \hat{\xi}^i\| < \varepsilon$. Then

$$\hat{\sigma}_0^2 = \frac{\hat{v}}{(n - m)} \quad (3.3.10)$$

The algorithm seems to converge to the TLS solution in most cases although its efficiency (convergence rate, convergence radius, etc.) still needs to be further investigated. Nevertheless, the establishment of equivalence between the Euler-Lagrange approach and the minimum eigenvalue method, provides an insight into the process that will support future developments.

Example: The following 6-parameter coordinate transformation problem was used to test the procedure and compare it with the standard least-squares approach.

No.	\mathbf{y}_1	\mathbf{y}_2	\mathbf{x}_1	\mathbf{x}_2
1	20	20	275	160
2	80	20	403	70
3	80	100	550	210
4	20	100	390	310
Mean value	50	60	404.5	187.5

(Data Quelle: Schaffrin 2005)

Here, the coordinates set $\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{y}_1 \end{bmatrix}$ is transformed into the second coordinates set $\begin{bmatrix} \mathbf{x}_2 \\ \mathbf{y}_2 \end{bmatrix}$,

thus:

$$\begin{bmatrix} \mathbf{x}_2 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{y}_1 \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} \Rightarrow \begin{bmatrix} \mathbf{x}_1 & \mathbf{y}_1 & 0 & 0 & 1 & 0 \\ 0 & 0 & \mathbf{x}_1 & \mathbf{y}_1 & 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \end{bmatrix}$$

Because there is no error that belongs to design matrix element “1”, the translation parameters shall disappear by centering the equation.

$$\begin{bmatrix} \underline{\mathbf{x}}_2 \\ \underline{\mathbf{y}}_2 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \underline{\mathbf{x}}_1 \\ \underline{\mathbf{y}}_1 \end{bmatrix}$$

with

$$\underline{\mathbf{x}}_2 = \mathbf{x}_2 - \text{mean}(\mathbf{x}_2), \underline{\mathbf{x}}_1 = \mathbf{x}_1 - \text{mean}(\mathbf{x}_1), \underline{\mathbf{y}}_2 = \mathbf{y}_2 - \text{mean}(\mathbf{y}_2), \underline{\mathbf{y}}_1 = \mathbf{y}_1 - \text{mean}(\mathbf{y}_1)$$

Then, the four parameters of the affine transformation matrix are estimated individually with the approaches from Van Huffel and Schaffrin.

Parameter	a	b	c	d
TLS solution with SVD	2.4214	1.6418	-1.590	1.8111
TLS solution by Schaffrin	2.4214	1.6418	-1.590	1.8177
Standard LS solution	2.40	1.6375	-1.583	1.8125

The resulting TLS sum of squares is:

$$\hat{\mathbf{e}}_{TLS}^T \mathbf{e}_{TLS} + \hat{\mathbf{E}}_A^T \hat{\mathbf{E}}_A = 18.37$$

The resulting LS sum of squares is:

$$\hat{\mathbf{e}}_{LS}^T \hat{\mathbf{e}}_{LS} = 281.0$$

From above, it is clearly that the two TLS approaches have the same results.

3.4. Properties of the Total Least Squares problem

Let us now discuss some of the main properties of the TLS method. The properties of TLS are best understood by comparing them with those of LS. First, a lot of insight can be gained by comparing their analytical expressions, given by:

$$\text{LS: } \xi_{LS} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y} \quad (3.3.11)$$

$$\text{TLS: } \xi_{TLS} = (\mathbf{A}^T \mathbf{A} - \sigma_{m+1}^2 \mathbf{I})^{-1} \mathbf{A}^T \mathbf{y} \quad (3.3.12)$$

With \mathbf{A} of full rank and σ_{m+1} the smallest singular value of the augmented design matrix $[\mathbf{A}; \mathbf{y}]$ (Van Huffel and Lemmerling, 2002).

From a numerical analyst's point of view, these formulas tell us that the TLS solution is more ill conditioned than the LS solution since it has a higher condition number. This implies that errors in the data more likely affect the TLS solution than the LS solution. This is particularly true under worst-case perturbations. Hence, TLS can be considered as a kind of deregularizing procedure. However, from a statistical point of view, these formulas tell us that TLS is doing the right thing in the presence of independently and identically-distributed (i.i.d.) equally sized errors since it removes (asymptotically) the bias by subtracting the error covariance matrix (estimated by $\sigma_{m+1}^2 I$) from the data covariance matrix $\mathbf{A}^T \mathbf{A}$. Secondly, while LS minimizes a sum of squared residuals, TLS minimizes a sum of squared normalized residuals, expressed as follows:

$$\text{LS:} \quad \min_{\xi} \|\mathbf{A}\xi - \mathbf{y}\|^2 \quad (3.3.13)$$

$$\text{TLS:} \quad \min_{\mathbf{z} = \begin{bmatrix} \xi \\ -1 \end{bmatrix}} \frac{\|[\mathbf{A}; \mathbf{y}]\mathbf{z}\|^2}{\|\mathbf{z}\|^2} = \min_{\xi} \frac{\|\mathbf{A}\xi - \mathbf{y}\|^2}{\|\xi\|^2 + 1} \quad (3.3.14)$$

From a numerical analyst's point of view, we say that TLS minimizes the Rayleigh quotient. From a statistical point of view, we say that we normalize the residuals by multiplying them with the inverse of their covariance matrix (up to a scaling factor) in order to derive consistent estimates (Van Huffel and Lemmerling, 2002).

Other properties of TLS, which were studied in the field of numerical analysis, are its sensitivity in the presence of errors on all data. Differences between the LS and TLS solution are shown to increase when the ratio $\sigma_m([\mathbf{A}; \mathbf{y}]) / \sigma_m(\mathbf{A})$ is growing. This is the case when the set of equations $\mathbf{y} \approx \mathbf{A}\xi$ becomes less compatible, when the Vector \mathbf{y} is growing in length and when \mathbf{A} tends to be rank-deficient. Assuming i.i.d. equally sized errors, the improved accuracy of the TLS solution compared to that of LS is maximal when the orthogonal projection of \mathbf{y} is parallel with the n th singular vector of \mathbf{A} , corresponding to its smallest singular value $\sigma_m(\mathbf{A})$. Additional algebraic connections and sensitivity properties of the TLS and LS problem, as well as many more statistical properties of the TLS estimators, based on knowledge of the distribution of the errors in the data, are analyzed in the two books on TLS (Van Huffel and Lemmerling, 2002).

4. Case studies: Applications of the TLS method in Baden-Württemberg

4.1. Six parameter Affine transformation models (2D)

We apply here the TLS solution to the centralized 6-parameter affine transformation model where the translation parameters are vanished. Let rewrite the 6-parameter affine transformation in six parameters as

$$\begin{aligned} \begin{bmatrix} N \\ E \end{bmatrix} &= \begin{bmatrix} \lambda_H \cos \alpha & -\lambda_R \sin \beta \\ \lambda_H \sin \alpha & \lambda_R \cos \beta \end{bmatrix} \begin{bmatrix} H \\ R \end{bmatrix} + \begin{bmatrix} t_N \\ t_E \end{bmatrix} \\ &=: \begin{bmatrix} \xi_{11} & \xi_{21} \\ \xi_{12} & \xi_{22} \end{bmatrix} \begin{bmatrix} H \\ R \end{bmatrix} + \begin{bmatrix} \xi_{31} \\ \xi_{32} \end{bmatrix} = \begin{bmatrix} H & R & 0 & 0 & 1 & 0 \\ 0 & 0 & H & R & 0 & 1 \end{bmatrix} \begin{bmatrix} \xi_{11} \\ \xi_{21} \\ \xi_{12} \\ \xi_{22} \\ \xi_{31} \\ \xi_{32} \end{bmatrix} \end{aligned} \quad (4.1.1)$$

Because there is no error that belongs to design matrix element “1”, the translation parameters shall disappear by centering this equation. Thus, after the centering the coordinates in the midpoint, the translation parameters t_N and t_E will automatically vanish. Then the observation and old coordinates are centered on their average values in the form:

$$\begin{bmatrix} \underline{N} \\ \underline{E} \end{bmatrix} =: \begin{bmatrix} \xi_{11} & \xi_{21} \\ \xi_{12} & \xi_{22} \end{bmatrix} \begin{bmatrix} \underline{H} \\ \underline{R} \end{bmatrix} \quad (4.1.2)$$

with

$$\begin{aligned} \underline{N} &= N - \text{mean}(N), \quad \underline{E} = E - \text{mean}(E) \\ \underline{H} &= H - \text{mean}(H), \quad \underline{R} = R - \text{mean}(R) \end{aligned}$$

For the n couple of coordinates we have the transformation model, which is suited for the application of TLS solution

$$E \begin{Bmatrix} \underline{N}_1 \\ \dots \\ \underline{N}_n \\ \underline{E}_1 \\ \dots \\ \underline{E}_n \end{Bmatrix} = E \begin{Bmatrix} \underline{H}_1 & \underline{R}_1 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ \underline{H}_n & \underline{R}_n & 0 & 0 \\ 0 & 0 & \underline{H}_1 & \underline{R}_1 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \underline{H}_n & \underline{R}_n \end{Bmatrix} \begin{Bmatrix} \xi_{11} \\ \xi_{21} \\ \xi_{12} \\ \xi_{22} \end{Bmatrix} \quad (4.1.3)$$

$$(\mathbf{y} - \mathbf{e}) = (\mathbf{A} - \mathbf{E}_A)\xi \quad (4.1.4)$$

$$\mathbf{e}^T \mathbf{e} + \mathbf{E}_A^T \mathbf{E}_A = \min(\mathbf{e}, \mathbf{E}_A, \xi) \quad (4.1.5)$$

Solution of the TLS problem by using the singular value decomposition (SVD).

$$\xi_{TLS} = (\mathbf{A}^T \mathbf{A} - \sigma_{m+1}^2 \mathbf{I})^{-1} \mathbf{A}^T \mathbf{y} \quad (4.1.6)$$

with σ_{m+1} the smallest singular value of the augmented design matrix $[\mathbf{A}; \mathbf{y}]$:

$$[\mathbf{A}; \mathbf{y}] = \mathbf{U} \Sigma \mathbf{V}^T = \sum_{i=0}^{m+1} \sigma_i \mathbf{u}_i \mathbf{v}_i^T, \quad \sigma_1 \geq \dots \geq \sigma_{m+1} \geq 0.$$

The best TLS approximation $[\hat{\mathbf{A}}; \hat{\mathbf{y}}]$ of $[\mathbf{A}; \mathbf{y}]$ is give by

$$[\hat{\mathbf{A}}; \hat{\mathbf{y}}] = \mathbf{U} \hat{\Sigma} \mathbf{V}^T, \quad \text{with } \hat{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_m, 0)$$

and with corresponding TLS correction matrix

$$[\hat{\mathbf{E}}_A; \hat{\mathbf{e}}] = [\mathbf{A}; \mathbf{y}] - [\hat{\mathbf{A}}; \hat{\mathbf{y}}] = \sigma_{m+1} \mathbf{u}_{m+1} \mathbf{v}_{m+1}^T.$$

with MATLAB function $[\mathbf{U}, \mathbf{S}, \mathbf{V}] = \text{svd}(\mathbf{X})$ these procedures can be implemented easily.

Through the TLS solution where the errors in the design matrix \mathbf{A} are considered the remaining transformation coordinate residuals of collocated DHDN points in Baden-Württemberg are reduced from 11.8 cm to 3.9 cm, which are illustrated in figure 8 in comparison with figure 2 in detail. The statistics of these residual and the transformation parameters in comparison with LS methods are listed in Table 2 and Table 3. The following statistical terms shows us the difference between the quadratics sums of the residuals $\hat{\mathbf{e}}_{LS}^T \hat{\mathbf{e}}_{LS}$ related LS and $\hat{\mathbf{e}}_{TLS}^T \hat{\mathbf{e}}_{TLS}$ related TLS, together with the quadratics sums of the errors of vectorized design matrix:

$$\hat{\mathbf{e}}_{LS}^T \hat{\mathbf{e}}_{LS} = 3.678308 \text{ (m}^2\text{)}$$

$$\hat{\mathbf{e}}_{TLS}^T \hat{\mathbf{e}}_{TLS} = 0.409136 \text{ (m}^2\text{)}$$

$$\hat{\mathbf{E}}_A^T \hat{\mathbf{E}}_A = 0.817619 \text{ (m}^2\text{)}$$

$$\hat{\mathbf{e}}_{TLS}^T \hat{\mathbf{e}}_{TLS} + \hat{\mathbf{E}}_A^T \hat{\mathbf{E}}_A = 1.226756 \text{ (m}^2\text{)}$$

Table 2 Numerical comparison of 6-P affine transformation with LS and TLS estimator

Transformation models	Collocated sites	Absolute mean Residuals (m)		Max. of absolute residuals (m)		RMS (m)	Standard deviation of unit weight (m)
		[V _N]	[V _E]	[V _N]	[V _E]		
LS	B-W 131	0.1049	0.0804	0.3288	0.3226	0.1187	0.1199
TLS	B-W 131	0.0350	0.0268	0.1097	0.1076	0.0396	0.0400

Table 3 Comparison of 6-p affine transformation parameters with LS and TLS

131 BWREF points	6-parameter affine transformation GK (DHDN) – UTM (ETRS89)					
	t_N (m)	t_E (m)	α (")	β (")	$d\lambda_H$ ($\times 10^{-4}$)	$d\lambda_R$ ($\times 10^{-4}$)
LS	437.194567	119.756709	0.165368	-0.196455	-3.996797	-3.988430
TLS	437.194554	119.756712	0.165368	-0.196455	-3.996797	-3.988430
TLS-LS	-0.000013	0.000003	0	0	0	0

In order to test the practicality of the TLS method in Baden-Württemberg, we select 10 points as checkpoints from the 131 points. Then transformation parameters are estimated depending on the rest 121 points.

The process is summarized here:

- 1) Calculate transformation parameters (121 points)

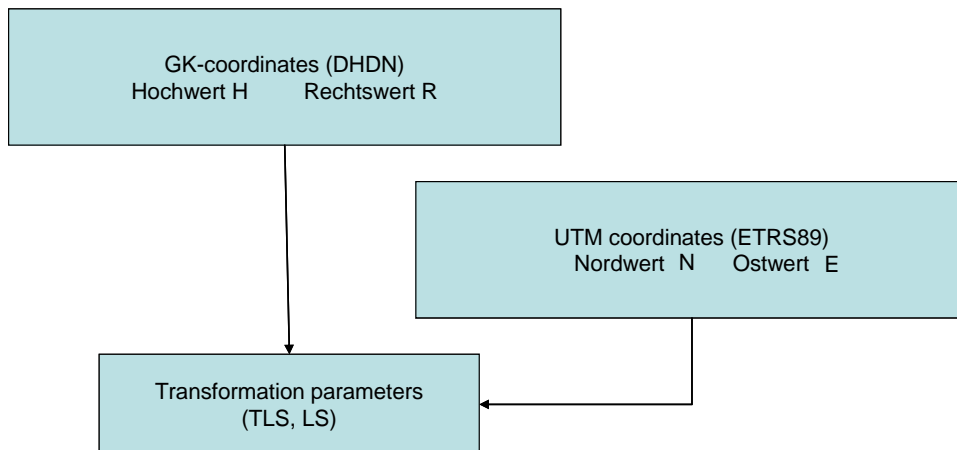


Figure 6 Calculation of the transformation parameters

- 2) Use the following formula to calculate the estimated UTM coordinates: (10 points)

$$\begin{bmatrix} \hat{N} \\ \hat{E} \end{bmatrix} = \begin{bmatrix} \xi_{11} & \xi_{21} \\ \xi_{12} & \xi_{22} \end{bmatrix} \begin{bmatrix} H \\ R \end{bmatrix} + \begin{bmatrix} \xi_{31} \\ \xi_{32} \end{bmatrix} \quad (4.1.7)$$

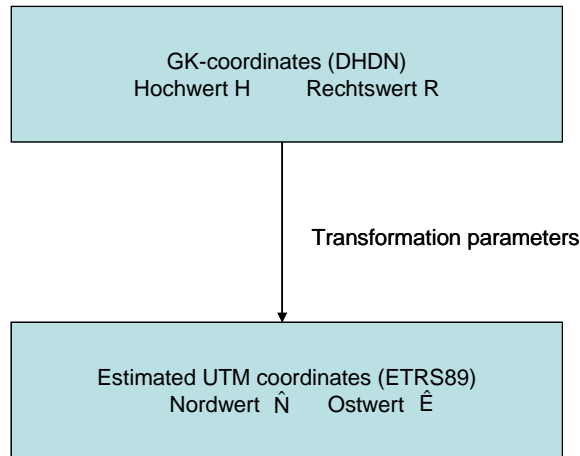


Figure 7 Calculation of the estimated UTM coordinates (ETRS89)

3) Compare with the original value (10 points)

$$\begin{bmatrix} d\hat{N} \\ d\hat{E} \end{bmatrix} = \begin{bmatrix} N - \hat{N} \\ E - \hat{E} \end{bmatrix} \quad (4.1.8)$$

The results are in table 4.

Table 4 Numerical comparison between LS and TLS

Points number	LS (m)		TLS (m)	
	$d\hat{N}$	$d\hat{E}$	$d\hat{N}$	$d\hat{E}$
652000308	+0.0920	+0.1934	+0.0920	+0.1934
672500108	-0.1465	+0.0357	-0.1465	+0.0357
701600288	+0.1289	-0.0213	+0.1289	-0.0213
722000300	+0.0551	+0.0003	+0.0551	+0.0003
722600208	-0.1468	-0.1983	-0.1468	-0.1983
751300208	+0.0218	-0.0409	+0.0218	-0.0409
761900108	+0.1243	+0.1138	+0.1243	+0.1138
792311808	+0.0406	-0.0141	+0.0406	-0.0141
811300108	-0.0875	-0.0945	-0.0875	-0.0945
811800108	-0.1544	+0.1293	-0.1544	+0.1293

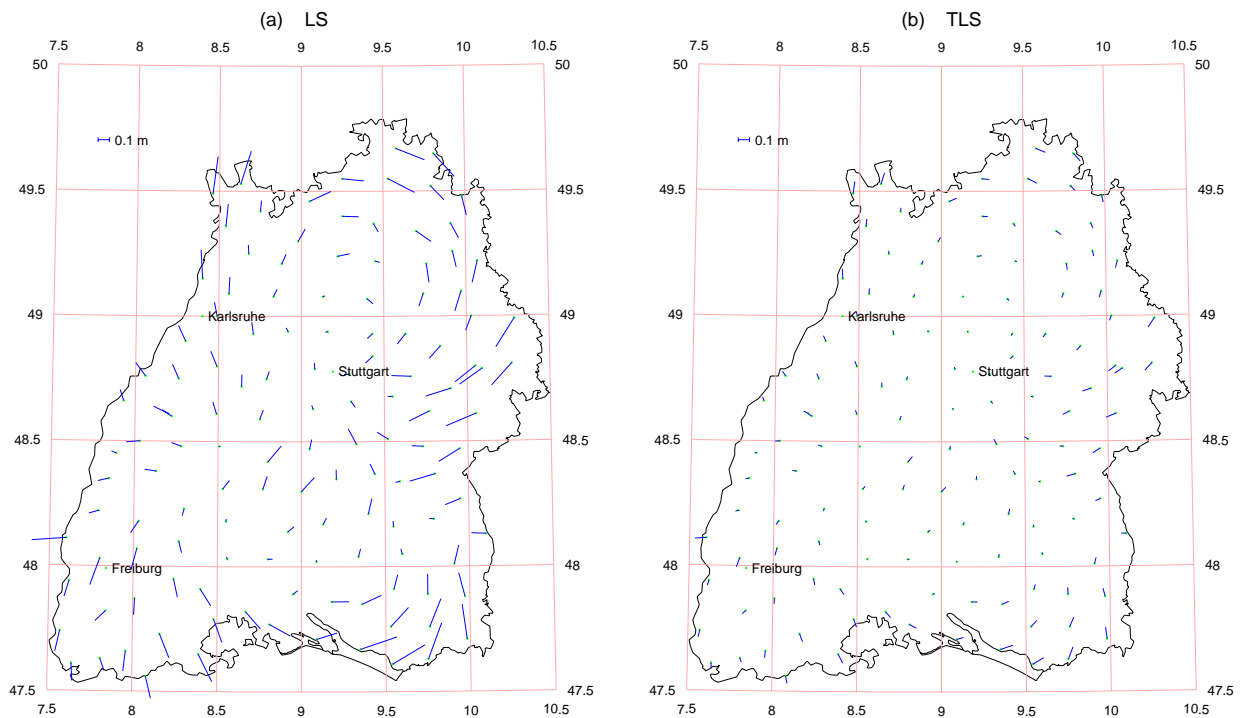


Figure 8 Horizontal residuals after 6-P affine transformation with TLS estimator in Baden-Württemberg Network (a) LS (b) TLS

It should be pointed out, though the residuals are reduced with TLS solution, the estimated 6 transformation parameters with TLS method haven't significant difference in comparison with those with LS methods. For these points with classic DHDN Gauss-Krüger coordinates, which are not collocated with new global technique, such as GPS, the so-called new points in coordinate transformation. When they are transformed with the estimated parameters with TLS method, the residuals of transformation are also remain larger, which are in the level of the residuals after transformation with LS method.

From table 4 we find that the residuals after transformation with LS and TLS method are almost identical and less than 2 dm. i.e., in practical situation TLS method hasn't obvious advantage in comparison with LS method.

For clearly display, these selected 10 points are transformed and their residuals are shown in the figure 9(b) with red color.

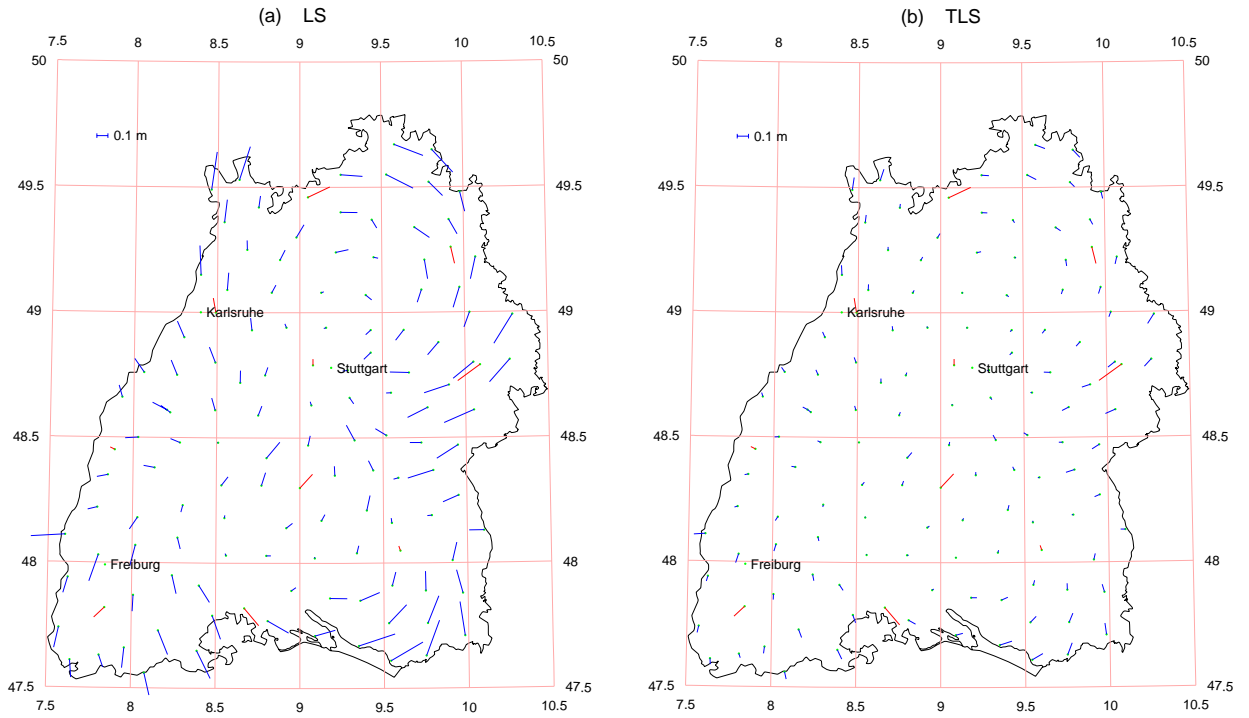


Figure 9 Horizontal residuals of 10 check points (a) LS (b) TLS

4.2. Seven parameter Helmert transformation model (3D)

The following formula has been used for the estimation of the parameters in seven-parameter Helmert transformation:

$$\begin{pmatrix} X_G \\ Y_G \\ Z_G \end{pmatrix} = (1+d\lambda) \begin{bmatrix} 1 & \gamma & -\beta \\ -\gamma & 1 & \alpha \\ \beta & -\alpha & 1 \end{bmatrix} \begin{pmatrix} X_L \\ Y_L \\ Z_L \end{pmatrix} + \begin{pmatrix} T_X \\ T_Y \\ T_Z \end{pmatrix}_L \quad (4.2.1)$$

$$\mathbf{X}_G = \lambda \begin{bmatrix} 1 & \gamma & -\beta \\ -\gamma & 1 & \alpha \\ \beta & -\alpha & 1 \end{bmatrix} \mathbf{X}_L + \mathbf{T}_L$$

Where λ is scale factor, α , β , γ are rotation angels. The translation terms T_X, T_Y, T_Z are the coordinates of the origin of the 3-D network.

After the linearization, the formula is rewritten:

$$\underbrace{\begin{bmatrix} X_G \\ Y_G \\ Z_G \end{bmatrix}}_{\mathbf{Y}_i} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & -Z_L & Y_L & X_L \\ 0 & 1 & 0 & Z_L & 0 & -X_L & Y_L \\ 0 & 0 & 1 & -Y_L & X_L & 0 & Z_L \end{bmatrix}}_{\bar{\mathbf{A}}_i} \underbrace{\begin{bmatrix} T_X \\ T_Y \\ T_Z \\ \delta\alpha \\ \delta\beta \\ \delta\gamma \\ \lambda \end{bmatrix}}_{\boldsymbol{\eta}} \quad (4.2.2)$$

$$\mathbf{Y}_i = \bar{\mathbf{A}}_i \boldsymbol{\eta} \quad (4.2.3)$$

After centering the coordinates in the midpoint, the translation parameter T_X, T_Y, T_Z will disappear, and then the observations and old coordinates are centered on their average values. This will be assumed in the following:

$$\underbrace{\begin{bmatrix} x_g \\ y_g \\ z_g \end{bmatrix}}_{\mathbf{y}_i} = \underbrace{\begin{bmatrix} 0 & -z_l & y_l & x_l \\ z_l & 0 & -x_l & y_l \\ -y_l & x_l & 0 & z_l \end{bmatrix}}_{\mathbf{A}_i} \underbrace{\begin{bmatrix} \delta\alpha \\ \delta\beta \\ \delta\gamma \\ \lambda \end{bmatrix}}_{\boldsymbol{\xi}} \quad (4.2.4)$$

$$\mathbf{y}_i = \mathbf{A}_i \cdot \boldsymbol{\xi} \quad (4.2.5)$$

$$\begin{bmatrix} x_g \\ y_g \\ z_g \end{bmatrix}_i = \begin{bmatrix} X_G \\ Y_G \\ Z_G \end{bmatrix}_i - \text{mean} \begin{bmatrix} X_G \\ Y_G \\ Z_G \end{bmatrix}_i, \quad \begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix}_i = \begin{bmatrix} X_L \\ Y_L \\ Z_L \end{bmatrix}_i - \text{mean} \begin{bmatrix} X_L \\ Y_L \\ Z_L \end{bmatrix}_i \quad (4.2.6)$$

Then, we have the transformation model which is suited for the application of TLS solution:

$$E \left\{ \begin{bmatrix} x_{g1} \\ \dots \\ x_{gn} \\ y_{g1} \\ \dots \\ y_{gn} \\ z_{g1} \\ \dots \\ z_{gn} \end{bmatrix} \right\} = E \left\{ \begin{bmatrix} 0 & -z_{l1} & y_{l1} & x_{l1} \\ \dots & \dots & \dots & \dots \\ 0 & -z_{ln} & y_{ln} & x_{ln} \\ z_{l1} & 0 & -x_{l1} & y_{l1} \\ \dots & \dots & \dots & \dots \\ z_{ln} & 0 & -x_{ln} & y_{ln} \\ -y_{l1} & x_{l1} & 0 & z_{l1} \\ \dots & \dots & \dots & \dots \\ -y_{ln} & x_{ln} & 0 & z_{ln} \end{bmatrix} \right\} \begin{bmatrix} \delta\alpha \\ \delta\beta \\ \delta\gamma \\ \lambda \end{bmatrix} \quad (4.2.7)$$

$$(\mathbf{y} - \mathbf{e}) = (\mathbf{A} - \mathbf{E}_A) \boldsymbol{\xi} \quad (4.2.8)$$

$$\mathbf{e}^T \mathbf{e} + \mathbf{E}_A^T \mathbf{E}_A = \min(\mathbf{e}, \mathbf{E}_A, \boldsymbol{\xi}), \quad (4.2.9)$$

Solution of the TLS problem by using the singular value decomposition (SVD).

$$\xi_{TLS} = (\mathbf{A}^T \mathbf{A} - \sigma_{m+1}^2 \mathbf{I})^{-1} \mathbf{A}^T \mathbf{y} \quad (4.2.10)$$

with σ_{m+1} the smallest singular value of the augmented design matrix $[\mathbf{A}; \mathbf{y}]$:

$$[\mathbf{A}; \mathbf{y}] = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T = \sum_{i=0}^{m+1} \sigma_i \mathbf{u}_i \mathbf{v}_i^T, \quad \sigma_1 \geq \dots \geq \sigma_{m+1} \geq 0.$$

The best TLS approximation $[\hat{\mathbf{A}}; \hat{\mathbf{y}}]$ of $[\mathbf{A}; \mathbf{y}]$ is given by

$$[\hat{\mathbf{A}}; \hat{\mathbf{y}}] = \mathbf{U} \hat{\mathbf{\Sigma}} \mathbf{V}^T, \quad \text{with } \hat{\mathbf{\Sigma}} = \text{diag}(\sigma_1, \dots, \sigma_m, 0)$$

and with corresponding TLS correction matrix

$$[\hat{\mathbf{E}}_A; \hat{\mathbf{e}}] = [\mathbf{A}; \mathbf{y}] - [\hat{\mathbf{A}}; \hat{\mathbf{y}}] = \sigma_{m+1} \mathbf{u}_{m+1} \mathbf{v}_{m+1}^T.$$

with MATLAB function $[\mathbf{U}, \mathbf{S}, \mathbf{V}] = \text{svd}(\mathbf{X})$ these procedures can be implemented easily.

Through the TLS solution where the errors in the design matrix \mathbf{A} are considered the remaining transformation coordinate residuals of collocated DHDN points in Baden-Württemberg are reduced from 12.4 cm to 6.2 cm, which are illustrated in figure 13 in comparison with figure 1 in detail. The statistics of these residuals and the transformation parameters in comparison with LS methods are listed in Table 5 and Table 6. The following statistical terms shows us the difference between the quadratics sums of the residuals $\hat{\mathbf{e}}_{LS}^T \hat{\mathbf{e}}_{LS}$ related LS and $\hat{\mathbf{e}}_{TLS}^T \hat{\mathbf{e}}_{TLS}$ related TLS, together with the quadratics sums of the errors of vectorized design matrix:

$$\hat{\mathbf{e}}_{LS}^T \hat{\mathbf{e}}_{LS} = 4.063234 \text{ (m}^2\text{)}$$

$$\hat{\mathbf{e}}_{TLS}^T \hat{\mathbf{e}}_{TLS} = 1.015790 \text{ (m}^2\text{)}$$

$$\hat{\mathbf{E}}_A^T \hat{\mathbf{E}}_A = 1.015808 \text{ (m}^2\text{)}$$

$$\hat{\mathbf{e}}_{TLS}^T \hat{\mathbf{e}}_{TLS} + \hat{\mathbf{E}}_A^T \hat{\mathbf{E}}_A = 2.031598 \text{ (m}^2\text{)}$$

Table 5 Numerical comparison of 7-P Helmert transformation with LS and TLS estimator

Transformation models	Collocated sites	Absolute mean Residuals (m)		Max. of absolute residuals (m)		RMS (m)	Standard deviation of unit weight (m)
		[V _N]	[V _E]	[V _N]	[V _E]		
LS	B-W 131	0.1055	0.0839	0.4211	0.3553	0.1241	0.1026
TLS	B-W 131	0.0527	0.0420	0.2105	0.1777	0.0621	0.0513

Table 6 Comparison of 7-p Helmert transformation parameters with LS and TLS

131 BWREF points	7-parameter Helmert transformation GK (DHDN) – UTM (ETRS89)						
	T_x (m)	T_y (m)	T_z (m)	α (")	β (")	γ (")	$d\lambda$ ($\times 10^{-5}$)
LS	582.901711	112.168080	405.603061	-2.255032	-0.335003	2.068369	9.117208
TLS	582.901702	112.168078	405.603051	-2.255032	-0.335003	2.068369	9.117210
TLS-LS	-0.000009	-0.000002	-0.000010	0	0	0	0.000002

Similarly, we select 10 points as checkpoints out from the 131 points. Then transformation parameters are calculated depending on the rest 121 points.

The process is summarized as follows:

- 1) Calculate transformation parameters (121 points)

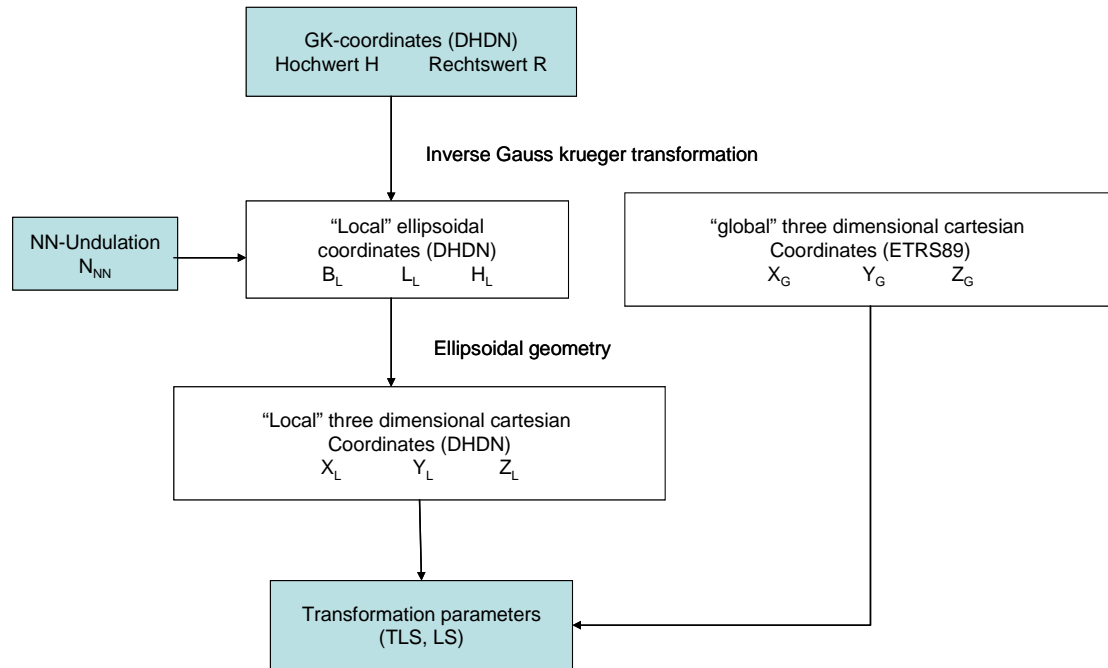


Figure 10 Calculation of the transformation parameters

- 2) Firstly, transfer the Gauss-Krüger coordinate system into X_L, Y_L, Z_L coordinate system.

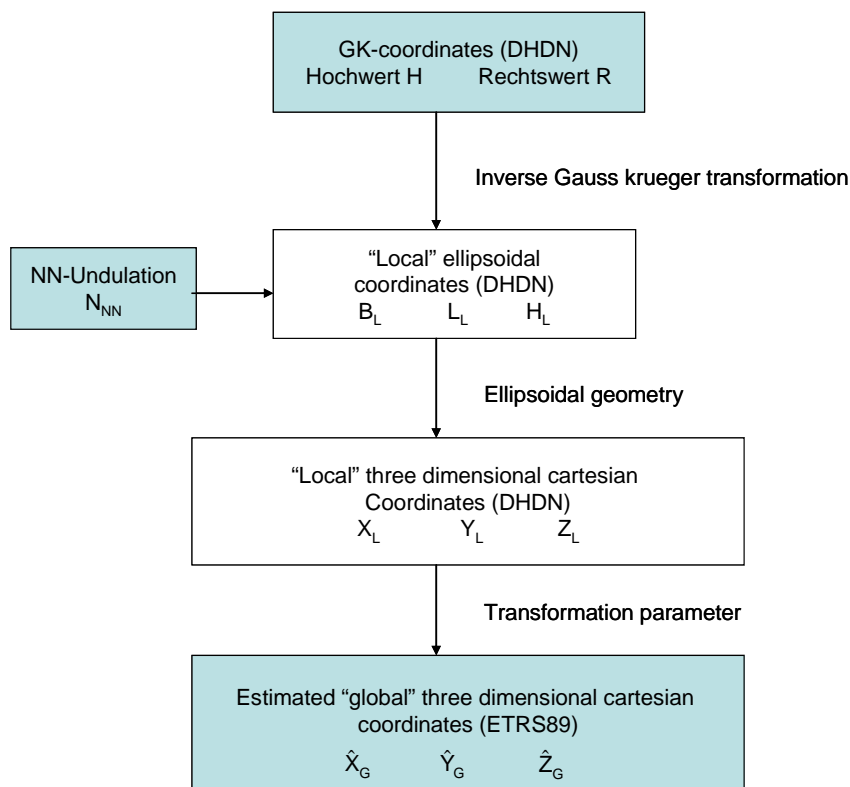


Figure 11 Calculation of the 3-D cartesian coordinates (ETRS89)

Then by using the transformation parameters, we calculate the

$$\begin{pmatrix} \hat{X}_G \\ \hat{Y}_G \\ \hat{Z}_G \end{pmatrix} = \lambda \mathbf{R}(\alpha, \beta, \gamma) \begin{pmatrix} X_L \\ Y_L \\ Z_L \end{pmatrix} + \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix}_L \quad (4.2.11)$$

by the inverse transformation we can obtain the estimated UTM coordinate system (\hat{N}, \hat{E}) (10 points)

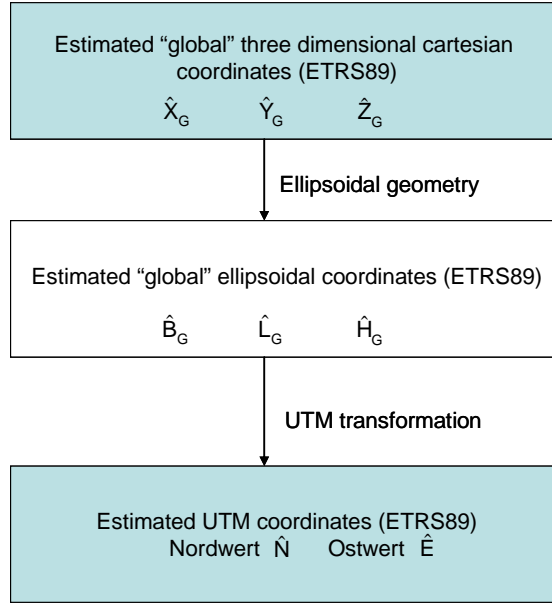


Figure 12 Calculation of the estimated UTM coordinates (ETRS89)

3) Compare with the original value (10 points)

$$\begin{bmatrix} d\hat{N} \\ d\hat{E} \end{bmatrix} = \begin{bmatrix} N - \hat{N} \\ E - \hat{E} \end{bmatrix} \quad (4.2.12)$$

The results are in table 7.

Table 7 Numerical comparison between LS and TLS

Points number	LS (m)		TLS (m)	
	$d\hat{N}$	$d\hat{E}$	$d\hat{N}$	$d\hat{E}$
652000308	-0.1353	+0.3165	-0.1353	+0.3165
672500108	-0.2147	+0.2435	-0.2147	+0.2435
701600288	-0.1503	+0.2967	-0.1503	+0.2967
722000300	-0.0860	-0.3590	-0.0860	-0.3590
722600208	+0.3374	+0.0763	+0.3374	+0.0763
751300208	+0.4139	-0.0714	+0.4139	-0.0714
761900108	+0.4164	+0.0786	+0.4164	+0.0786
792311808	-0.0801	+0.2510	-0.0801	+0.2510
811300108	+0.0184	+0.2470	+0.0184	+0.2470
811800108	+0.1787	+0.2680	+0.1787	+0.2680

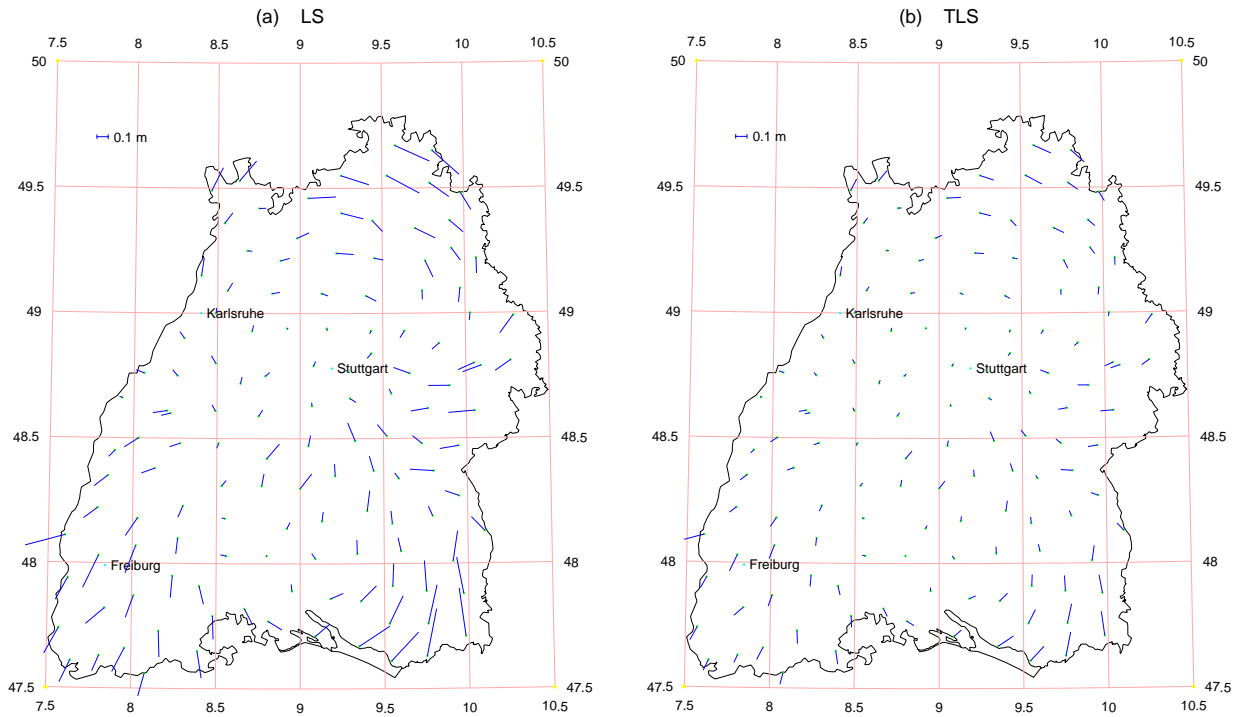


Figure 13 Horizontal residuals after 7-P Helmert transformation with TLS estimator in Baden-Württemberg Network (a) LS (b) TLS

From table 7 we can make the same conclusion as 6-parameter affine transformation. The residuals after transformation with LS and TLS method are almost identical and less than 4.5 dm. i.e., in practical situation TLS method has not obvious advantage for new points in comparison with LS method. See Figure 14. That is since that the estimated transformation parameters $\hat{\xi}_{TLS}$ and $\hat{\xi}_{LS}$ in table 6 are almost identical.

To sum up, as long as the data satisfy the assumed EIV (error-in-variables) model, TLS gives better estimates of the true model parameters ξ than does LS, regardless of the common error distribution, and should be preferred to LS. For small sample sizes, small errors, or small ξ , there is not much difference between TLS and LS: the mean squared error of the LS estimates may be slightly larger than that of the TLS estimates but, on the other hand, TLS is less robust than LS and has larger variances. However, when the sample size is increased, the TLS estimates are clearly more accurate than the LS estimates. When the data significantly violate the assumptions of the model, e.g., when outliers are present, the TLS estimates are very unstable and their accuracy is strongly affected. LS also encounters stability problems, although less dramatically. In these situations, robust procedures, which are quite efficient and rather insensitive to outliers, should be applied (Van Huffel, 1991).

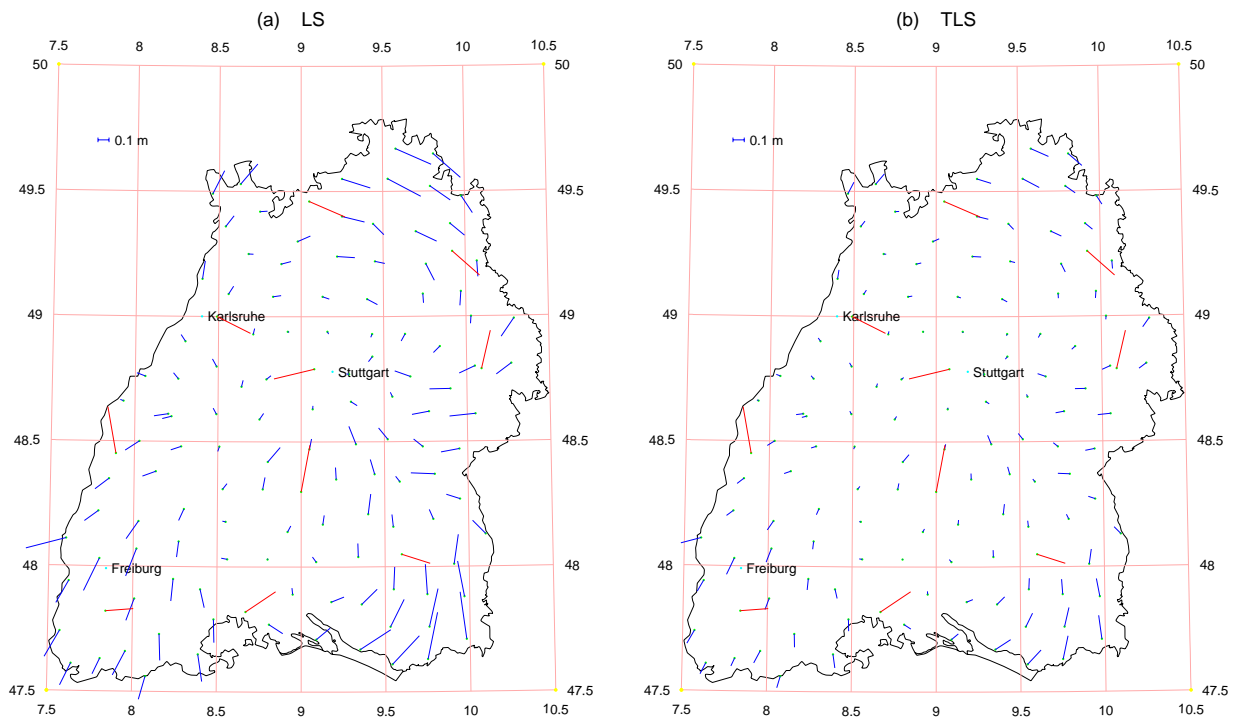


Figure 14 Horizontal residuals of 10 check points (a) LS (b) TLS

5. Conclusions and further studies

The traditional techniques used for solving the linear estimation problems are based on classical LS. Even though some robust methods based on L_1 -norm do exist, either the LS or robust estimation techniques assume that only the observation vector contains errors. However, this assumption is not valid for every case. One example for such a problem is the coordinate transformation (Acar and Özlüdemir, 2006). Considering the coordinate transformation between two coordinate sets, we can easily see that the observations and the partly the design matrix of the transformation are erroneous. One solution to this problem is the application of TLS method for estimation of the transformation parameters. TLS method considers the erroneous design matrix as well as their covariance information in computations.

For the concrete “Introduction of ETRS89 into Baden-Württemberg” has an alternative transformation procedure with the Total Least Squares method been applied in the transformation with the models of the 7-parameter Helmert transformation and 6-parameter affine transformation using the 131 collocated points, and the results have been tested and discussed here. The results have been also analyzed and compared with these results of the conventional LS method. Based on these analyses and comparisons with different estimation methods the following points can be concluded (Cai, 2006):

- TLS method is mathematically rigorous, especially for the coordinate transformations. With Van Huffel’s SVD method, TLS is simply implemented through MATLAB and allows meaningful application in data analysis. Moreover, if we use Schaffrin’s algorithm, the calculation of SVD is unnecessary.
- In our case study, the TLS solution is not differ from the standard LS solution, since the estimated transformation parameters $\hat{\xi}_{TLS}$ and $\hat{\xi}_{LS}$ almost identical. There are no significant differences between the residuals after transformation with LS and TLS method, i.e., for new points the application of TLS method should be carefully considered.
- For further studies, we should study how to apply the observation weights in TLS.

Finally, it is hoped that this study will aid and stimulate the reader to apply TLS in his or her own applications and problems.

REFERENCES

- Acar, M., Oezluedemir, M. T., Akzilmaz, O., Celik, R. N. and Ayan, T. (2006): Deformation analysis with Total Least Squares, Istanbul Technical University, Division of Geodesy, Istanbul, Turkey.
- Adcock, R. J. (1878): A problem in least squares. *The Analyst*, 5, 53-54.
- Badekas, J. (1969): Investigations related to the establishment of a world geodetic system, Report 124, Department of Geodetic Science, Ohio State University, Columbus.
- Bursa, M. (1962): The theory of the determination of the nonparallelism of the minor axis of the reference ellipsoid, polar axis of the Earth, and initial astronomical and geodetic meridians from observations of artificial Earth satellites. *Studia Geophysica et Geodaetica* 6, 209-214.
- Cai, J. (2000): The systematic analysis of the transformation between the German Geodetic Reference System (DHDN, DHHN) and the ETRF System (DREF91). *Earth Planets Space* 52, 947-952.
- Cai, J. (2006): Systematical analysis of the transformation between Gauss-Krüger-Coordinate/DHDN and UTM-Coordinate/ETRS89 in Baden-Württemberg with different estimation methods. Submitted to the Proceeding of IAG Symposium Geodetic Reference Frame 2006.
- Cai, J. (2003): Gutachten zur "Einführung von ETRS89 in Baden-Württemberg". Geodätisches Institut der Universität Stuttgart, Stuttgart 2003.
- Gleser, L. J. (1981): Estimation in a multivariate 'errors in variables' regression model: Large sample results. *Ann. Statist.* 9, 24-44.
- Golub, G. H. (1973): Some modified matrix eigenvalue problems. *SIAM Rev.* 15, 318-344.
- Golub, G. H. and Van Loan, C. F. (1980): An Analysis of the total least squares problem. *Siam J. Numer. Anal.* Vol 17, No.6., December 1980.
- Grafarend, E. (2006): *Linear and Nonlinear Models: Fixed Effects, Random Effects, and Mixed Models.* de Gruyter, Berlin; New York.
- Koopmans, T. C. (1937): *Linear Regression Analysis of Economic Time Series.* De Erven F. Bohn, Haarlem, The Netherlands.
- Levin, M. J. (1964): Estimation of a system pulse transfer function in the presence of noise. *IEEE Trans. Automat. Control* 9, 229-235.
- Madansky, A. (1959): The fitting of straight lines when both variables are subject to error. *J. Amer. Statist. Assoc.* 54, 173-205.
- Molodensky, M., Eremeev, S. and Yurkina, V. F. (1960): *Metody izucheniya vnesnego gravitatsionnogopolya I figure Zemli.* Tr. CNIIGAIK, Moscow, 131.
- Mühlich, M. and Mester, R. (1999): Subspace methods and equilibration in computer vision. Submitted to Citeseer 2000. Institute for Applied Physics, J. W. Goethe-Universität Frankfurt, Frankfurt, Germany.
- Pearson, K. (1901): On lines and planes of closest fit to points in space, *Philos. Mag.*, 2, pp. 559-572.
- Schaffrin, B. and Felus, Y. (2003): On Total Least-Squares adjustment with constraints, In Proceedings to IUGG General Ass. Department of Civil and Environmental Engineering and geodetic Science, The Ohio State University, Columbus.
- Schaffrin, B. (2003): A Note on Constrained Total Least-Squares Estimation, In *Computational Statistics and Data Analysis*, eingereicht Okt 2003.
- Schaffrin, B. (2005): Generalizing the Total Least-Squares Estimation for Empirical Coordinate Transformation, lecture in Geodetic Inst. Stuttgart.
- Strang, G. (1988): *Linear Algebra and its Applications*, 3rd edition, Harcourt Brace Jovanovich, San Diego, pp. 505.
- Van Huffel, S. and Zha, H. (1993): The Total Least-Squares problem, in *Handbook of Statistics*:

- Computational Statistics, (Rao, C. R., ed.), vol. 9, North Holland Publishing Company, Amsterdam, The Netherlands, 1993, pp. 377-408.
- Van Huffel, S. and Lemmerling, P. (2002): Total Least Squares and Errors-in-Variables Modeling, Kluwer Academic Publishers, Netherlands.
- Van Huffel, S. and Vandewalle, J. (1991): The Total Least Squares Problem Computational Aspects and Analysis, Society for Industrial and Applied Mathematics, Philadelphia.
- Veis, G. (1960): Geodetic uses of artificial satellites, Smithsonian Contributions to Astrophysics 3, 95-161.
- Wolf, H. (1963): Geometric connection and re-orientation of three-dimensional triangulation nets, Bulletin Geodesique, 68, 165-169.
- Wolf, R. P. and Ghilani, D. C. (1997): Adjustment computations: statistics and least squares in surveying and GIS, Wiley, New York, 347.
- York, D. (1966): Least squares fitting of a straight line. Canad. J. Phys. 44, 1079-1086.

Appendix I

Matlab-Program

```

1. function [t, r, k, VPV, w, A, res]=affine_TLS(N, E, H, R)
% calculate the affine transformation parameters with TLS method
%-----
% | N |      | cos(alpha) -sin(beta) || m1  0 | | H | | tN |
% |   | = |      | | | | | |
% | E |      | sin(alpha)  cos(beta) || 0  m2 | | R | | tE |
%
%          | m1 cos(alpha) -m2 sin(beta) | | H | | tN |
%   =      |      | | | | | + | |
%          | m1 sin(alpha)  m2 cos(beta) | | R | | tE |
%
%          |      a      b      | | H | | tN |
%   =      |      | | | | | + | |
%          |      g      f      | | R | | tE |
%-----
% input          : N ... northvalue in UTM system
%                E ... eastvalue in Gauss-Krueger system
%                H ... hochwert in Gauss-Krueger system
%                R ... rechtwert in Gauss-Krueger system
% output         : t ... 2 x 1-Vektor of Translation parameter
%                r ... 2 x 2-Rotationsmatrix
%                k ... 2 x 1-Vektor mit dem Mass-Staeben
%                VPV.. square residual (e*e^T+E_A* E_A^T)
%                w ... 3 x 1-Vektor mit w(1)=Rotationswinkel, w(2)=Rotationswinkel
%                w(3)=Gewichtseinheitsfehler(sigma0)
%                A ... 2 x 2-Matrix mit den Parametern a, b, g, f
%                res 2 x n-Matrix, row 1 are northvalue residuals
%                row 2 are ostvalue residuals

n=size(N, 1);
% Schwerpunktzentrierte Koordinaten
Ns=mean(N);
Es=mean(E);
Hs=mean(H);
Nq=N-Ns;
Eq=E-Es;
Hq=H-Hs;
Rq=R-fix(R/1e6)*1e6;
Rs=mean(Rq);
Rq=Rq-Rs;
% TLS method
TLY=[Nq;Eq];

```

```

for i=1:n
    TLA(i, :)= [Hq(i), Rq(i), 0, 0, Nq(i)];
    TLA(i+n, :)= [0, 0, Hq(i), Rq(i), Eq(i)];    % A Matrix
end
[UE, SE, VE]=svd(TLA); % Singulaerwertzerlegung(SVD)
[mm, nn] = size(VE);
mmm=mm-1;
TLX=-(1/VE(mn, mn))*VE(1:mmm, mm)';
TLX=TLX';
TLEA=SE(mm, mm)*UE(:, mm)*VE(:, mmm)';
[mm, nn]=size(TLEA);
TLE=TLEA(:, nn);                % e
TLEA3=[TLEA(:, 1), TLEA(:, 2), TLEA(:, 3), TLEA(:, 4)];
anzahl=2*n*4;
TLEA=reshape(TLEA3, anzahl, 1);    % EA
XX=reshape(TLX, 2, 2);
t=[Ns;Es]-XX'*[Hs;Rs] ;          % 2 x 1-Translationsvektor tH, tR

X(1, 1)=t(1);
X(1, 2)=t(2);
X(2, 1)=TLX1(1);
X(3, 1)=TLX1(2);
X(2, 2)=TLX1(3);
X(3, 2)=TLX1(4);

m1=sqrt(X(2, 1)^2+X(2, 2)^2);      % m1=sqrt(a^2+g^2): Mass-Stab m1 für die H-Achse  m1=sqrt(a^2+g^2)
m2=sqrt(X(3, 1)^2+X(3, 2)^2);      % m2=sqrt(b^2+f^2): Mass-Stab m2 für die R-Achse
k=[m1;m2];                        %                : 2 x 1-Mass-Stabsvektor m1, m2
w(1)= atan2(X(2, 2), X(2, 1));      % w(1)= atan(g/a) : Rotationswinkel [rad] für die H-Achse
w(2)=-atan2(X(3, 1), X(3, 2));      % w(2)=-atan(b/f) : Rotationswinkel [rad] für die R-Achse
r(1, 1)=X(2, 1)/m1;                % =a/m1    2 x 2-Rotationsmatrix
r(1, 2)=X(3, 1)/m2;                % =b/m2
r(2, 1)=X(2, 2)/m1;                % =g/m1
r(2, 2)=X(3, 2)/m2;                % =f/m2
A=X;                                % Koeffizientenmatrix
if n > 3
    res1=reshape(TLE, n, 2);
    res=res1';                      % das ist nun eine 2 x n-Matrix, die zurück gegeben werden kan
                                    % Zeile 1 sind Nord- bzw. Hochwertresiduen
                                    % Zeile 2 sind Ost- bzw. Rechtswertresiduen
VPV=TLE'*TLE+TLEA'*TLEA;
w(3)=sqrt(VPV/(2*n-6));
else
    w(3)=NaN;

```

```

    res=zeros(2,n);
end

2. function [tx, rotx, massstabx, winkelx, resx, VPV, SIGMAo]=TrafPara_TLS(xg, yg, zg, xl, yl, zl)
% 7 parameter Hermelt transformation with TLS method
%-----
%          | XG |          | XL | | TX |
%          | YG | = Scalar * Rot * | YL | + | TY |
%          | ZG |          | ZL | | TZ |L
%-----
% input          :   xg, yg, zg   global three dimensional cartesian coordinates XG YG ZG..
%                :   xl, yl, zl   local three dimensional cartesian coordinates XL YL ZL
% output         :   tx           3X1 TranslationsArray
%                :   rotx        3X3 Rotations matrix [rad]
%                :   massstabx   same Massstab for three axeses (Scalar)
%                :   winkelx     3X1 rotation angles alfa, betta, gama [Rad]
%                :   resx        nX3 (n=number of points). Residuals of control points
%                               vx, vy, vz [m]
%                :   VPV         square residual (e*eT+eA* eAT)
%                :   SIGMAo     standard deviation of unit weight [m]

n=max(size(xg));
n1=max(size(xl));
if n~=n1
    error('Anzahl der lokalen Punkte und globalen Punkte nicht gleich!')
end
[x, SIGMAo, VPV, e, ea]=Helmut_TLS(xg, yg, zg, xl, yl, zl, n);

tx=[x(1), x(2), x(3)]';
rot=[0, -x(6), x(5);
     x(6), 0, -x(4);
     -x(5), x(4), 0];

rotx=-rot+eye(3). *x(7)+eye(3);          % 3X3 Rotationsmatrix [rad]
winkelx=[x(4), x(5), x(6)]';           % winkelx
massstabx=x(7);                         % massstabx
dummy=reshape(e, n, 3);
resx=dummy;                             % resx

```

```

3. function [x, SIGMA0, VPV, TLE, TLEA]=Helmut_TLS(xg, yg, zg, xl, yl, zl, n)
% sub-function for 7 parameter Hermelt transformation with TLS method
%-----
%           y=Ax                y      observation
%                               A      disign matrix
%                               x      transformation parameters
%           [UE, SE, VE]=svd(TLA)    singular value decomposition.
%                               TLA    design matrix with TLS method TLA=[A;y]
%                               UE     left singular vector
%                               VE     right singular vector
%                               SE     singular value matrix
%-----

% input           :   xg, yg, zg   global three dimensional cartesian coordinates XG YG ZG..
%                 :   xl, yl, zl   local three dimensional cartesian coordinates XL YL ZL
%                 :   n            number of points
% output          :   x            7 transformation parameters : 3 translation, 3 rotation,
%                               1 scale
%                 :   SIGMA0      standard deviation of unit weight [m]
%                 :   VPV         square residual (e*eT+EA* EAT)
%                 :   TLE         residuals of observation y

L=[xg;yg;zg];
%Schwerpunktzentrierung
R=[xg yg zg xl yl zl];
R1=mean(R);
TLL=[xg-R1(1);yg-R1(2);zg-R1(3)];;
%design matrix
zlq=zl-R1(6);
ylq=yl-R1(5);
xlq=xl-R1(4);
% TLS method
for i=1:n
    TLA(i, :)= [0, -zlq(i), ylq(i), xlq(i), TLL(i)];
    TLA(i+n, :)= [zlq(i), 0, -xlq(i), ylq(i), TLL(i+n)];
    TLA(i+2*n, :)= [-ylq(i), xlq(i), 0, zlq(i), TLL(i+2*n)];           % A Matrix
end
[UE, SE, VE]=svd(TLA);           % Singulaerwertzerlegung(SVD)
[mm, nn] = size(VE);
mmm=mm-1;
TLX=-(1/VE(mn, mn))*VE(1:mmm, mm)';
TLX=TLX';
TLEA=SE(mn, mn)*UE(:, mn)*VE(:, mn)';
[mm, nn] = size(TLEA);

```



```

TLE=TLEA(:, 5); % e
TLEA3=[TLEA(:, 1), TLEA(:, 2), TLEA(:, 3), TLEA(:, 4)];
anzahl=3*n*4;
TLEA=reshape(TLEA3, anzahl, 1); % EA
TLXX=[TLX(4), TLX(3), -TLX(2); -TLX(3), TLX(4), TLX(1); TLX(2), -TLX(1), TLX(4)];
TLT=[R1(1); R1(2); R1(3)]-TLXX*[R1(4); R1(5); R1(6)]; % 3 x 1-Translationsvektor
VPV=TLE'*TLE+TLEA'*TLEA;
SIGMAo=sqrt(VPV/(3*n-7));
x=[TLT(1); TLT(2); TLT(3); TLX(1); TLX(2); TLX(3); (TLX(4)-1)];

```

Appendix II

Data output

```
%-----
% residuals:      vv(e) = eTe; vv(ea) = eATeA; vv(e, ea) = VPV = eTe + eATeA; vv(e + ea) = (e + eA)T(e + eA)
%-----
```

1. Transformationsparameter des Gesamtnetzes aus Affine2D (6-P 121 Punkte mit LS)

```
=====
```

```
tN      : 437.389561 [m]
tE      : 119.889076 [m]
Alpha   : 0.159846 ["]
Beta    : -0.207380 ["]
massstab1: 0.9996002792
massstab2: 0.9996011784
Sigma0(e) : 0.120751 [m]
vv(e)    : 3.441096 [m]
```

Statistische Angaben Ueber die 121 Passpunkte nach der Transformation

```
=====
```

Feld	Punkte	Abs.Mittel (m)		Abs.Max (m)		QMR (m)	Sigma0 (m)
		[VN]	[VE]	max[VN]	max[VE]		
	121	0.105315	0.080063	0.334795	0.320312	119492	0.120751

2. Transformationsparameter des Gesamtnetzes aus Affine2D (6-P 121 Punkte mit TLS)

```
=====
```

```
tN      : 437.389548 [m]
tE      : 119.889079 [m]
Alpha   : 0.159846 ["]
Beta    : -0.207380 ["]
massstab1: 0.9996002792
massstab2: 0.9996011784
Sigma0(e) : 0.040272 [m]
vv(e)    : 0.382751 [m]
vv(ea)   : 0.764891 [m]
vv(e, ea) : 1.147643 [m]
vv((e+ea)) : 1.530394 [m]
```

Statistische Angaben Ueber die 121 Passpunkte nach der Transformation

Feld	Punkte	Abs.Mittel (m)		Abs.Max (m)		QMR (m)	Sigma0 (m)
		[VN]	[VE]	max[VN]	max[VE]		
	121	0.035124	0.026702	0.111658	0.106828	0.039852	0.040272

3. Transformationsparameter des Gesamtnetzes aus Affine2D (6-P 131 Punkte mit LS)

tN	:	437.194567 [m]
tE	:	119.756709 [m]
Alpha	:	0.165368 ["]
Beta	:	-0.196455 ["]
massstab1:		0.9996003203
massstab2:		0.9996011570
Sigma0(e)	:	0.119868 [m]
vv(e)	:	3.678308 [m]

Statistische Angaben Ueber die 131 Passpunkte nach der Transformation

Feld	Punkte	Abs.Mittel (m)		Abs.Max (m)		QMR (m)	Sigma0 (m)
		[VN]	[VE]	max[VN]	max[VE]		
	131	0.104860	0.080369	0.328825	0.322622	0.118715	0.119868

4. Transformationsparameter des Gesamtnetzes aus Affine2D (6-P 131 Punkte mit TLS)

tN	:	437.194554 [m]
tE	:	119.756712 [m]
Alpha	:	0.165368 ["]
Beta	:	-0.196455 ["]
massstab1:		0.9996003203
massstab2:		0.9996011570
Sigma0(e)	:	0.039977 [m]
vv(e)	:	0.409136 [m]
vv(ea)	:	0.817619 [m]
vv(e, ea)	:	1.226756 [m]
vv((e+ea))	:	1.635892 [m]

Statistische Angaben Ueber die 131 Passpunkte nach der Transformation

Feld	Punkte	Abs. Mittel (m)		Abs. Max (m)		QMR (m)	Sigma0 (m)
		[VN]	[VE]	max[VN]	max[VE]		
	131	0.034972	0.026804	0.109667	0.107598	0.039593	0.039977

5. Transformationsparameter des Gesamtnetzes aus Helmert3D (7-P 121 Punkte mit LS)

Tx	:	582.953910 [m]
Ty	:	112.260037 [m]
Tz	:	405.722189 [m]
Alpha	:	-2.260711 ["]
Beta	:	-0.336845 ["]
Gamma	:	2.065813 ["]
Massstab:		9.096290e-006
Sigma0(e)	:	0.103619 [m]
vv(e)	:	3.822341 [m]

Statistische Angaben Ueber die 121 Passpunkte nach der Transformation

Feld	Punkte	Abs. Mittel (m)		Abs. Max (m)		QMR (m)	Sigma0 (m)
		[VN]	[VE]	max[VN]	max[VE]		
	121	0.106692	0.083776	0.415881	0.357934	0.125309	0.103619

6. Transformationsparameter des Gesamtnetzes aus Helmert3D (7-P 121 Punkte mit TLS)

Tx	:	582.953900 [m]
Ty	:	112.260036 [m]
Tz	:	405.722178 [m]
Alpha	:	-2.260711 ["]
Beta	:	-0.336845 ["]
Gamma	:	2.065813 ["]
Massstab:		9.096293e-006
Sigma0(e)	:	0.051809 [m]
vv(e)	:	0.955568 [m]
vv(ea)	:	0.955585 [m]
vv(e, ea)	:	1.911153 [m]
vv((e+ea))	:	3.822306 [m]

Statistische Angaben Ueber die 121 Passpunkte nach der Transformation

Feld	Punkte	Abs.Mittel (m)		Abs.Max (m)		QMR (m)	Sigma0 (m)
		[VN]	[VE]	max[VN]	max[VE]		
	121	0.053345	0.041887	0.207939	0.178965	0.062654	0.051809

7. Transformationsparameter des Gesamtnetzes aus Helmert3D (7-P 131 Punkte mit LS)

Tx : 582.901711 [m]
 Ty : 112.168080 [m]
 Tz : 405.603061 [m]
 Alpha : -2.255032 ["]
 Beta : -0.335003 ["]
 Gamma : 2.068369 ["]
 Massstab: 9.117208e-006

Sigma0(e) : 0.102599 [m]
 vv(e) : 4.063234 [m]

Statistische Angaben Ueber die 131 Passpunkte nach der Transformation

Feld	Punkte	Abs.Mittel (m)		Abs.Max (m)		QMR (m)	Sigma0 (m)
		[VN]	[VE]	max[VN]	max[VE]		
	131	0.105497	0.083931	0.421063	0.355342	0.124101	0.102599

8. Transformationsparameter des Gesamtnetzes aus Helmert3D (7-P 131 Punkte mit TLS)

Tx : 582.901702 [m]
 Ty : 112.168078 [m]
 Tz : 405.603051 [m]
 Alpha : -2.255032 ["]
 Beta : -0.335003 ["]
 Gamma : 2.068369 ["]
 Massstab: 9.117210e-006

Sigma0(e) : 0.051299 [m]
 vv(e) : 1.015790 [m]
 vv(ea) : 1.015808 [m]
 vv(e, ea) : 2.031598 [m]
 vv((e+ea)) : 4.063197 [m]

Statistische Angaben Ueber die 131 Passpunkte nach der Transformation

Feld	Punkte	Abs. Mittel (m)		Abs. Max (m)		QMR (m)	Sigma0 (m)
		[VN]	[VE]	max[VN]	max[VE]		
	131	0.052748	0.041965	0.210530	0.177669	0.062050	0.051299