WISSENSCHAFTLICHE ARBEITEN DER FACHRICHTUNG
GEODÄSIE UND GEOINFORMATIK DER LEIBNIZ UNIVERSITÄT HANNOVER
ISSN 0174-1454

Nr. 294

Xing Fang

Weighted Total Least Squares Solutions for Applications in Geodesy

HANNOVER 2011

WISSENSCHAFTLICHE ARBEITEN DER FACHRICHTUNG
GEODÄSIE UND GEOINFORMATIK DER LEIBNIZ UNIVERSITÄT HANNOVER
ISSN 0174-1454

Nr. 294

Weighted Total Least Squares Solutions for Applications in Geodesy

Von der Fakultät für Bauingenieurwesen und Geodäsie
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung der Grades

DOKTOR-INGENIEUR

genehmigte Dissertation

von

Dipl.-Ing Xing Fang

geboren am 26.Nov.1981, in Hubei VR China

HANNOVER 2011

Vorsitzender der Prüfungskommission:	Prof. Dr.-Ing Christian Heipke, Leibniz Universität Hannover
Referent:	Prof. Dr.-Ing Hansjörg Kutterer, Leibniz Universität Hannover
Korreferent:	Prof. Dr.-Ing Wolf-Dieter Schuh, Universität Bonn
	Prof. Dr.-Ing Steffen Schön, Leibniz Universität Hannover
Tag der mündlichen Prüfung:	27. Apr. 2011

Schlagworte: Total Least Squares; nichtlineares Modell; iterative Algorithmen

Keywords: total least-squares; nonlinear model; iterative algorithms

Contents

Abstract

Although data processing in geodetic applications often relies on the least-squares method, the Gauss-Markov model with uncertain model matrix has to be solved rigorously using the total least-squares (TLS) technique. Recently, a large number of reports have been published to adjust the errors-in-variables model. However, the general solutions and the computational advantages of the TLS problem are mostly unknown in various scientific domains.

In this contribution the auxiliary Lagrange multipliers are used to give some solutions of the TLS problem, where the variance covariance matrix of the extended observation vector is considered as a fully populated matrix in the adjustment. In contrast to solving the problem using the nonlinear Gauss-Helmert model, the solutions proposed in this thesis do not require any linearization. Furthermore, it is widely agreed that the method of Lagrange multipliers or the nonlinear Gauss-Helmert model (implicitly using Lagrange multipliers) yield only necessary conditions for optimality in the constrained problems. However, the second derivative of the objective function with respect to the parameter vector representing the sufficient condition of the optimization is reasonable to be presented. Based on the aforementioned second derivative the Newton algorithm is designed for the optimization problem. In contrast to the Gauss-Newton algorithm, which is popularly applied for the weighted TLS problem, the Newton algorithm works more efficiently in the final stage. In addition, the Newton or the Gauss Newton method can be modified via the combination with, for example the steepest descent method obtained by the first derivative.

After given the theoretical development of the fully weighted TLS problem, some extensions are presented. The weighted TLS problem with fixing columns is taken into account, where the model matrix with fixing columns in the weighted TLS problem can be separated into the deterministic and stochastic parts. In this case, the parameters corresponding to the fixed columns are eliminated based on the normal equation system. In the more general case, fixing elements is also solved by means of the non-linear iterative Gauss-Helmert model. Moreover, Lagrange multipliers are applied to solve the constrained weighted TLS problems and the weighted TLS problem, in which the parameters and the conventional observations are expressed matrix-wise instead of vector-wise. Undoubtedly, the iterative Gauss-Helmert model method can solve a lot of non-linear TLS problems due to its simplicity. However, it should be generalized for the weighted TLS problem by integrating the nonlinear constraints of parameters and the observation equation simultaneously.

Based on the solutions discussed in previous chapters, some geodetic applications are demonstrated. The purpose of the orthogonal regression is to show the solutions leading to identical results. The performances of the solutions are compared with current methods with respect to the convergence behavior and the weight information. In addition, the 3D similarity transformation considering the errors in the model matrix is solved by Gauss Newton method in this study. At the later part the weighted TLS solution is investigated in the quadratic form analysis and the free stationing with stochastic parameter within geodetic networks with the weak datum.

Kurzfassung

Obwohl die Datenverarbeitung in geodätischen Anwendungen häufig auf der Methode der kleinsten Quadrate basiert, ist es nötig, das Gauß-Markov-Modell mit unsicherer Modellmatrix mit der Total Least Squares (TLS) Technik zu lösen. Viele Lösungen für die Berechnung der TLS wurden in letzter Zeit in verschiedenen wissenschaftlichen Bereichen publiziert, die allgemeinen Lösungen und die numerische Vorteile müssen allerdings noch untersucht werden.

In diesem Beitrag werden die Lagrange-Multiplikatoren als Hilfsmittel eingesetzt, um einige Lösungen des TLS Problems zu präsentieren, bei dem die Kovarianzmatrix des vollständigen Beobachtungsvektors als vollbesetzte Matrix gegeben ist. Im Gegensatz zur Lösung mit Hilfe des nichtlinearen Gauss-Helmert-Modell erfordern die vorgeschlagenen Lösungen keine Linearisierung. Im Weiteren ist bekannt, dass die Methode der Lagrange-Multiplikatoren oder des nichtlinearen Gauss-Helmert-Modells (implizit mit Lagrange-Multiplikatoren) nur die notwendige Bedingung für Optimalität im vorliegenden Restriktionsproblem liefern. Deshalb ist es sinnvoll, die zweite Ableitung der Zielfunktion in Bezug auf den Parametervektor herzuleiten, weil sie die hinreichende Bedingung für die Optimierung repräsentiert. Basierend auf der oben erwähnten zweiten Ableitung ist der Newton Algorithmus für das Optimierungsproblem zu entwickeln. Im Gegensatz zum Gauss-Newton-Algorithmus, der für das (gewichtete) TLS Problem am häufigsten angewendet wird, konvergiert der Newton Algorithmus effizienter in der Endphase. Darüber hinaus kann das Newton- oder Gauß-Newton-Verfahren durch die Kombination mit z.B. der steilsten Abstiegsmethode, die durch die erste Ableitung erhalten wird, modifiziert werden.

Nach der theoretischen Entwicklung des voll gewichteten TLS-Problems werden ein paar Erweiterungen vorgestellt. Das gewichtete TLS-Problem, bei dem die Spalten festgelegt sind, ist zu berücksichtigen. In dem Fall besteht die Modellmatrix aus einem deterministischen und stochastischen Anteil. Die Parameter, die den fixierten Spalten entsprechen, können durch das Normalgleichungssystem eliminiert werden. Im allgemeineren Fall wird das Problem mit den festgehaltenen Elementen auch durch das nichtlineare Gauss-Helmert-Modell gelöst. Darüber hinaus werden die Lagrange-Multiplikatoren angewendet, um das gewichtete TLS-Problem mit der linearen Restriktion und das gewichtete TLS Problem, in dem die Parameter und die herkömmlichen Beobachtungen in Matrixweise anstatt in Vektorweise dargestellt sind, zu lösen. Das iterative Gauß-Helmert-Modell Verfahren kann offensichtlich eine Menge von nicht-linearen TLS Problemen lösen. Es sollte jedoch für das TLS-Problem der gleichzeitigen Integration der nichtlinearen Nebenbedingungen der Parameter und der Beobachtungsgleichung verallgemeinert werden.

Basierend auf der theoretischen Entwicklung werden einige geodätische Anwendungen dargestellt. Der Zweck der orthogonalen Regression ist zu prüfen, ob die Lösungen die gleichen Ergebnisse liefern. Das Verhalten der Lösungen wird mit den aktuellen Methoden in Bezug auf das Konvergenzverhalten und die Gewichtsinformationen verglichen. Weiterhin wird die 3D Ähnlichkeits-Transformation unter Berücksichtigung der Fehler in der Modellmatrix durch das Gauss-Newton-Verfahren gelöst. Bei Letzterem wird die gewichtete TLS Lösung auf die Analyse quadratischer Formen und die freie Stationierung mit stochastischen Parametern innerhalb geodätischer Netze mit schwachem Datum angewendet.

1 Introduction

1.1 Background

The traditional geodetic task consists of measuring and representing the earth mostly through a mathematical model. Based on sufficient observations the unknown parameters of the mathematical model can be estimated. The method of least-squares (LS), which has been developed by C.F. Gauss and A.M. Legendre in the nineteenth century, is applied to approximate solutions of these overdetermined systems. These overdetermined systems can be usually expressed as $\mathbf{y}+\mathbf{v}_y=\mathbf{A}\xi$, where the traditional observation vector $\mathbf{y}$ is affected by errors having the corresponding residual vector $\mathbf{v}_y$, and ξ is an unknown parameter vector.

However, the principal hypothesis of the certain model matrix $\mathbf{A}$ is not necessarily fulfilled in geodetic applications. A type of models with an uncertain model matrix is known in the literature (e.g. Gleser 1981) as errors-in-variables (EIV) models. The problem was studied already in Adcock 1877, and has been rediscovered many times independently in statistics (see Markovsky and van Huffel 2007). In 1980 the total least-squares (TLS) was introduced to adjust the EIV model by Golub and van Loan in the field of numerical analysis. Nowadays the terminology TLS has been widely used as a standard technique of the estimation method for the EIV model in many fields in science and engineering.

The classical TLS method is to find the solution for an overdetermined system of equation $\mathbf{y}+\mathbf{v}_y=(\mathbf{A}+\mathbf{V}_A)\xi$, where the matrix $\mathbf{A}$ is also affected by errors. The residuals $\mathbf{V}_A$ and $\mathbf{v}_y$ have an independent and identical distribution in this case. The unweighted TLS can be alternatively formulated as a problem of low rank matrix approximation (e.g. Kupferer 2005). This problem usually has a unique solution, which can be obtained in a closed form in terms of the singular value decomposition (SVD) of the data matrix (e.g., Golub and van Loan 1980, Van Huffel and Vandewalle 1989). More general problems taking different errors size and correlation into consideration have been discussed by many authors. In the generalized TLS estimator, the residual matrix $[\mathbf{V}_A, \mathbf{v}_y]$ are assumed to be row-wise independent and correlated within the rows with identical variance covariance matrix (vcm). However, the so-called generalized TLS (GTLS) does not refer to a general vcm. Some research groups (e.g. Schaffrin and Wieser 2008) call it 'equilibrated TLS' in order to avoid confusions. The solution for the so-called generalized TLS using a three-step algorithm can be found in Van Huffel and Vandewalle (1989). Further generalization where elements of the residual matrix $[\mathbf{V}_A, \mathbf{v}_y]$ are independent, but not identically distributed with element-wise different variances is called as the element-wise-weighted TLS (EW-TLS). Under this circumstance, solving the EIV model in this case has been proved to be a non-convex optimization problem, and there is no analytical-form solution (cf. Markovsky et al. 2006).

Recently, the investigation about the TLS estimation has been shown in quite a number of publications in geodesy. From the methodological point of view, the most frequently used approach, which rigorously adjusts the EIV model, is the closed form solution in terms of the SVD of the data matrix (e.g., Teunissen 1988, Felus 2004, Akyilmaz 2007, Schaffrin and Felus 2008). Another kind of methods uses auxiliary Lagrange multipliers to rearrange the TLS problem as a constrained minimization optimization, where the covariance matrix of the observations including the conventional observations and the observation in the model matrix is fairly general (e.g., Felus and Burtch 2009, Schaffrin and Felus 2008, Schaffin and Wieser 2008, Schaffrin and Wieser 2009). Furthermore, avoiding pitfalls the non-linear Gauss Helmert model (GHM) method proposed by Pope (1972) can solve the TLS and weighted TLS problem without any limitation of the vcm. This proves the TLS adjustment not referring to a new adjustment method but the adjustment for the model containing the uncertain model matrix (see, Neitzel and Petrovic 2008, Neitzel 2010, Schaffrin and Snow 2010). However, in order to hold the consistency we will still use the term TLS throughout the thesis as a synonym for the LS estimation of the parameters in the EIV model. From the theoretical point of view: The TLS problem (e.g., Kupferer 2006), the multivariate TLS problem (Schaffrin and Felus 2008), weighted TLS problem (e.g., Neitzel 2010), the regularized TLS problem (e.g., Schaffrin and Snow 2010) and the constrained TLS (Schaffrin and Felus 2009) has been discussed.

1.2 Scope and outlines of the thesis

The special structure of the weight matrix for the vector $vec([\mathbf{A} \quad \mathbf{y}])$ (vec‘ denotes the operator that stacks one column of a matrix underneath the previous one) results in special weighted TLS problems. The following Figure 1.1 shows a hierarchical classification of the problems with various weighted matrices. The weighted TLS problem with a fully populated vcm is called as the WTLS Problem, which is in the top hierarchy of the figure.

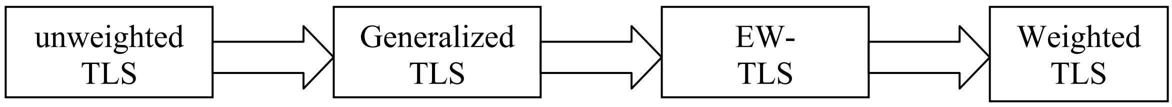


Figure 1.1 Hierarchy of weighted TLS Problems according to the structure of the weight matrix

Such problems as an orthogonal regression, the EIV model or the TLS have been discussed over a hundred years in different research fields and occurred in quite a number of publications. However, the solution and computational advantages of the WTLS problem is not widely discussed (see Chapter 4 in detail). Thus, it makes sense that this thesis elaborates general solutions for the WTLS problem. The solutions based on different principles are rigorously presented, and the advantages of solutions in comparison to the existing methods are discussed. Furthermore, extensions and applications of the WTLS solution are also demonstrated in the thesis.

The rest of the thesis is organized as follows: In Chapter 2 the estimation properties, which provide qualitatively good estimates of unknown parameter, are explained. The Gauss-Markov model (GMM) as well as the GHM and the popular estimations for this linear model, e.g. Best Linear Unbiased Estimation (BLUE) and LS estimation, are briefly presented. Chapter 3 gives an overview of the classical TLS, which has a closed form expression in terms of the SVD. We then extend the classical TLS to fixing columns, constrained TLS, structured TLS, generalized and element-wise TLS. In Chapter 4 the WTLS problem is defined and solved based on different principles, such as, the Lagrange multipliers, the iterative non-linear GHM method, Newton, Gauss-Newton and steepest descend methods. Meanwhile, the sufficient conditions for the optimization problem are given based on the second derivative of the objective function w.r.t. the parameter vector. In Chapter 5, the extensions such as fixing column, fixing elements, and the linear constrained problem are studied. In addition, an integrated nonlinear EIV model (nonlinear GHM plus nonlinear constraints of parameters and condition equations) is presented and solved by the Gauss-Newton algorithm. Chapter 6 presents some numerical geodetic applications, which are solved by the method proposed in the thesis. Finally, Chapter 7 concludes this thesis and presents the recommendations for further works.

2 Basic Knowledge for Parameter Estimation

One important task of geodesy is to estimate the unknown parameters of mathematical model representing real objects in a statistical sense. In this chapter we firstly review the theory of LS estimation technique in inconsistent linear models where the inconsistency is caused by errors in the observations. Considering the parameter vector is required to be estimated from the observation vector, the essential estimation problem is therefore to find a function $\mathbf{s}(\mathbf{y})$. The function $\mathbf{s}(\mathbf{y})$ called an estimator of $\mathbf{h}(\xi)$ is a random function, while the estimate of $\mathbf{h}(\xi)$ is a realized value of the estimator and thus a deterministic vector. After choosing a proper function $\mathbf{s}(\mathbf{y})$, all results can be obtained via the normal equations system. Moreover, the solutions for non-linear models (non-linear GMM and GHM) will be mentioned. The fundamental concept of the estimation methods presented through the whole chapter can be also primarily found in Koch (1999).

2.1 Optimal Properties of Estimation

Since $\mathbf{s}(\mathbf{y})$ depends on the chosen function $\mathbf{s}$, one must set criteria for the optimal estimation function. In the following we list some desirable properties for the function $\mathbf{s}(\mathbf{y})$:

Unbiasedness

The estimator $\mathbf{s}(\mathbf{y})$ is defined as an unbiased estimator of $\mathbf{h}(\xi)$ only if the mathematical expectation of the estimation error expressed as the difference $\mathbf{s}(\mathbf{y})-\mathbf{h}(\xi)$ is zero. An unbiased estimator is therefore

$$E\left(\mathbf{s}(\mathbf{y})-\mathbf{h}(\xi)\right)=\mathbf{0}, \tag{2.1}$$

where E denotes the expectation operator. Here and later on the boldface letters represent the vectors and matrices, the non-boldface letters stand for variables.

Mean square error

The second moment of estimation error $E\left(\left\|\mathbf{s}(\mathbf{y})-\mathbf{h}(\xi)\right\|^2\right)$ is the mean square error of the estimation. This should be expected to be the smallest quantity which corresponds to the minimum variance representing the best estimator in the absence of bias.

Best unbiased estimator

Hence, we obtain a best unbiased estimator if

$$E\left(\mathbf{s}(\mathbf{y})-\mathbf{h}(\xi)\right)=\mathbf{0} \quad \text{and} \quad E\left(\left\|\mathbf{s}(\mathbf{y})-\mathbf{h}(\xi)\right\|^2\right)\to \min . \tag{2.2}$$

If the parameter vector ξ is not stochastic, we can express this property as follows

$$E\left(\mathbf{s}(\mathbf{y})\right)=\mathbf{h}(\xi) \quad \text{and} \quad E\left(\left\|\mathbf{s}(\mathbf{y})\right\|^2\right)\rightarrow \min \tag{2.3}$$

or

$$E\left(\mathbf{h}\left(\hat{\xi}\right)\right)=\mathbf{h}(\xi) \quad \text{and} \quad E\left(\left\|\mathbf{h}\left(\hat{\xi}\right)\right\|^2\right)\rightarrow \min . \tag{2.4}$$

Both properties are criteria for measuring closeness to the true value according to the first and second moment for the distribution of $\mathbf{h}\left(\hat{\xi}\right)$. If such an unbiased estimator with minimum variance exists for all the parameters, it is called a uniformly best unbiased estimator.

Note that the other properties such as consistency, resistance, robust and sufficiency are not described in the thesis.

2.2 Gauss Markov Model

After introducing the criteria for the optimal properties, we present the GMM which is defined as

$$\begin{aligned} E(\mathbf{y}) &= \mathbf{A}\xi \quad \text{or} \quad \mathbf{y}+\mathbf{v}_y = \mathbf{A}\xi \\ \mathbf{\Sigma}_{yy} &= D(\mathbf{y}) = \sigma_0^2\mathbf{Q}_{yy} = \sigma_0^2\mathbf{P}_{yy}^{-1} \end{aligned} \tag{2.5}$$

where $\mathbf{A}$ is a $n\times u$ deterministic model matrix with a full column rank, i.e. $rank(\mathbf{A})=u$, $\mathbf{\Sigma}_{yy}$ is the vcm of the observation vector $\mathbf{y}$. The difference $\mathbf{v}_y=\mathbf{A}\xi-\mathbf{y}$ is called the residual vector and also the vector of corrections in the least-squares adjustment. The symbol D denotes the dispersion operator. $\mathbf{Q}_{yy}$ and $\mathbf{P}_{yy}$ are the $n\times n$ positive definite cofactor matrix and the weight matrix according to the observation vector $\mathbf{y}$, respectively. The properties of a positive definite matrix can be found in Koch (1999). σ_0^2 is the unknown variance component in the GMM.

In some situations there is no linear relationship between parameters and observations. However, in this case one can connect the expectation of the observation vector $\mathbf{y}$ and the parameter vector ξ through a non-linear, differentiable equation.

$$E(\mathbf{y})=\mathbf{f}(\xi) \quad \text{or} \quad \mathbf{y}+\mathbf{v}_y=\mathbf{f}\left(\xi^0+d\xi\right), \tag{2.6}$$

where ξ^0 is the initial value of the parameter. $d\xi=\xi-\xi^0$ denotes the increment of the parameter vector.

The model matrix can be obtained by the partial derivative of the function $\mathbf{f}$ with respect to ξ as follows

$$\mathbf{A}^0=\left.\frac{\partial\mathbf{f}}{\partial\xi^T}\right|_{\xi^0}. \tag{2.7}$$

The model is linearized at the position ξ^0 and expressed as

$$\mathbf{y}+\mathbf{v}_y=\mathbf{f}\left(\xi^0\right)+\mathbf{A}^0 d\xi, \tag{2.8}$$

The reduced observation vector $\mathbf{y}-\mathbf{f}\left(\xi^0\right)$ is often called observed minus computed (O-C) in geodetic applications (e.g., Xu 2007). The LS solution of the non-linear GMM may be iteratively processed and has been explained in detail in Lenzmann and Lenzmann (2007). Some standard estimation techniques presented in this part are utilized in the linear (or linearized) GMM to give out the estimated parameter vector and other variables. In addition, the corresponding cofactor matrices and the unbiased estimator of the unit of weight are also shown.

Least-squares estimation

The weighted LS solution in the linear or linearized GMM is defined as follows

$$\hat{\xi} = \arg\min_{\xi} \left(\mathbf{A}\xi - \mathbf{y}\right)^T \mathbf{P}_{yy} \left(\mathbf{A}\xi - \mathbf{y}\right). \tag{2.9}$$

Through a partial derivative one can have the normal equation as follows

$$\mathbf{A}^T \mathbf{P}_{yy} \left(\mathbf{A}\hat{\xi} - \mathbf{y}\right) = \mathbf{0}. \tag{2.10}$$

Then, we list the vectors of the estimated parameters, observations and residuals as follows

$$\hat{\xi} = \left(\mathbf{A}^T \mathbf{P}_{yy} \mathbf{A}\right)^{-1} \mathbf{A}^T \mathbf{P}_{yy} \mathbf{y}, \tag{2.11}$$

$$\hat{\mathbf{y}} = \mathbf{R}\mathbf{y}, \tag{2.12}$$

$$\hat{\mathbf{v}}_y = -\left(\mathbf{I}_n - \mathbf{R}\right)\mathbf{y}, \tag{2.13}$$

where the matrix $\mathbf{R}$ is known as the 'hat matrix' since it projects the observation vector $\mathbf{y}$ into the vector of adjusted observation $\hat{\mathbf{y}}$ (Schaffrin 1997). The matrix $\mathbf{I}_n - \mathbf{R} = \mathbf{I}_n - \mathbf{A}\left(\mathbf{A}^T \mathbf{P}_{yy} \mathbf{A}\right)^{-1} \mathbf{A}^T \mathbf{P}_{yy}$ is the well-known projection matrix whose geometrical meaning is mentioned in, e.g., Koch (1999), Vennebusch et al. (2009). The matrix $\mathbf{I}_n - \mathbf{R}$ is also called the reliability matrix (Shan 1989) or redundancy matrix (Schaffrin 1997). $\mathbf{I}_n$ is the $n \times n$ identity matrix.

The cofactor matrices of the estimated vectors are obtained by error propagation as follows

$$\mathbf{Q}_{\hat{\xi}\hat{\xi}} = \left(\mathbf{A}^T \mathbf{P}_{yy} \mathbf{A}\right)^{-1}, \tag{2.14}$$

$$\mathbf{Q}_{\hat{y}\hat{y}} = \mathbf{R}\mathbf{Q}_{yy}\mathbf{R}^T, \tag{2.15}$$

$$\mathbf{Q}_{\hat{v}_y\hat{v}_y} = \left(\mathbf{I}_n - \mathbf{R}\right)\mathbf{Q}_{yy}\left(\mathbf{I}_n - \mathbf{R}\right)^T. \tag{2.16}$$

The unbiasedness property of the parameter vector, observation vector, residual vector and mean squared error of the estimated parameter vector can be shown as follows

$$E\left(\hat{\xi}\right) = \left(\mathbf{A}^T \mathbf{P}_{yy} \mathbf{A}\right)^{-1} \mathbf{A}^T \mathbf{P}_{yy} \mathbf{A}\xi = \xi, \tag{2.17}$$

$$E\left(\hat{\mathbf{y}}\right) = E\left(\mathbf{A}\hat{\xi}\right) = \mathbf{A}E\left(\hat{\xi}\right) = \mathbf{A}\xi = E\left(\mathbf{y}\right), \tag{2.18}$$

$$E\left(\hat{\mathbf{v}}_y\right) = E\left(-\left(\mathbf{I}_n - \mathbf{R}\right)\mathbf{y}\right) = -\left(\mathbf{I}_n - \mathbf{R}\right)\mathbf{A}\xi = \mathbf{0} = E\left(\mathbf{v}_y\right), \tag{2.19}$$

$$E\left(\left\|\hat{\xi} - \xi\right\|^2\right) = tr\left(E\left(\left(\hat{\xi} - \xi\right)\left(\hat{\xi} - \xi\right)^T\right)\right) = \sigma_0^2 tr\left(\mathbf{Q}_{\hat{\xi}\hat{\xi}}\right), \tag{2.20}$$

where the symbol tr denotes the trace operation.

The unbiased and best estimator of the unit of weight can be obtained as

$$\hat{\sigma}_0^2 = \frac{\left(\mathbf{A}\hat{\xi} - \mathbf{y}\right)^T \mathbf{P}_{yy} \left(\mathbf{A}\hat{\xi} - \mathbf{y}\right)}{n - u} = \frac{\mathbf{y}^T \left(\mathbf{I}_n - \mathbf{R}\right)^T \mathbf{P}_{yy} \left(\mathbf{I}_n - \mathbf{R}\right)\mathbf{y}}{n - u}. \tag{2.21}$$

which can be expressed without the estimated parameter vector.

Best linear unbiased estimation

The BLUE (cf. Koch 1999) is obtained according to the optimal properties $E\left(s(\mathbf{y})-h(\xi)\right)=\mathbf{0}$ and $E\left(\left\|s(\mathbf{y})-h(\xi)\right\|^2\right)\rightarrow \min$. Through the LS estimator and BLUE in linear model based on the different principles, both estimators lead to the identical estimator for the unknown parameters.

Parameter elimination

The general approach for the elimination of unknown parameters is a block-wise reduction of the functional model. In this case, the parameter vector and model matrix are decomposed to two sub-parameter vectors and sub model matrices

$$\mathbf{y}+\mathbf{v}_{\mathrm{y}}=\begin{bmatrix}\mathbf{A}_1 & \mathbf{A}_2\end{bmatrix}\begin{bmatrix}\xi_1\\ \xi_2\end{bmatrix} \tag{2.22}$$

Inserting the decomposed model matrix and the parameter vector to the normal equation (2.10), we will have

$$\begin{bmatrix}\mathbf{A}_1^T\mathbf{P}_{\mathrm{yy}}\mathbf{A}_1 & \mathbf{A}_1^T\mathbf{P}_{\mathrm{yy}}\mathbf{A}_2\\ \mathbf{A}_2^T\mathbf{P}_{\mathrm{yy}}\mathbf{A}_1 & \mathbf{A}_2^T\mathbf{P}_{\mathrm{yy}}\mathbf{A}_2\end{bmatrix}\begin{bmatrix}\hat{\xi}_1\\ \hat{\xi}_2\end{bmatrix}=\begin{bmatrix}\mathbf{A}_1 & \mathbf{A}_2\end{bmatrix}^T\mathbf{P}_{\mathrm{yy}}\mathbf{y}\,. \tag{2.23}$$

Through the matrix identities (Niemeier 2002 p. 287), the subparameter vector $\hat{\xi}_2$ can be eliminated. The subparameter vector $\hat{\xi}_1$ can be estimated with the following formula

$$\mathbf{A}_1^T\mathbf{P}_{\mathrm{yy}}\left(\mathbf{I}_n-\mathbf{A}_2\left(\mathbf{A}_1^T\mathbf{P}_{\mathrm{yy}}\mathbf{A}_1\right)^{-1}\mathbf{A}_2^T\mathbf{P}_{\mathrm{yy}}\right)\mathbf{A}_1\hat{\xi}_1=\mathbf{A}_1^T\mathbf{P}_{\mathrm{yy}}\left(\mathbf{I}_n-\mathbf{A}_2\left(\mathbf{A}_1^T\mathbf{P}_{\mathrm{yy}}\mathbf{A}_1\right)^{-1}\mathbf{A}_2^T\mathbf{P}_{\mathrm{yy}}\right)\mathbf{y}\,. \tag{2.24}$$

where the matrix $\mathbf{I}_n-\mathbf{A}_2\left(\mathbf{A}_1^T\mathbf{P}_{\mathrm{yy}}\mathbf{A}_1\right)^{-1}\mathbf{A}_2^T\mathbf{P}_{\mathrm{yy}}$ is idempotent and plays an important role in the parameter elimination process. Then, the subparameter vector $\hat{\xi}_2$ can be given.

GMM with linear constraints

The GMM with linear constraints is frequently encountered in geodetic applications. There are some approaches to solve the problem (e.g., Lagrange multipliers or using the pseudo observation equation). The functional model can be described as follows

$$\begin{aligned}\mathbf{y}+\mathbf{v}_{\mathrm{y}}&=\mathbf{A}\xi\\ \kappa_0&=\mathbf{K}\xi\end{aligned}\,, \tag{2.25}$$

where $\mathbf{K}$ is the deterministic constraints matrix and κ_0 is the constraints constant.

Applying the Lagrange multiplier vector $\hat{\boldsymbol{\mu}}$, the normal equation can be arranged:

$$\begin{bmatrix} \mathbf{A}^T\mathbf{P}_{yy}\mathbf{A} & \mathbf{K}^T \\ \mathbf{K} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \hat{\boldsymbol{\xi}} \\ \hat{\boldsymbol{\mu}} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^T\mathbf{P}_{yy}\mathbf{y} \\ \boldsymbol{\kappa}_0 \end{bmatrix}, \tag{2.26}$$

If one inverts the normal matrix, the parameter vector can be obtained.

2.3 Gauss-Helmert Model

A generalized model $\mathbf{f}\left(E(\mathbf{l}),\boldsymbol{\xi}\right)=\mathbf{0}$ in comparison to $E(\mathbf{l})=\mathbf{f}(\boldsymbol{\xi})$ is frequently encountered. One can connect the expectation of the observation vector $\mathbf{y}$ and the parameter vector $\boldsymbol{\xi}$ through a non-linear, differentiable equation

$$\mathbf{f}\left(E(\mathbf{l}),\boldsymbol{\xi}\right)=\mathbf{0} \quad \text{or} \quad \mathbf{f}\left(\mathbf{l}+\mathbf{v},\boldsymbol{\xi}^0+d\boldsymbol{\xi}\right)=\mathbf{0}. \tag{2.27}$$

which is termed as the non-linear GHM.

The Jacobian matrices (model matrices) with respect to parameters and observations can be expressed as

$$\begin{aligned} \mathbf{A}^0 &= \left.\frac{\partial \mathbf{f}}{\partial \boldsymbol{\xi}^T}\right|_{\mathbf{l}^0,\boldsymbol{\xi}^0} \\ \mathbf{B}^0 &= \left.\frac{\partial \mathbf{f}}{\partial \mathbf{l}^T}\right|_{\mathbf{l}^0,\boldsymbol{\xi}^0} \end{aligned}. \tag{2.28}$$

where $\mathbf{l}^0$ is the initial value of the observation vector. Note the vector must be deterministic so that both Jacobian matrices are non-random. The Jacobian matrices $\mathbf{A}^0$ and $\mathbf{B}^0$ have the dimension $n\times u$ and $n\times m$, where n and m denote the number of equations and observations.

Then, the model can be linearized and rewritten as a linear (or linearized) GHM as follows

$$\begin{aligned} &\mathbf{A}^0 d\boldsymbol{\xi}+\mathbf{B}^0\mathbf{v}+\mathbf{w}^0=\mathbf{0} \\ &E(\mathbf{v})=\mathbf{0} \quad D(\mathbf{l})=\sigma_0^2\mathbf{P}^{-1} \end{aligned} \tag{2.29}$$

where $\mathbf{w}^0=\mathbf{f}\left(\mathbf{l}^0+\mathbf{v}^0,\boldsymbol{\xi}^0\right)-\mathbf{B}^0\mathbf{v}^0$.

Here we note that matrix $\mathbf{A}^0$ has full column rank, and matrix $\mathbf{B}^0$ has full row rank. According to the minimization of the objective function $\mathbf{v}^T\mathbf{P}\mathbf{v}$, the LS solution is produced as follows (e.g., Koch 1999)

$$d\hat{\boldsymbol{\xi}}=\left(\left(\mathbf{A}^0\right)^T\left(\mathbf{B}^0\mathbf{Q}_{ll}\left(\mathbf{B}^0\right)^T\right)^{-1}\mathbf{A}^0\right)^{-1}\left(\mathbf{A}^0\right)^T\left(\mathbf{B}^0\mathbf{Q}_{ll}\left(\mathbf{B}^0\right)^T\right)^{-1}\left(-\mathbf{w}^0\right) \tag{2.30}$$

More detailed information on the linear or non-linear GHM can be found in Pope (1972), Wolf (1978) or Lenzmann and Lenzmann (2004). This special case of the general mixed model that each residual occurs only in one observation equation is called the quasi indirect error adjustment (in German: Quasi-vermittelnde Ausgleichung, see Wolf 1968 p. 105). In this case the row rank of the model matrix $\mathbf{B}^0$ equals to the row number of the matrix, which guarantees the matrix $\left(\mathbf{B}^0\mathbf{Q}_{ll}\left(\mathbf{B}^0\right)^T\right)^{-1}$ exists. The problem introduced in Chapter 4 can be also classified as this adjustment category.

3 Total Least-Squares: A Review

3.1 Classical TLS and Errors-in-Variables model

In this chapter the classical Total Least-Squares method is introduced. The attribute 'classical' refers to an unweighted case, in which there is a unique solution.

Motivation and formulation of TLS problem

It is well-known that for the solution of the overdetermined system $\mathbf{A}\xi = \mathbf{y} + \mathbf{v}_{\mathrm{y}}$ LS is used as standard method. Unfortunately, the model does not always match the reality. In Wicki (1998), Yang (1999) et al, it is shown that the measurement errors do not necessarily fulfill the principal hypothesis of normal distribution in geodetic applications. For this case one can add parameters to model the outliers or reduce the influence on the parameters from the residuals. The observation errors may preclude the possibility of knowing the model matrix $\mathbf{A}$ exactly. If the model matrix $\mathbf{A}$ is mathematically inexact or rather contaminated with errors, the model should be dealt with another strategy.

The terminology TLS was introduced by Golub and Van Loan (1980) and is considered as an estimation problem where the observation vector $\mathbf{y}$ and the model matrix $\mathbf{A}$ are erroneous. The mathematical model of the classical TLS is

$$\min_{[\mathbf{V}_{\mathrm{A}} \;\; \mathbf{v}_{\mathrm{y}}]} \left\| \begin{bmatrix} \mathbf{V}_{\mathrm{A}} & \mathbf{v}_{\mathrm{y}} \end{bmatrix} \right\|_F \quad \text{subject to } \mathbf{y} + \mathbf{v}_{\mathrm{y}} = (\mathbf{A} + \mathbf{V}_{\mathrm{A}})\xi \tag{3.1}$$

where the full rank matrix $\mathbf{A}$ affected by errors and the vector $\mathbf{y}$ have the residuals $\mathbf{V}_{\mathrm{A}}$ and $\mathbf{v}_{\mathrm{y}}$, and ξ is an unknown parameter vector. $\| \;\|_F$ denotes the Frobenius norm, which is defined as $\|\mathbf{H}\|_F = \sqrt{tr(\mathbf{H}^T\mathbf{H})}$. The model of this type $\mathbf{y} + \mathbf{v}_{\mathrm{y}} = (\mathbf{A} + \mathbf{V}_{\mathrm{A}})\xi$ is known in the literature (e.g. Gleser 1981) as EIV model.

Solution

The key role of LS for the GMM is the same as one of TLS in EIV modeling. The model can be rewritten as

$$\begin{bmatrix} \mathbf{A} + \mathbf{V}_{\mathrm{A}} & \mathbf{y} + \mathbf{v}_{\mathrm{y}} \end{bmatrix} \begin{bmatrix} \xi \\ -1 \end{bmatrix} = \mathbf{0}. \tag{3.2}$$

In the generic case this problem has a unique solution, which can be obtained in a closed form in terms of the SVD of the data matrix. The SVD decomposition of this augmented matrix is

$$\begin{bmatrix} \mathbf{A} & \mathbf{y} \end{bmatrix} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \tag{3.3}$$

where

$$\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n] \in \Re^{n \times n}$$
$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{u+1}] = [v_{ik}] \in \Re^{(u+1)\times(u+1)}, \tag{3.4}$$

and

$$\mathbf{\Sigma} = diag(\sigma_1, \sigma_2, ..., \sigma_{u+1}) = [\sigma_{ij}] \tag{3.5}$$

is an $n \times (u+1)$ matrix. Diagonal elements are equal to the singular values and off-diagonal elements are zero. i.e. $\sigma_{ii} = \sigma_i$ if $i = 1,...,u+1$ and $\sigma_{ij} = 0$ if $\mathrm{i} \neq \mathrm{j}$. The augmented matrix denotes the matrix $\mathbf{A}$ that has an extra column containing the right-hand side (rhs) terms $\mathbf{y}$ of the equation system.

The conditions for the existence and uniqueness of the TLS solution are

1. A TLS solution exists if and only if $v_{u+1,u+1} \neq 0$.
2. A TLS solution is unique if and only if $\sigma_u \neq \sigma_{u+1}$.

The equivalent condition for the existence and uniqueness of the TLS solution can be described as $\gamma_u > \sigma_{u+1}$, where γ_u is the minimum singular value of the matrix $\mathbf{A}^T\mathbf{A}$ (e.g., Van Huffel and Vandewalle 1991). If the solution is not unique, this problem is often called nongeneric TLS problem. If the conditions are satisfied, the estimated data matrix is given by

$$[\hat{\mathbf{A}} \quad \hat{\mathbf{y}}] = \mathbf{U}\mathbf{\Sigma}^*\mathbf{V}^T, \tag{3.6}$$

where $\mathbf{\Sigma}^* = diag(\sigma_1, \sigma_2, ..., \sigma_u, 0)$.

Then, the solution is solved by using the Eckart-Young-Mirsky theorem (e.g., Eckart and Young 1936) to let the data matrix to reduce the rank of data matrix and have minimal norm $\left\| [\hat{\mathbf{A}} \quad \hat{\mathbf{y}}] - [\mathbf{A} \quad \mathbf{y}] \right\|_F$. Because of $N([\hat{\mathbf{A}} \quad \hat{\mathbf{y}}]) = R(\mathbf{v}_{u+1})$ we have the estimated parameter after scaling the last component $v_{u+1,u+1}$ of $\mathbf{v}_{u+1}$ to -1 to identify the vector $\begin{bmatrix} \xi \\ -1 \end{bmatrix}$

$$\hat{\xi}_{TLS} := -[v_{1,u+1}, ..., v_{u,u+1}]^T / v_{u+1,u+1} \tag{3.7}$$

and the residual matrix

$$[\hat{\mathbf{V}}_{\mathbf{A}} \quad \hat{\mathbf{v}}_{\mathbf{y}}] = [\hat{\mathbf{A}} \quad \hat{\mathbf{y}}] - [\mathbf{A} \quad \mathbf{y}] = -\sigma_{u+1}\mathbf{u}_{u+1}\mathbf{v}_{u+1}^T, \tag{3.8}$$

where the symbol N and R denote the null space and column space.

The classical TLS problems can be solved by the strategy introduced above. Other methods to solve this problem are also widely discussed in geodesy. The first investigation taking uncertain model matrix into account in geodesy was carried out by Teunissen (1988). He derived a closed form solution for the EIV model from a two-step procedure. Afterward, Schaffrin (2006) solved the classical TLS problem using a non-linear Lagrange function approach. In addition, an unweighted multivariate TLS approach was presented by Schaffrin and Felus (2008) for the geodetic transformation. The geometrical interpretation and alternative definition of the TLS problem can be found in Kupferer (2005).

The TLS problem has an analytical expression (e.g., Schaffrin 2006)

$$\hat{\xi}_{TLS} = \left(\mathbf{A}^T\mathbf{A} - \sigma_{u+1}^2\mathbf{I}_u\right)^{-1}\mathbf{A}^T\mathbf{y}, \tag{3.9}$$

where the σ_{u+1} is the smallest singular value, as mentioned above.

The alternative form indicates that the TLS solution is more ill-conditioned than the LS solution due to the higher condition number (It must be noted that the both solutions are not certainly ill-conditioned. Here, 'more ill-conditioned' only denotes a relative comparison). The reason is that the form $\left(\mathbf{A}^T\mathbf{A}-\sigma_{u+1}^2\mathbf{I}_u\right)^{-1}\mathbf{A}^T\mathbf{y}$ is also the solution of the minimization problem for $\left\|\mathbf{A}\xi-\mathbf{y}\right\|_2^2+\left(-\sigma_{u+1}^2\right)\left\|\xi\right\|_2^2$ considered as a reverse ridge regression (Golub and Van Loan 1980) or as deregularizing procedure (Markovsky and Van Huffel 2007). It means that errors in the data can exert a more considerable influence on the LS solution. It is also shown by Van Huffel and Vandewalle (1991) that differences between the LS and TLS solution increase if the ratio between the second smallest singular value of $\left[\mathbf{A}\quad\mathbf{y}\right]$ and the smallest singular value of $\mathbf{A}$ is growing.

3.2 Fixing columns

In many geodetic applications (e.g., geodetic transformation) some of the columns of $\mathbf{A}$ are known exactly. Consequently, for the estimation of parameters in this case one should keep the corresponding columns of $\mathbf{A}$ unchanged since they are fixed.

We separate the model matrix $\mathbf{A}$ into two parts: the uncertain, known (and hence deterministic) part $\mathbf{A}_1$ and the stochastic part $\mathbf{A}_2$. This separation is often called fixing columns of the design matrix in TLS (e.g., Schaffrin and Felus 2008) or mixed LS-TLS problem (e.g., Kupferer 2005). We write the optimization problem as follows

$$\begin{aligned}&\min\left\|\left[\mathbf{V}_2\quad\mathbf{v}_y\right]\right\|_F\\&\textit{subject to }\mathbf{y}+\mathbf{v}_y=\left(\mathbf{A}+\mathbf{V}_A\right)\xi=\mathbf{A}_1\xi_1+\mathbf{A}_2\xi_2+\mathbf{V}_2\xi_2\end{aligned},\tag{3.10}$$

where $\mathbf{A}=\left[\mathbf{A}_1,\mathbf{A}_2\right]$, $\mathbf{V}_A=\left[\mathbf{0},\mathbf{V}_2\right]$, $\xi=\left[\xi_1;\xi_2\right]$. The ξ_2 vector is the subparameter vector with $u-t$ elements corresponding to the sub matrix $\mathbf{A}_2$; ξ_1 with t elements is corresponding to the deterministic part $\mathbf{A}_2$ of $\mathbf{A}$.

In order to solve this problem the QR decomposition of the augmented matrix $\left[\mathbf{A}_1\quad\mathbf{A}_2\quad\mathbf{y}\right]$ is applied as follows

$$\left[\mathbf{A}_1\quad\mathbf{A}_2\quad\mathbf{y}\right]=\mathbf{Q}\cdot\mathbf{R}\tag{3.11}$$

or with the matrix transpose of $\mathbf{Q}$ as

$$\mathbf{Q}^T\left[\mathbf{A}_1\quad\mathbf{A}_2\quad\mathbf{y}\right]=\mathbf{Q}^T\mathbf{Q}\mathbf{R}=\mathbf{R}=\begin{bmatrix}\mathbf{R}_{11}&\mathbf{R}_{12}&\mathbf{r}_{1b}\\\mathbf{0}&\mathbf{R}_{22}&\mathbf{r}_{2b}\end{bmatrix}\begin{matrix}t\\n-t\end{matrix},\quad\begin{matrix}t&u-t&1\end{matrix}\tag{3.12}$$

which leads the problem to

$$\begin{bmatrix}\mathbf{R}_{11}&\mathbf{R}_{12}\\\mathbf{0}&\mathbf{R}_{22}\end{bmatrix}\begin{bmatrix}\xi_1\\\xi_2\end{bmatrix}=\begin{bmatrix}\mathbf{r}_{1b}\\\mathbf{r}_{2b}\end{bmatrix}.\tag{3.13}$$

Thus, the solution of this mixed problem has two steps:

1) The subparameters ξ_2 can be estimated as

$$\mathbf{R}_{22}\hat{\xi}_2\approx\mathbf{r}_{2b},\tag{3.14}$$

which can be solved with classical TLS approach.

2) The subparameters ξ_1 can be estimated as

$$\mathbf{R}_{11}\hat{\xi}_1 = \mathbf{r}_{1b} - \mathbf{R}_{12}\hat{\xi}_2 \tag{3.15}$$

with a regular inversion.

3.3 Constrained TLS

Recently, Schaffrin and Felus (2005), Schaffrin (2006) and Schaffrin and Felus (2009) described a strategy to solve the classical TLS problem with linear (stochastic and deterministic) and quadratic constraints. The problem with linear constraints will be described in detail as follows

$$\begin{aligned} &\mathbf{y} + \mathbf{v}_y = (\mathbf{A} + \mathbf{V}_A)\xi \\ &\kappa_0 = \mathbf{K}\xi \end{aligned} , \tag{3.16}$$

which leads to

$$\begin{aligned} &\mathbf{y} + \mathbf{v}_y = (\mathbf{A}_1 + \mathbf{V}_{A_1})\xi_1 + (\mathbf{A}_2 + \mathbf{V}_{A_2})\xi_2 \\ &rank\mathbf{A} = rank[\mathbf{A}_1, \mathbf{A}_2] = u < n \\ &\kappa_0 = \mathbf{K}_2\xi_2 + \mathbf{K}_1\xi_1, \mathbf{K}_1 \text{ invertible} \end{aligned} , \tag{3.17}$$

where q is the number of constraints, $rank[\mathbf{K}_1, \mathbf{K}_2] = q < u$. $\mathbf{K}$ is the deterministic constraints matrix and κ_0 is the constraints constant. $\mathbf{K}_1$ and $\mathbf{K}_2$ are the fixed constraints matrix correspondent to ξ_1 and ξ_2, respectively.

Part of the parameters can be presented from the other parameter part via constraints

$$\xi_1 = \mathbf{K}_1^{-1}(\kappa_0 - \mathbf{K}_2\xi_2). \tag{3.18}$$

By inserting these parameters back to the model we have

$$\left[(\mathbf{A}_2 - \mathbf{A}_1\mathbf{K}_1^{-1}\mathbf{K}_2) + (\mathbf{V}_{A_2} - \mathbf{V}_{A_1}\mathbf{K}_1^{-1}\mathbf{K}_2)\right]\xi_2 = (\mathbf{y} - \mathbf{A}_1\mathbf{K}_1^{-1}\kappa_0) + (\mathbf{v}_y - \mathbf{V}_{A_1}\mathbf{K}_1^{-1}\kappa_0), \tag{3.19}$$

which leads to a weighted TLS problem. The stochastic property for the vectors $vec(\mathbf{V}_{A_2} - \mathbf{V}_{A_1}\mathbf{K}_1^{-1}\mathbf{K}_2)$ and $\mathbf{v}_y - \mathbf{V}_{A_1}\mathbf{K}_1^{-1}\kappa_0$ can be obtained through the error propagation.

This constrained problem can be solved with the traditional Lagrange multiplier as follows:

$$\Phi(\mathbf{V}_A, \mathbf{v}_y, \lambda, \xi, \mu) = tr(\mathbf{V}_A\mathbf{V}_A^T) + \mathbf{v}_y^T\mathbf{v}_y + 2\lambda^T(\mathbf{y} + \mathbf{v}_y - (\mathbf{A} + \mathbf{V}_A)\xi) + 2\mu^T(\kappa_0 - \mathbf{K}\xi), \tag{3.20}$$

where μ is the Lagrange multiplier for the constraints.

After some simplifications (see Schaffrin and Felus 2005) the non-linear normal equation can be expressed as

$$\begin{bmatrix} \mathbf{A}^T\mathbf{A} & \mathbf{A}^T\mathbf{y} & \mathbf{K} \\ \mathbf{y}^T\mathbf{A} & \mathbf{y}^T\mathbf{y} & \kappa_0 \\ \mathbf{K} & \kappa_0 & \hat{v}\mathbf{I}_q \end{bmatrix} \begin{bmatrix} \hat{\xi} \\ -1 \\ \hat{\mu} \end{bmatrix} = \begin{bmatrix} \hat{\xi} \\ -1 \\ \hat{\mu} \end{bmatrix} \hat{v}, \tag{3.21}$$

where $\hat{v} = (\mathbf{y} - \mathbf{A}\xi)^T(\mathbf{y} - \mathbf{A}\xi)/(1 + \xi^T\xi)$.

Obviously, this system can be interpreted implicitly as an eigenvalue problem. Based on it a converging algorithm has been proposed in Schaffrin and Felus (2005).

3.4 Structured TLS

If the structure of the model matrix $\mathbf{A}$ is composed in the way that some variables appear twice or more, this is called Structured TLS problem. The name Structured TLS (STLS) was first introduced by De Moor (1993). The STLS problem occurs frequently in signal processing applications (e.g. Markovsky and Van Huffel 2007). In geodesy a practical coordinate transformation problem is presented to demonstrate this technique by Felus and Schaffrin (2005), and a comparison with the LS is made. In this part we give an example to demonstrate this problem. A six parameter transformation (cf. Schaffrin and Felus 2008) is presented as follows

$$\begin{aligned}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} &\approx \begin{bmatrix} s_1 \cos\beta & -s_2 \sin(\beta+\varepsilon) \\ s_1 \sin\beta & s_2 \cos(\beta+\varepsilon) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \\ &=: \begin{bmatrix} \xi_{11} & \xi_{21} \\ \xi_{12} & \xi_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \xi_{31} \\ \xi_{32} \end{bmatrix}\end{aligned}, \tag{3.22}$$

where s_1 and s_2 are the two scale factors, β and $\beta+\varepsilon$ are the respective rotation angels, t_1 and t_2 are the shifts along the respective axes, y_1 and y_2 are the transformed coordinates, x_1 and x_2 are the original coordinates. The six physical (geometric) parameters s_1, s_2, β, ε, t_1 and t_2 are usually replaced by six mathematical parameters ξ_{11}, ξ_{21}, ξ_{31}, ξ_{12} ξ_{22}, ξ_{32} in the adjustment.

If some identical points in both coordinate systems are measured, the transformation equation system can be written as follows

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{1}_n & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{1}_n \end{bmatrix} \begin{bmatrix} \xi_{11} \\ \xi_{21} \\ \xi_{31} \\ \xi_{12} \\ \xi_{22} \\ \xi_{32} \end{bmatrix}, \tag{3.23}$$

where $\mathbf{1}_n$ is a $n\times1$ vector of ones. $\mathbf{x}_1, \mathbf{x}_2$ are the coordinate vectors in the original coordinate system, whereas $\mathbf{y}_1$, $\mathbf{y}_2$ are the coordinate vectors in the transformed coordinate system.

It is obvious that the model matrix has the special structure, and the correspondent residuals for the model matrix should also have this structure as follows

$$\mathbf{V}_{\mathbf{A}} = \begin{bmatrix} \mathbf{v}_{\mathbf{x}_1} & \mathbf{v}_{\mathbf{x}_2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{v}_{\mathbf{x}_1} & \mathbf{v}_{\mathbf{x}_2} & \mathbf{0} \end{bmatrix}. \tag{3.24}$$

Note that the residual vector should be the vector of zeros for the deterministic variables.

Hence we cannot obtain a correct result with the classical TLS estimator, because in the classical case the residual vector does not appear twice or more in the residual matrix. The STLS solution put forward by Felus and Schaffrin (2005) has replaced the model matrix by a vector of independent elements in the structured matrix multiplying the characteristic matrix, and used an iterative SVD process. The solution has been employed to estimate the rotation and the scale parameters of a coordinate transformation problem. The estimation process is usually called Cadzow's algorithm (see Schaffrin et al 2009). The modified version has been presented by Schaffrin et al (2009). The modified Cadzow's algorithm presented the identical parameter estimates (except the variance component) as the exact solution, which has been solved by the iterative GHM method proposed (e.g., in Neitzel 2010). However, the computation of the (modified) Cadzow's algorithm is expensive since the SVD method has to be applied many times.

3.5 Generalized TLS and element-wise TLS

In this section we introduce extensions of the classical TLS problem, in which the errors have different size.

In the generalized TLS estimator, the residual matrix $\left[\mathbf{V}_A, \mathbf{v}_y\right]$ is assumed to be row-wise independent and correlated within the rows with identical vcm. Yet, the so-called generalized TLS does not refer to a general vcm. Some authors (e.g., Schaffrin and Wieser 2008) call it 'equilibrated TLS' to avoid the confusions. The solution for the so-called generalized TLS using a three-step algorithm has been propounded by Van Huffel and Vandewalle (1991), Felus (2004). The optimization problem can be expressed as follows

$$\begin{aligned} &\min_{[\mathbf{V}_2 \quad \mathbf{v}_y]} \left\| \mathbf{D}\left[\mathbf{V}_2 \quad \mathbf{v}_y\right]\mathbf{C} \right\|_F \\ &\textit{subject to } \mathbf{y}+\mathbf{v}_y = \left(\mathbf{A}+\mathbf{V}_A\right)\boldsymbol{\xi} = \mathbf{A}_1\boldsymbol{\xi}_1 + \mathbf{A}_2\boldsymbol{\xi}_2 + \mathbf{V}_2\boldsymbol{\xi}_2 \end{aligned}, \tag{3.25}$$

where $\mathbf{D}$ is an $n \times n$ diagonal row-scaling matrix, and $\mathbf{C}$ is a $(u-t+1)\times(u-t+1)$ diagonal column-scaling matrix. The elements of the matrix $\mathbf{D}$ represent the inverse standard deviations associated with rows. The elements of the matrix $\mathbf{C}$ represent the inverse standard deviations associated with the columns of $\left[\mathbf{A} \quad \mathbf{y}\right]$ (see Schaffrin and Felus 2008).

In this case the GTLS problem can be solved by a three step algorithm as follows

1. Find the QR factorization of the augmented matrix $\mathbf{D}\left[\mathbf{A},\mathbf{y}\right]$

$$\mathbf{Q}^T\mathbf{D}\left[\mathbf{A}_1 \quad \mathbf{A}_2 \quad \mathbf{y}\right] = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \mathbf{r}_{1b} \\ \mathbf{0} & \mathbf{R}_{22} & \mathbf{r}_{2b} \end{bmatrix} \begin{matrix} t \\ n-t \end{matrix}. \quad \begin{matrix} t & u-t & 1 \end{matrix} \tag{3.26}$$

2. Compute the classical TLS solution $\hat{\boldsymbol{\xi}}_2$ for a reduced system

$$\left[\mathbf{R}_{22} \quad \mathbf{r}_{2b}\right]\mathbf{C}\left(\mathbf{C}^{-1}\begin{bmatrix} \boldsymbol{\xi}_2 \\ -1 \end{bmatrix}\right) \approx \mathbf{0}. \tag{3.27}$$

where $\mathbf{C}$ is the diagonal column-weighted matrix.

3. Use SVD to obtain the $\hat{\boldsymbol{\xi}}_2$ and $\hat{\boldsymbol{\xi}}_1$

$$\left[\mathbf{R}_{22} \quad \mathbf{r}_{2b}\right]\mathbf{C} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T, \tag{3.28}$$

$$\hat{\boldsymbol{\xi}}_2 = -\frac{1}{c_{22}v_{u+1,u+1}}\mathbf{C}_{11}\left[v_{1,u+1},\ldots,v_{u,u+1}\right]^T, \tag{3.29}$$

$$\mathbf{R}_{11}\hat{\boldsymbol{\xi}}_1 = \mathbf{r}_{1b} - \mathbf{R}_{12}\hat{\boldsymbol{\xi}}_2, \tag{3.30}$$

where $\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \\ & c_{22} \end{bmatrix}$ with diagonal matrix $\mathbf{C}_{11}$.

When the elements of the residual matrix $\left[\mathbf{V}_{\mathbf{A}}, \mathbf{v}_{\mathbf{y}}\right]$ are independent, but not identically distributed with element-wise different variances, this further generalization is called element-wise-weighted TLS. Solving EIV model in this case has been proved as a non-convex optimization problem (e.g., in the simplest case the optimization problem can be described as $\min \frac{(x-1)^2}{(1+x^2)}$), and there is no analytical solution of EW-TLS (Markovsky et al. 2006, Schaffrin and Wieser 2008). In Markovsky et al. (2006) this optimization problem was solved as n independent optimization problems. An iterative algorithm was designed which led to a local minimum. Lately, Schaffrin and Wieser (2008) introduced a solution with a certain block structured vcm using Lagrange multiplier which is only a special type of EW-TLS.

The problems, such as the special weighted TLS problem, the TLS problem with fixing column, the TLS problem with linear constraints, introduced in this chapter will be generalized in the case that the vcm is fully populated. The general weighted TLS solution which can treat the fully correlated data will be presented in Chapter 4. The fact that the general solution cannot be established by the presented strategies (e.g., SVD method) is proved. Thus, the methods which have been presented in this chapter do not explicitly relate to the approaches applied in the general weighted TLS Problem. The corresponding extension (e.g., with fixing column and linear constraints) will be discussed in Chapter 5.

4 Weighted TLS solutions

In the previous chapter we reviewed the development and extension of the classical TLS technique. In addition, solutions have been presented for its extension to structured, constrained and weighted problems. The classical TLS has a unique analytical solution in the generic case, whereas the structured, constrained and weighted TLS problems have no such closed solution and are currently solved numerically to a local minimum. However, the weighted TLS problem with the general vcm has not been widely discussed until now.

The WTLS solution is useful for, e.g., the rigorous modeling of point clouds for surface reconstruction. There the vcm is usually obtained from the error propagation of the original error source, which varies dependent on, e.g. distance, angles, intensity, etc. The typical surface reconstruction is also very important in the geosciences and used to adjust gravity or magnetic observations and to remove the global trend and systematic effects (Schaffrin and Felus 2009). For another geodetic application, 'the empirical coordinate transformation', the vcm mentioned in the present TLS literature takes the correlation between the columns of the data matrix into account rather than the vcms of the point coordinates in the coordinate set resulting from network adjustment (Akyilmaz 2007). Meanwhile, Schaffrin and Felus (2005) and Schaffrin and Felus (2009) have shown that the WTLS can be used as a promising tool to solve the constrained TLS problem. Furthermore, the data sets can provide large temporal or regional correlations, and the character of uncertainty should be expressed as a general vcm. One common statistical model for many types of geodetic and geophysical signals may be described as a power-law process, whose variance covariance matrix structure can be seen in Williams (2003). The efficient algorithms for the structured TLS with large block matrices are given by Markovsky and Van Huffel (2005). In the aspect of combination of the surface reconstruction and signal processing, the rigorous collocation method used, e.g., in the gravity domain could be also concerned.

Although the error adjustment problem has been studied as TLS, EIV model and orthogonal regression over a hundred years in different science domains and occurred in numerous literature, the general solution of a Total Least-Squares problem is still not available (cf. Markovsky et al., 2006 and Schaffrin and Wieser 2008). The situation of the research on the general solution of the fully weighted TLS can be also found in other geodetic literature:

- The solution of a fully weighted TLS problem is still an open question (Schaffrin and Felus 2009).
- A TLS problem that can handle differently weighted and correlated measurements is still under investigation (Felus and Burtch 2009).
- No solution to such a problem, which takes purely the variance covariance matrices of the two coordinate sets into account, has been achieved yet, even in mathematical science (Akyilmaz 2007).
- It is emphasized, however, that the notion of a suitable weight choice is still unclear and needs to be investigated in the future (Schaffrin 2007).
- After avoiding the pitfalls, the problem can be solved without the limitation of the vcm using the non-linear GHM. However, the approach is only a particular method to solve the non-linear normal equation (Schaffrin 2007).

However, from the statistician's point of view, a general approach and thorough understanding for adjusting the EIV model is required. From the geodesist's point of view, the estimation method with general vcm is an urgent task, as the different precision and correlation of the observations is encountered nowadays in many geodetic applications. Hence, it makes sense to solve this problem not only for theoretical understanding, but also for practical purposes.

In this chapter the fully weighted TLS problem is studied. Firstly, the various objective functions are presented and proved as being equivalent. In addition, after giving the standard definition for the weighted TLS, the traditional Lagrange approach is applied to solve the fully weighted univariate EIV model. Based on the full analysis of the non-linear normal equations, three solutions are displayed and the correspondent algorithms are designed. Although one of our solutions, which do not require the linearization, is obtained by the different starting points from the LS estimator (nonlinear GHM method) proposed by e.g., Neitzel (2010), one can see that they have the identical closed-form expression. The alternative definition for the WTLS problem, which can be easily compared with that of the LS estimator, is convenient to calculate the first and second derivative of the objective function w.r.t. the parameter vector from the numerical analysis aspect, the latter of which represents the sufficient conditions of the WTLS solutions. Based on the second derivative one can obtain a better understanding about the non-convex optimization problem and design an algorithm of the Newton type, which converges more efficiently than the algorithms of Gauss-Newton type at least in the final stage of the iteration. Furthermore, the WTLS problem with the quadratic objective function can be interpreted as the standard Gauss-Newton type problem from the view of the numerical analysis. After comprising the algorithms designed in the thesis with the existing methods, e.g., Schaffrin and Wieser (2008), Neitzel (2010) and Shen et al. (2010) the advantages of the general solutions are presented.

4.1 Objective function and general solution based on SVD decomposition

Objective function

The objective function of the classical TLS to solve the EIV model can be expressed in various forms. The original LS estimation minimizes the sum of squared residuals whereas the classical total least-squares optimization problem is to find the minimal sum of weighted squared residuals: $\min_{\xi} \frac{\|\mathbf{A}\xi - \mathbf{y}\|^2}{\|\xi\|^2 + 1}$ (cf. Markovsky and Van Huffel 2007). The objective function can be understood as the squared residuals multiplying the correspondent weight matrix if the errors are independent and identically distributedThe other formulation of the objective function is the minimum Frobenius norm $\left\|\left[\mathbf{V_A}, \mathbf{v_y}\right]\right\|_F$ of the residual matrix (see Chapter 3), or as minimization of $vec^T\left(\mathbf{V_A}\right)vec\left(\mathbf{V_A}\right) + \mathbf{v_y}^T\mathbf{v_y}$ after the maximum likelihood method if similarly expressed (see Schaffrin and Felus 2008).

The objective function about the generalized TLS is defined as $\min_{\left[\mathbf{V_A} \quad \mathbf{v_y}\right]} \left\|\mathbf{D}\left[\mathbf{V_A} \quad \mathbf{v_y}\right]\mathbf{C}\right\|_F$. Both scaling matrices can be defined as $\mathbf{C} = \mathbf{I}$ and $\mathbf{D} = diag\left(1/s_1, \ldots, 1/s_n\right)$ (cf. Felus 2004). With the matrix property $vec^T\left(\mathbf{T}^T\right)\left(\mathbf{S}^T \otimes \mathbf{U}\right)vec\left(\mathbf{V}\right) = tr\left(\mathbf{UVST}\right)$ (e.g., Amiri-simkooei 2007, Appendix A1) and the definition of the Frobenius norm (see chapter 3.1) we see

$$\begin{aligned}\left\|\mathbf{D}\left[\mathbf{V_A} \quad \mathbf{v_y}\right]\mathbf{C}\right\|_F &= \left[tr\left(\mathbf{D}^2\left[\mathbf{V_A} \quad \mathbf{v_y}\right]\mathbf{C}^2\left[\mathbf{V_A} \quad \mathbf{v_y}\right]^T\right)\right]^{1/2} \\ &= \left[vec^T\left(\left[\mathbf{V_A} \quad \mathbf{v_y}\right]\right)\left(\mathbf{C}^2 \otimes \mathbf{D}^2\right)vec\left(\left[\mathbf{V_A} \quad \mathbf{v_y}\right]\right)\right]^{1/2} \\ &= \left[\mathbf{v}^T\mathbf{P}_{GTLS}\mathbf{v}\right]^{1/2}\end{aligned} \quad , \tag{4.1}$$

where $\mathbf{v} = vec\left(\begin{bmatrix}\mathbf{V}_\mathbf{A} & \mathbf{v}_\mathbf{y}\end{bmatrix}\right)$ is the extended residual vector. $\mathbf{P}_{GTLS} = \mathbf{C}^2 \otimes \mathbf{D}^2$ is the weighted matrix in this generalized TLS case. The operator $\otimes$ is the 'Kronecker-Zehfuss product' (e.g., Grafarend and Schaffrin 1993 p. 409), which is defined by $\mathbf{B} \otimes \mathbf{A} := \left[b_{ij} \cdot \mathbf{A}\right]$ if $\mathbf{B} = \left[b_{ij}\right]$.

We can easily see that the Frobenius norm objective function is the same as the maximum likelihood objective function $vec^T(\mathbf{V}_\mathbf{A})vec(\mathbf{V}_\mathbf{A}) + \mathbf{v}_\mathbf{y}{}^T\mathbf{v}_\mathbf{y}$ in the classical TLS case and GTLS case.

Since $vec(\mathbf{ABC}) = \left(\mathbf{C}^T \otimes \mathbf{A}\right)vec(\mathbf{B})$ (see e.g., Koch 1999), we see that

$$\mathbf{y} - \mathbf{A}\boldsymbol{\xi} = \mathbf{V}_\mathbf{A}\boldsymbol{\xi} - \mathbf{v}_\mathbf{y} = \left[\boldsymbol{\xi}^T \otimes \mathbf{I}_n, -\mathbf{I}_n\right]\begin{bmatrix} vec(\mathbf{V}_\mathbf{A}) \\ \mathbf{v}_\mathbf{y}\end{bmatrix} = \left[\boldsymbol{\xi}^T \otimes \mathbf{I}_n, -\mathbf{I}_n\right]\mathbf{v}, \tag{4.2}$$

which can be regarded as a condition equation adjustment problem $\mathbf{Bv} + \mathbf{w} = \mathbf{0}$ if $\mathbf{B} = \left[\boldsymbol{\xi}^T \otimes \mathbf{I}_n, -\mathbf{I}_n\right]$ and $\mathbf{w} = -(\mathbf{y} - \mathbf{A}\boldsymbol{\xi})$. Then, one can transform $\begin{bmatrix} vec^T\left(\hat{\mathbf{V}}_\mathbf{A}\right) & \hat{\mathbf{v}}_\mathbf{y}^T\end{bmatrix}\begin{bmatrix} vec\left(\hat{\mathbf{V}}_\mathbf{A}\right) \\ \hat{\mathbf{v}}_\mathbf{y}\end{bmatrix}$ to the weighted sum of squared residuals $\dfrac{\left\|\mathbf{A}\hat{\boldsymbol{\xi}} - \mathbf{y}\right\|^2}{\left\|\hat{\boldsymbol{\xi}}\right\|^2 + 1}$ in the unweighted TLS case (see Eq (4.19) for detail).

After recognizing the identity of the Frobenius norm objective function and the maximum likelihood objective function, the objective function expressed as sum of weighted squared residuals will be proven in Eq (4.19) as being equivalent to both mentioned objective functions; it leads to identical results.

The solution based on SVD decomposition

The solution for weighted TLS based on SVD has been given in the literature (e.g., Van Huffel and Vandewalle, 1991, Felus 2004, Akyilmaz 2007 and Schaffrin and Felus 2008). The weighted matrix for the extended observation vector can be expressed as $\mathbf{P}_{GTLS} = \mathbf{C}^2 \otimes \mathbf{D}^2$, where both scaling matrices are diagonal.

In this part the general solution based on the SVD decomposition is given (the both scaling matrices are not diagonal). The functional model $(\mathbf{A} + \mathbf{V}_\mathbf{A})\boldsymbol{\xi} - (\mathbf{y} + \mathbf{v}_\mathbf{y}) = \mathbf{0}$ can be rewritten as follows

$$\mathbf{D}\begin{bmatrix}\mathbf{A} + \mathbf{V}_\mathbf{A} & \mathbf{y} + \mathbf{v}_\mathbf{y}\end{bmatrix}\mathbf{C}\mathbf{C}^{-1}\begin{bmatrix}\boldsymbol{\xi} \\ -1\end{bmatrix} = \mathbf{0} \tag{4.3}$$

or

$$\mathbf{D}\begin{bmatrix}\mathbf{A} + \mathbf{V}_\mathbf{A} & \mathbf{y} + \mathbf{v}_\mathbf{y}\end{bmatrix}\mathbf{C}\begin{bmatrix}\mathbf{C}_1^{-1}\boldsymbol{\xi}c_2 \\ -1\end{bmatrix}c_2^{-1} = \mathbf{0}, \tag{4.4}$$

where $\mathbf{D}^2$ is a symmetric positive definite matrix. $\mathbf{C} = \begin{bmatrix}\mathbf{C}_1 & \mathbf{0} \\ \mathbf{0} & c_2\end{bmatrix}$, $\mathbf{C}_1^{-1}$ exists and $c_2 \neq 0$. $\mathbf{C}^2 \otimes \mathbf{D}^2$ is also a symmetric positive definite matrix.

This system can then be solved with the SVD approach and the solution can be obtained as follows

$$\begin{aligned}&\mathbf{D}[\mathbf{A} \quad \mathbf{y}]\mathbf{C} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T \\ &\hat{\boldsymbol{\xi}} := -\mathbf{C}_1\left[v_{1,u+1}, \ldots, v_{u,u+1}\right]^T / \left(v_{u+1,u+1}c_2\right),\end{aligned} \tag{4.5}$$

where the structure of the matrix $\mathbf{V}$ can be found in Eq (3.4).

The vector $\begin{bmatrix}\mathbf{C}_1^{-1}\boldsymbol{\xi} & -c_2^{-1}\end{bmatrix}$ is the eigenvector associated with the smallest eigenvalue of $\mathbf{C}[\mathbf{A}\quad\mathbf{y}]^T\mathbf{D}^2[\mathbf{A}\quad\mathbf{y}]\mathbf{C}$ yielding the following eigenvector equation

$$\begin{aligned}&\mathbf{C}[\mathbf{A}\quad\mathbf{y}]^T\mathbf{D}^2[\mathbf{A}\quad\mathbf{y}]\mathbf{C}\begin{bmatrix}\mathbf{C}_1^{-1}\boldsymbol{\xi} & -c_2^{-1}\end{bmatrix}^T\\ &=\begin{bmatrix}\mathbf{C}_1\mathbf{A}^T\mathbf{D}^2\mathbf{A}\mathbf{C}_1 & \mathbf{C}_1\mathbf{A}^T\mathbf{D}^2\mathbf{y}c_2\\ c_2\mathbf{y}^T\mathbf{D}^2\mathbf{A}\mathbf{C}_1 & c_2\mathbf{y}^T\mathbf{D}^2\mathbf{y}c_2\end{bmatrix}\begin{bmatrix}\mathbf{C}_1^{-1}\boldsymbol{\xi}\\ -c_2^{-1}\end{bmatrix}\\ &=\sigma_{u+1}^2\begin{bmatrix}\mathbf{C}_1^{-1}\boldsymbol{\xi}\\ -c_2^{-1}\end{bmatrix}\end{aligned} . \tag{4.6}$$

After some rearrangement of Eq (4.6) we have the normal equations of TLS estimation

$$\mathbf{A}^T\mathbf{D}^2\mathbf{A}\hat{\boldsymbol{\xi}}-\sigma_{u+1}^2\mathbf{C}_1^{-2}\hat{\boldsymbol{\xi}}=\mathbf{A}^T\mathbf{D}^2\mathbf{y} . \tag{4.7}$$

The TLS estimator based on SVD with the most general weight matrix is

$$\hat{\boldsymbol{\xi}}:=\left(\mathbf{A}^T\mathbf{D}^2\mathbf{A}-\sigma_{u+1}^2\mathbf{C}_1^{-2}\right)^{-1}\mathbf{A}^T\mathbf{D}^2\mathbf{y} . \tag{4.8}$$

In this case the weighted matrix for the extended observation vector should have the structure $\mathbf{P}=\begin{bmatrix}\mathbf{C}_1^2\otimes\mathbf{D}^2 & \mathbf{0}\\ \mathbf{0} & c_2^2\mathbf{D}^2\end{bmatrix}$. In comparison with the present TLS solution the matrices $\mathbf{C}_1$ and $\mathbf{D}$ could be fully populated.

4.2 Definition of the WTLS problem

In the chapters 2 and 3 we have reviewed the LS and TLS method and the corresponding models. The least-squares estimation is the best linear unbiased estimation when the design matrix $\mathbf{A}$ is free of noise and the observation vector $\mathbf{y}$ is affected by errors. This kind of estimation has frequently been applied in the Gauss-Markov model for the error adjustment. In contrast, an EIV model is a model similar to GMM except that the elements of the design matrix are observed with errors. The LS adjustment is statistically motivated as a maximum likelihood estimator in a linear GMM, and the TLS as maximum likelihood estimator in the EIV model. The definition of the WTLS can be expressed as an optimization problem

$$\begin{aligned}&\mathbf{v}^T\mathbf{P}\mathbf{v}=\min\\ &\textit{subject to }\mathbf{y}+\mathbf{v}_{\mathbf{y}}=\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}\right)\boldsymbol{\xi}\end{aligned} . \tag{4.9}$$

With $\mathbf{v}=vec\left(\begin{bmatrix}\mathbf{V}_{\mathbf{A}} & \mathbf{v}_{\mathbf{y}}\end{bmatrix}\right)$. Note that there is an alternative formulation, which uses the error vector and matrix instead of the correction vector and matrix. i.e. error matrix $\mathbf{E}_{\mathbf{A}}=-\mathbf{V}_{\mathbf{A}}$, and error vector $\mathbf{e}_{\mathbf{y}}=-\mathbf{v}_{\mathbf{y}}$ (see, e.g., Schaffrin and Wieser 2008).

If one wants to take the stochastic property of all errors into account, the observations may be written in an extended vector. Thus, the uncertainty vector and the stochastic properties of the uncertainty can be characterized by the extended dispersion matrix

$$\mathbf{l} = \begin{bmatrix} vec(\mathbf{A}) \\ \mathbf{y} \end{bmatrix}, \qquad \mathbf{v} = \begin{bmatrix} vec(\mathbf{V}_\mathbf{A}) \\ \mathbf{v}_\mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_\mathbf{A} \\ \mathbf{v}_\mathbf{y} \end{bmatrix},$$
$$D(\mathbf{l}) = \sigma_0^2 \mathbf{Q}_{\mathbf{ll}} = \sigma_0^2 \begin{bmatrix} \mathbf{Q}_{\mathbf{AA}} & \mathbf{Q}_{\mathbf{Ay}} \\ \mathbf{Q}_{\mathbf{yA}} & \mathbf{Q}_{\mathbf{yy}} \end{bmatrix} = \sigma_0^2 \mathbf{P}^{-1}, \tag{4.10}$$

with

$$\mathbf{Q}_{\mathbf{ll}} = \begin{bmatrix} \mathbf{Q}_1 \\ \cdots \\ \mathbf{Q}_k \\ \cdots \\ \mathbf{Q}_{u+1} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_{11} & \cdots & \mathbf{Q}_{1j} & \cdots & \mathbf{Q}_{1(u+1)} \\ \cdots & & \cdots & & \cdots \\ \mathbf{Q}_{k1} & \cdots & \mathbf{Q}_{kj} & \cdots & \mathbf{Q}_{k(u+1)} \\ \cdots & & \cdots & & \cdots \\ \mathbf{Q}_{(u+1)1} & \cdots & \mathbf{Q}_{(u+1)j} & \cdots & \mathbf{Q}_{(u+1)(u+1)} \end{bmatrix}. \tag{4.11}$$

$\mathbf{Q}_k$ is a $n \times n(u+1)$ matrix representing the variances and covariances between the elements of the k'th column of the augmented matrix $[\mathbf{A} \quad \mathbf{y}]$ and all elements, $\mathbf{Q}_{kj}$ is the $n \times n$ matrix corresponding to the vcm of the k'th column ($k = j$) or the covariance matrix between the k'th and j'th columns of the augmented matrix $[\mathbf{A} \quad \mathbf{y}]$. σ_0^2 is the unknown variance component. $\mathbf{l}$ is the extended observation vector, which includes the elements of $\mathbf{A}$ and the conventional observation vector $\mathbf{y}$, $\mathbf{v}$ is the corresponding correction vector of the extended observation vector. $\mathbf{Q}_{\mathbf{ll}}$ and $\mathbf{P}$ are the symmetric and positive definite cofactor matrix and the weighted matrix of $\mathbf{l}$, respectively. $\mathbf{Q}_{\mathbf{AA}}$ is the cofactor matrix for $\mathbf{v}_\mathbf{A}$, and the cofactor matrices $\mathbf{Q}_{\mathbf{Ay}}$ and $\mathbf{Q}_{\mathbf{yA}}$ refers to correlations of $\mathbf{v}_\mathbf{A}$ and $\mathbf{v}_\mathbf{y}$.

4.3 General solutions using Lagrange multipliers

In this section we will show how to solve the fully weighted TLS with 'Lagrange multipliers'. According to the traditional Lagrange approach we form the target function as follows

$$\Phi(\mathbf{v}, \boldsymbol{\lambda}, \boldsymbol{\xi}) = \mathbf{v}^T \mathbf{P} \mathbf{v} + 2\boldsymbol{\lambda}^T (\mathbf{y} - \mathbf{A}\boldsymbol{\xi} - \mathbf{V}_\mathbf{A}\boldsymbol{\xi} + \mathbf{v}_\mathbf{y}), \tag{4.12}$$

where $\boldsymbol{\lambda}$ is the Lagrange multipliers vector.

Setting the partial derivatives of the target function w.r.t. $\boldsymbol{\xi}, \mathbf{v}, \boldsymbol{\lambda}$ equal to $\mathbf{0}$, gives the necessary conditions as

$$\frac{1}{2} \frac{\partial \Phi}{\partial \boldsymbol{\xi}} \Big|_{\hat{\xi}, \hat{v}, \hat{\lambda}} = -\mathbf{A}^T \hat{\boldsymbol{\lambda}} - \hat{\mathbf{V}}_\mathbf{A}^T \hat{\boldsymbol{\lambda}} = \mathbf{0}, \tag{4.13}$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial \mathbf{v}} \Big|_{\hat{\xi}, \hat{v}, \hat{\lambda}} = \mathbf{P}\hat{\mathbf{v}} - \left[\hat{\boldsymbol{\xi}}^T \otimes \mathbf{I}_n, -\mathbf{I}_n \right]^T \hat{\boldsymbol{\lambda}} = \mathbf{P}\hat{\mathbf{v}} - \hat{\mathbf{B}}^T \hat{\boldsymbol{\lambda}} = \mathbf{0}, \tag{4.14}$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial \boldsymbol{\lambda}} \Big|_{\hat{\xi}, \hat{v}, \hat{\lambda}} = \mathbf{y} - \mathbf{A}\hat{\boldsymbol{\xi}} - \hat{\mathbf{B}}\hat{\mathbf{v}} = \mathbf{0}, \tag{4.15}$$

where $\hat{\mathbf{B}}_{n \times n(u+1)} = \left[\hat{\boldsymbol{\xi}}^T \otimes \mathbf{I}_n, -\mathbf{I}_n \right]$ with the full row rank. The symbol 'hat' of the matrix $\hat{\mathbf{B}}$ is because of the use of the estimated parameter vector $\hat{\boldsymbol{\xi}}$.

The solution above represents the necessary condition for the minimum of the objective function. The sufficient condition is fulfilled for the residual vector by $\frac{1}{2}\frac{\partial\Phi}{\partial\mathbf{v}\mathbf{v}^T}=\mathbf{P}$ since the Hessian matrix of second derivatives of the objective function w.r.t. the extended residual vector is a positive semi-definite matrix. However, in the non-linear model all of the residuals and parameters as well as the auxiliary vector of Lagrange multipliers should be regarded as the (deterministic or stochastic) variables in the function (4.12). Only the semi-positive Hessian matrix of second derivatives of the objective function w.r.t. the extended residual vector cannot guarantee that the parameter estimates converge to the minimum of the objective function, because for this matrix it is not known whether it is positive definite or not. The analytical form and property of the matrix is not mentioned in the present literature and will be discussed in the later part of this chapter.

From Eq (4.14) we can have the estimated residual vector as follows

$$\hat{\mathbf{v}}=\begin{bmatrix}\hat{\mathbf{v}}_{\mathrm{A}}\\ \hat{\mathbf{v}}_{\mathrm{y}}\end{bmatrix}=\mathbf{Q}_{\mathrm{ll}}\hat{\mathbf{B}}^T\hat{\boldsymbol{\lambda}}, \tag{4.16}$$

Using Eq (4.16) in Eq (4.15) $\hat{\boldsymbol{\lambda}}$ can be expressed as follows

$$\hat{\boldsymbol{\lambda}}=\left(\hat{\mathbf{B}}\mathbf{Q}_{\mathrm{ll}}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\hat{\boldsymbol{\xi}}\right). \tag{4.17}$$

Then by reinserting Eq (4.17) to Eq (4.16) we have

$$\begin{aligned}\hat{\mathbf{v}}=\begin{bmatrix}\hat{\mathbf{v}}_{\mathrm{A}}\\ \hat{\mathbf{v}}_{\mathrm{y}}\end{bmatrix}&=\mathbf{Q}_{\mathrm{ll}}\hat{\mathbf{B}}^T\left(\hat{\mathbf{B}}\mathbf{Q}_{\mathrm{ll}}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\hat{\boldsymbol{\xi}}\right)\\ &=\begin{bmatrix}\left[\mathbf{Q}_{\mathrm{AA}}\quad \mathbf{Q}_{\mathrm{Ay}}\right]\hat{\mathbf{B}}^T\left(\hat{\mathbf{B}}\mathbf{Q}_{\mathrm{ll}}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\hat{\boldsymbol{\xi}}\right)\\ \left[\mathbf{Q}_{\mathrm{yA}}\quad \mathbf{Q}_{\mathrm{yy}}\right]\hat{\mathbf{B}}^T\left(\hat{\mathbf{B}}\mathbf{Q}_{\mathrm{ll}}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\hat{\boldsymbol{\xi}}\right)\end{bmatrix},\end{aligned} \tag{4.18}$$

where the residual vector can be obtained from the condition error adjustment problem $\mathbf{B}\mathbf{v}+\mathbf{w}=\mathbf{0}$ which is equivalent to $\mathbf{y}+\mathbf{v}_{\mathrm{y}}=\left(\mathbf{A}+\mathbf{V}_{\mathrm{A}}\right)\boldsymbol{\xi}$, if $\mathbf{B}=\left[\boldsymbol{\xi}^T\otimes\mathbf{I}_n,-\mathbf{I}_n\right]$ and $\mathbf{w}=-\left(\mathbf{y}-\mathbf{A}\boldsymbol{\xi}\right)$. This equation also indicates that

$$\hat{\mathbf{v}}^T\mathbf{P}\hat{\mathbf{v}}=\left(\mathbf{y}-\mathbf{A}\hat{\boldsymbol{\xi}}\right)^T\left(\hat{\mathbf{B}}\mathbf{Q}_{\mathrm{ll}}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\hat{\boldsymbol{\xi}}\right)=\hat{\boldsymbol{\lambda}}^T\hat{\mathbf{B}}\mathbf{Q}_{\mathrm{ll}}\hat{\mathbf{B}}^T\hat{\boldsymbol{\lambda}}, \tag{4.19}$$

which shows the equivalence of both minimization problems on the estimation level. Note that the matrix $\mathbf{A}$ is the original (observed) model matrix (not the estimated one). Furthermore, the Hessian matrix representing the second derivative of the target function w.r.t. the vector of Lagrange multipliers is $\hat{\mathbf{B}}\mathbf{Q}_{\mathrm{ll}}\hat{\mathbf{B}}^T$, which can be proven as positive definite matrix due to the full column rank of the transposed matrix $\hat{\mathbf{B}}^T$ (see Koch 1999 for detail). The identity of Eq (4.19) serves as the proof of the equivalence of both WTLS definition (the alternative definition is given in chapter 4.5).

We use the vectorization of a transposed vector to be the same vector and $vec\left(\mathbf{ABC}\right)=\left(\mathbf{C}^T\otimes\mathbf{A}\right)vec\left(\mathbf{B}\right)$ in order to obtain

$$\hat{\mathbf{V}}_{\mathrm{A}}^T\hat{\boldsymbol{\lambda}}=vec\left(\hat{\boldsymbol{\lambda}}^T\hat{\mathbf{V}}_{\mathrm{A}}\mathbf{I}_{\mathrm{u}}\right)=\left(\mathbf{I}_{\mathrm{u}}\otimes\hat{\boldsymbol{\lambda}}^T\right)\hat{\mathbf{v}}_{\mathrm{A}}. \tag{4.20}$$

With the help of Eqs (4.20), (4.13) and (4.17) we obtain

$$\left(\mathbf{I}_u\otimes\hat{\boldsymbol{\lambda}}^T\right)\hat{\mathbf{v}}_{\mathrm{A}}=\hat{\mathbf{V}}_{\mathrm{A}}^T\hat{\boldsymbol{\lambda}}=-\mathbf{A}^T\hat{\boldsymbol{\lambda}}=\mathbf{A}^T\left(\hat{\mathbf{B}}\mathbf{Q}_{\mathrm{ll}}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{A}\hat{\boldsymbol{\xi}}-\mathbf{y}\right). \tag{4.21}$$

Now, $\hat{\xi}$ can be derived by use of $\hat{\mathbf{V}}_{\mathbf{A}}^T\hat{\boldsymbol{\lambda}} = \mathbf{A}^T\left(\hat{\mathbf{B}}\mathbf{Q}_{ll}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{A}\hat{\xi}-\mathbf{y}\right)$ (see Eq (4.17)) as

$$\begin{aligned}\hat{\xi} &= \left(\mathbf{A}^T\left(\hat{\mathbf{B}}\mathbf{Q}_{ll}\hat{\mathbf{B}}^T\right)^{-1}\mathbf{A}\right)^{-1}\left(\hat{\mathbf{V}}_{\mathbf{A}}^T\hat{\boldsymbol{\lambda}} + \mathbf{A}^T\left(\hat{\mathbf{B}}\mathbf{Q}_{ll}\hat{\mathbf{B}}^T\right)^{-1}\mathbf{y}\right)\\ &= \left(\mathbf{A}^T\left(\hat{\mathbf{B}}\mathbf{Q}_{ll}\hat{\mathbf{B}}^T\right)^{-1}\mathbf{A}\right)^{-1}\left(\left(\mathbf{I}_u\otimes\hat{\boldsymbol{\lambda}}^T\right)\hat{\mathbf{v}}_{\mathbf{A}} + \mathbf{A}^T\left(\hat{\mathbf{B}}\mathbf{Q}_{ll}\hat{\mathbf{B}}^T\right)^{-1}\mathbf{y}\right)\\ &= \left(\mathbf{A}^T\left(\hat{\mathbf{B}}\mathbf{Q}_{ll}\hat{\mathbf{B}}^T\right)^{-1}\mathbf{A}\right)^{-1}\left(\left(\mathbf{I}_u\otimes\left(\left(\mathbf{y}-\mathbf{A}\hat{\xi}\right)^T\left(\hat{\mathbf{B}}\mathbf{Q}_{ll}\hat{\mathbf{B}}^T\right)^{-1}\right)\right)\hat{\mathbf{v}}_{\mathbf{A}} + \mathbf{A}^T\left(\hat{\mathbf{B}}\mathbf{Q}_{ll}\hat{\mathbf{B}}^T\right)^{-1}\mathbf{y}\right)\end{aligned} \quad . \qquad (4.22)$$

with $\left(\mathbf{I}_u\otimes\hat{\boldsymbol{\lambda}}^T\right)\hat{\mathbf{v}}_{\mathbf{A}} = \hat{\mathbf{V}}_{\mathbf{A}}^T\hat{\boldsymbol{\lambda}}$ (see Eq (4.21)) and $\hat{\boldsymbol{\lambda}} = \left(\hat{\mathbf{B}}\mathbf{Q}_{ll}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\hat{\xi}\right)$ (see Eq (4.17)).

If $\hat{\mathbf{V}}_{\mathbf{A}}$ is obtained through $\hat{\mathbf{V}}_{\mathbf{A}} = Ivec_{n\times u}(\hat{\mathbf{v}}_{\mathbf{A}})$, where the operator $Ivec_{n\times u}$ is the opposite of the 'vec' operator and reshapes the vector as the assigned matrix form (Matlab's reshape), one can compute this solution with other two closed-form expressions. From Eqs (4.13) and (4.17) we present the non-linear normal equation

$$\left(\mathbf{A}^T+\hat{\mathbf{V}}_{\mathbf{A}}^T\right)\hat{\boldsymbol{\lambda}} = \left(\mathbf{A}^T+\hat{\mathbf{V}}_{\mathbf{A}}^T\right)\left(\hat{\mathbf{B}}\mathbf{Q}_{ll}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{A}\hat{\xi}-\mathbf{y}\right) = \mathbf{0} \qquad (4.23)$$

leading to

$$\left(\mathbf{A}^T+\hat{\mathbf{V}}_{\mathbf{A}}^T\right)\left(\hat{\mathbf{B}}\mathbf{Q}_{ll}\hat{\mathbf{B}}^T\right)^{-1}\mathbf{A}\hat{\xi} = \left(\mathbf{A}^T+\hat{\mathbf{V}}_{\mathbf{A}}^T\right)\left(\hat{\mathbf{B}}\mathbf{Q}_{ll}\hat{\mathbf{B}}^T\right)^{-1}\mathbf{y}\,. \qquad (4.24)$$

Based on Eq (4.24) one solution should be

$$\hat{\xi} = \left(\left(\mathbf{A}^T+\hat{\mathbf{V}}_{\mathbf{A}}^T\right)\left(\hat{\mathbf{B}}\mathbf{Q}_{ll}\hat{\mathbf{B}}^T\right)^{-1}\mathbf{A}\right)^{-1}\left(\mathbf{A}^T+\hat{\mathbf{V}}_{\mathbf{A}}^T\right)\left(\hat{\mathbf{B}}\mathbf{Q}_{ll}\hat{\mathbf{B}}^T\right)^{-1}\mathbf{y}\,. \qquad (4.25)$$

If we add $\left(\mathbf{A}^T+\hat{\mathbf{V}}_{\mathbf{A}}^T\right)\left(\hat{\mathbf{B}}\mathbf{Q}_{ll}\hat{\mathbf{B}}^T\right)^{-1}\hat{\mathbf{V}}_{\mathbf{A}}\hat{\xi}$ to both sides of Eq (4.24), the other solution can be expressed as follows

$$\begin{aligned}&\left(\mathbf{A}^T+\hat{\mathbf{V}}_{\mathbf{A}}^T\right)\left(\hat{\mathbf{B}}\mathbf{Q}_{ll}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{A}+\hat{\mathbf{V}}_{\mathbf{A}}\right)\hat{\xi} = \left(\mathbf{A}^T+\hat{\mathbf{V}}_{\mathbf{A}}^T\right)\left(\hat{\mathbf{B}}\mathbf{Q}_{ll}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{y}+\hat{\mathbf{V}}_{\mathbf{A}}\hat{\xi}\right)\Rightarrow\\ &\hat{\xi} = \left(\left(\mathbf{A}^T+\hat{\mathbf{V}}_{\mathbf{A}}^T\right)\left(\hat{\mathbf{B}}\mathbf{Q}_{ll}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{A}+\hat{\mathbf{V}}_{\mathbf{A}}\right)\right)^{-1}\left(\mathbf{A}^T+\hat{\mathbf{V}}_{\mathbf{A}}^T\right)\left(\hat{\mathbf{B}}\mathbf{Q}_{ll}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{y}+\hat{\mathbf{V}}_{\mathbf{A}}\hat{\xi}\right)\end{aligned} \quad . \qquad (4.26)$$

In the non-linear problem, the estimated parameter vector cannot be separated from the predicted residual matrix $\hat{\mathbf{V}}_{\mathbf{A}}$ and even the estimated parameter vector per se. Based on the various closed-form expressions of the estimated parameter vector the algorithms can be designed correspondingly. Eq (4.22) using the original model matrix $\mathbf{A}$ is discussed in Björck et al (2000) and Schaffrin and Felus (2008) for the unweighted case, the latter of which shows that the algorithm converges monotonically (but slowly) to the solution. The algorithm based on the solution (4.26) can be interpreted as the standard Gauss-Newton algorithm (see chapter 4.6 for details), with which we expect linear convergence in general while this algorithm can also converge with a quadratic rate, provided that the initial values are close enough to the true solution. For the property of the convergence of the Gauss-Newton type we refer to Teunissen (1990) and Frandsen et al. (2004). The mixed form (4.25) of the original and estimated model matrix can be also applied to derive the desired solution through the iteration in this case although the exact convergence properties are still to be determined. It must be emphasized that in the non-linear case the normal matrix is not necessarily symmetric. Due to the non-convexity of the problem (Markovsky et al. 2006, Schuermans et al. 2007), the second derivative (the Hessian matrix) w.r.t. the parameter vector is not always positive definite. The sufficient condition for the optimization problem is not always fulfilled, so that the iterative procedures are not always guaranteed to converge to the minimum. The Hessian matrix representing the sufficient condition will be algebraically calculated in the next part of the chapter.

After deriving the general solutions, we resolve the solutions of WTLS and give the algebraic formulation based on the weighted LS solution. Firstly, the cofactor matrix is partitioned in two parts

$$\begin{aligned}\hat{\mathbf{B}}\mathbf{Q}_{\mathbf{ll}}\hat{\mathbf{B}}^T &= \left[\hat{\xi}^T \otimes \mathbf{I}_n, -\mathbf{I}_n\right]\begin{bmatrix}\mathbf{Q}_{\mathbf{AA}} & \mathbf{Q}_{\mathbf{Ay}} \\ \mathbf{Q}_{\mathbf{yA}} & \mathbf{Q}_{\mathbf{yy}}\end{bmatrix}\begin{bmatrix}\left(\hat{\xi}\otimes\mathbf{I}_n\right) \\ -\mathbf{I}_n\end{bmatrix} \\ &= \left(\hat{\xi}^T \otimes \mathbf{I}_n\right)\mathbf{Q}_{\mathbf{AA}}\left(\hat{\xi}\otimes\mathbf{I}_n\right) + \mathbf{Q}_{\mathbf{yA}}\left(\hat{\xi}\otimes\mathbf{I}_n\right) + \left(\hat{\xi}^T \otimes \mathbf{I}_n\right)\mathbf{Q}_{\mathbf{Ay}} + \mathbf{Q}_{\mathbf{yy}} \\ &= \mathbf{Q}_{\mathbf{yy}} + \hat{\mathbf{Q}}\end{aligned} \tag{4.27}$$

with $\hat{\mathbf{Q}} = \left(\hat{\xi}^T \otimes \mathbf{I}_n\right)\mathbf{Q}_{\mathbf{AA}}\left(\hat{\xi}\otimes\mathbf{I}_n\right) + \mathbf{Q}_{\mathbf{yA}}\left(\hat{\xi}\otimes\mathbf{I}_n\right) + \left(\hat{\xi}^T \otimes \mathbf{I}_n\right)\mathbf{Q}_{\mathbf{Ay}}$, which leads to

$$\begin{aligned}\left(\hat{\mathbf{B}}\mathbf{Q}_{\mathbf{ll}}\hat{\mathbf{B}}^T\right)^{-1} &= \left(\mathbf{Q}_{\mathbf{yy}} + \mathbf{I}_n\hat{\mathbf{Q}}\mathbf{I}_n\right)^{-1} \\ &= \mathbf{P}_{\mathbf{yy}} - \mathbf{P}_{\mathbf{yy}}\left(\hat{\mathbf{Q}}^{-1} + \mathbf{P}_{\mathbf{yy}}\right)^{-1}\mathbf{P}_{\mathbf{yy}} \\ &= \mathbf{P}_{\mathbf{yy}} - \Delta\hat{\mathbf{P}}_{\mathbf{yy}}\end{aligned} \quad . \tag{4.28}$$

using the matrix identity according to Koch (1999 p. 34).

Inserting Eq (4.28) into (4.23), the normal equation can be expressed as the following form

$$\begin{aligned}\mathbf{0} &= \left(\mathbf{A}^T + \hat{\mathbf{V}}_{\mathbf{A}}^T\right)\left(\mathbf{P}_{\mathbf{yy}} - \Delta\hat{\mathbf{P}}_{\mathbf{yy}}\right)\left(\mathbf{y} - \mathbf{A}\hat{\xi}\right) \\ &= \mathbf{A}^T\mathbf{P}_{\mathbf{yy}}\left(\mathbf{y} - \mathbf{A}\hat{\xi}\right) + \hat{\mathbf{V}}_{\mathbf{A}}^T\mathbf{P}_{\mathbf{yy}}\left(\mathbf{y} - \mathbf{A}\hat{\xi}\right) - \left(\mathbf{A}^T + \hat{\mathbf{V}}_{\mathbf{A}}^T\right)\Delta\hat{\mathbf{P}}_{\mathbf{yy}}\left(\mathbf{y} - \mathbf{A}\hat{\xi}\right) \\ &= \mathbf{A}^T\mathbf{P}_{\mathbf{yy}}\mathbf{y} - \mathbf{A}^T\mathbf{P}_{\mathbf{yy}}\mathbf{A}\hat{\xi} + \hat{\mathbf{V}}_{\mathbf{A}}^T\mathbf{P}_{\mathbf{yy}}\left(\mathbf{y} - \mathbf{A}\hat{\xi}\right) - \left(\mathbf{A}^T + \hat{\mathbf{V}}_{\mathbf{A}}^T\right)\Delta\hat{\mathbf{P}}_{\mathbf{yy}}\left(\mathbf{y} - \mathbf{A}\hat{\xi}\right)\end{aligned} \quad . \tag{4.29}$$

leading to

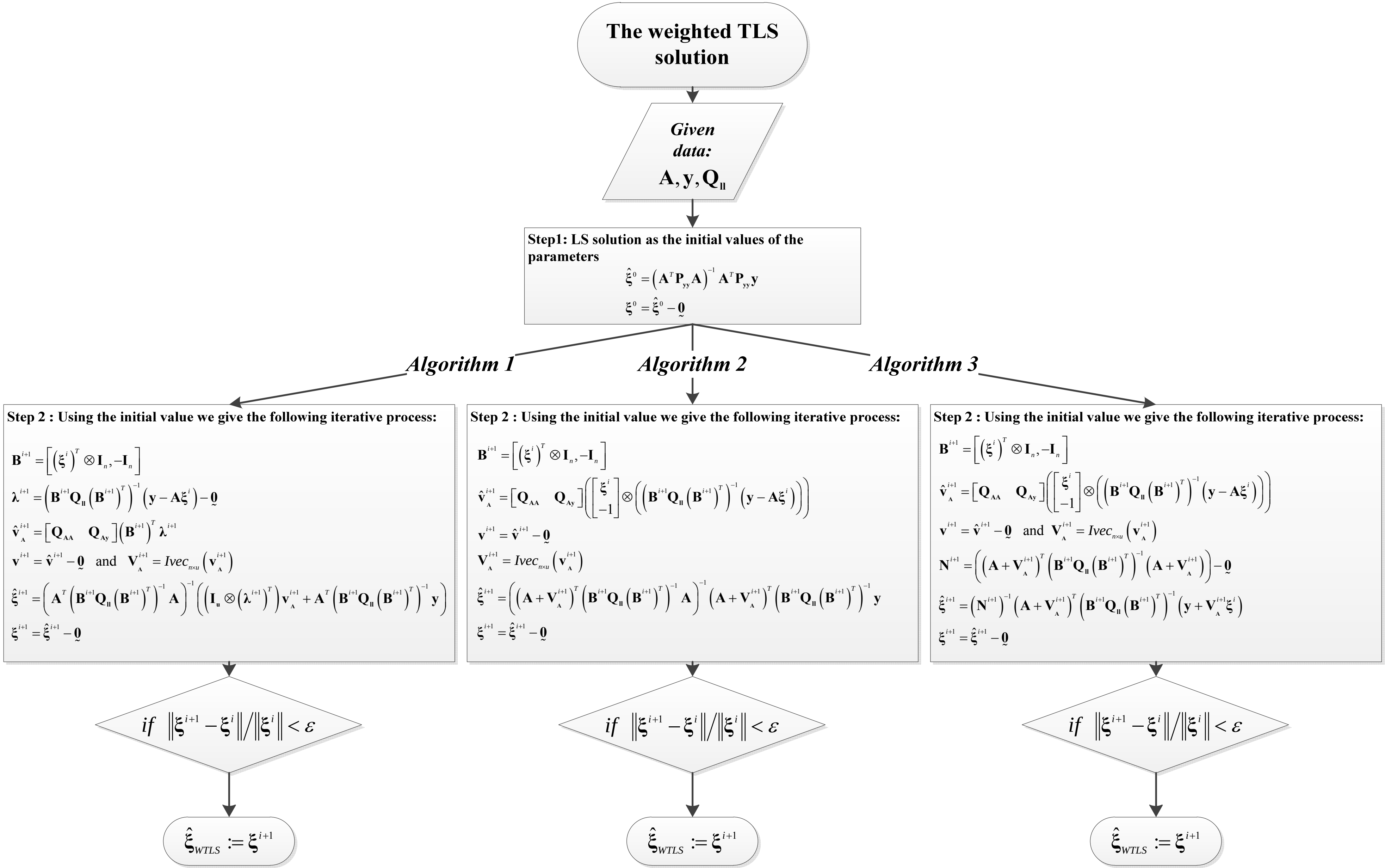
$$\mathbf{A}^T\mathbf{P}_{\mathbf{yy}}\mathbf{A}\hat{\xi} = \mathbf{A}^T\mathbf{P}_{\mathbf{yy}}\mathbf{y} + \hat{\mathbf{V}}_{\mathbf{A}}^T\mathbf{P}_{\mathbf{yy}}\left(\mathbf{y} - \mathbf{A}\hat{\xi}\right) - \left(\mathbf{A}^T + \hat{\mathbf{V}}_{\mathbf{A}}^T\right)\Delta\hat{\mathbf{P}}_{\mathbf{yy}}\left(\mathbf{y} - \mathbf{A}\hat{\xi}\right). \tag{4.30}$$

Based on Eq (4.30), we derive the WTLS solution formulated via the WLS solution

$$\hat{\xi} = \hat{\xi}_{LS} - \left(\mathbf{A}^T\mathbf{P}_{\mathbf{yy}}\mathbf{A}\right)^{-1}\left(\mathbf{A}^T + \hat{\mathbf{V}}_{\mathbf{A}}^T\right)\Delta\hat{\mathbf{P}}_{\mathbf{yy}}\left(\mathbf{y} - \mathbf{A}\hat{\xi}\right) + \left(\mathbf{A}^T\mathbf{P}_{\mathbf{yy}}\mathbf{A}\right)^{-1}\hat{\mathbf{V}}_{\mathbf{A}}^T\mathbf{P}_{\mathbf{yy}}\left(\mathbf{y} - \mathbf{A}\hat{\xi}\right). \tag{4.31}$$

which represents the difference of WLS and WTLS solution without any limitation of the covariance matrix of the extended observations.

On the basis of the formulas (4.22), (4.25) and (4.26) three algorithms can be designed here. The convergence depends on the initial approximation. The LS solution could be used as the start value to solve the EIV model (e.g., Schaffrin and Wieser 2008). In the following the index i is the number of the iteration step. The i'th (approximate) estimates are stripped of the randomness while keeping their numerical value. i.e. $\xi^{i+1} = \hat{\xi}^{i+1} - \underset{\sim}{\mathbf{0}}$ where the $\underset{\sim}{\mathbf{0}}$ denotes a random zero vector (or vector of 'pseudo-observation') of suitable size, in accordance with notion in Harville (1997). I.e. the deterministic term ξ^{i+1} is obtained by stripping the randomness of $\hat{\xi}^{i+1}$ (linear combinations of random variables) using $\underset{\sim}{\mathbf{0}}$ (see Schaffrin and Snow 2010 for detail). It is also for the estimated residual vector with $\mathbf{v}^{i+1} = \hat{\mathbf{v}}^{i+1} - \underset{\sim}{\mathbf{0}}$. ε is a sufficiently small threshold chosen to accomplish the iteration procedure. The variables e.g., $\mathbf{B}^{i+1}$ without the hat symbol mean that the random character is stripped. The three algorithms are designed in the following diagram:


The weighted TLS solution
Given data:
$\mathbf{A},\mathbf{y},\mathbf{Q}_{ll}$
Step1: LS solution as the initial values of the parameters
$\hat{\xi}^0=\left(\mathbf{A}^T\mathbf{P}_{yy}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{P}_{yy}\mathbf{y}$
$\xi^0=\hat{\xi}^0-\underset{\sim}{0}$
Algorithm 1
Algorithm 2
Algorithm 3
Step 2 : Using the initial value we give the following iterative process:
$\mathbf{B}^{i+1}=\left[\left(\xi^i\right)^T\otimes\mathbf{I}_n,-\mathbf{I}_n\right]$
$\lambda^{i+1}=\left(\mathbf{B}^{i+1}\mathbf{Q}_{ll}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi^i\right)-\underset{\sim}{0}$
$\hat{\mathbf{v}}_A^{i+1}=\left[\mathbf{Q}_{AA}\quad\mathbf{Q}_{Ay}\right]\left(\mathbf{B}^{i+1}\right)^T\lambda^{i+1}$
$\mathbf{v}^{i+1}=\hat{\mathbf{v}}^{i+1}-\underset{\sim}{0}$ and $\mathbf{V}_A^{i+1}=Ivec_{n\times u}\left(\mathbf{v}_A^{i+1}\right)$
$\hat{\xi}^{i+1}=\left(\mathbf{A}^T\left(\mathbf{B}^{i+1}\mathbf{Q}_{ll}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\mathbf{A}\right)^{-1}\left(\left(\mathbf{I}_u\otimes\left(\lambda^{i+1}\right)^T\right)\mathbf{v}_A^{i+1}+\mathbf{A}^T\left(\mathbf{B}^{i+1}\mathbf{Q}_{ll}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\mathbf{y}\right)$
$\xi^{i+1}=\hat{\xi}^{i+1}-\underset{\sim}{0}$
Step 2 : Using the initial value we give the following iterative process:
$\mathbf{B}^{i+1}=\left[\left(\xi^i\right)^T\otimes\mathbf{I}_n,-\mathbf{I}_n\right]$
$\hat{\mathbf{v}}_A^{i+1}=\left[\mathbf{Q}_{AA}\quad\mathbf{Q}_{Ay}\right]\left(\begin{bmatrix}\xi^i\\-1\end{bmatrix}\otimes\left(\left(\mathbf{B}^{i+1}\mathbf{Q}_{ll}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi^i\right)\right)\right)$
$\mathbf{v}^{i+1}=\hat{\mathbf{v}}^{i+1}-\underset{\sim}{0}$
$\mathbf{V}_A^{i+1}=Ivec_{n\times u}\left(\mathbf{v}_A^{i+1}\right)$
$\hat{\xi}^{i+1}=\left(\left(\mathbf{A}+\mathbf{V}_A^{i+1}\right)^T\left(\mathbf{B}^{i+1}\mathbf{Q}_{ll}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\mathbf{A}\right)^{-1}\left(\mathbf{A}+\mathbf{V}_A^{i+1}\right)^T\left(\mathbf{B}^{i+1}\mathbf{Q}_{ll}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\mathbf{y}$
$\xi^{i+1}=\hat{\xi}^{i+1}-\underset{\sim}{0}$
Step 2 : Using the initial value we give the following iterative process:
$\mathbf{B}^{i+1}=\left[\left(\xi^i\right)^T\otimes\mathbf{I}_n,-\mathbf{I}_n\right]$
$\hat{\mathbf{v}}_A^{i+1}=\left[\mathbf{Q}_{AA}\quad\mathbf{Q}_{Ay}\right]\left(\begin{bmatrix}\xi^i\\-1\end{bmatrix}\otimes\left(\left(\mathbf{B}^{i+1}\mathbf{Q}_{ll}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi^i\right)\right)\right)$
$\mathbf{v}^{i+1}=\hat{\mathbf{v}}^{i+1}-\underset{\sim}{0}$ and $\mathbf{V}_A^{i+1}=Ivec_{n\times u}\left(\mathbf{v}_A^{i+1}\right)$
$\mathbf{N}^{i+1}=\left(\left(\mathbf{A}+\mathbf{V}_A^{i+1}\right)^T\left(\mathbf{B}^{i+1}\mathbf{Q}_{ll}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\left(\mathbf{A}+\mathbf{V}_A^{i+1}\right)\right)-\underset{\sim}{0}$
$\hat{\xi}^{i+1}=\left(\mathbf{N}^{i+1}\right)^{-1}\left(\mathbf{A}+\mathbf{V}_A^{i+1}\right)^T\left(\mathbf{B}^{i+1}\mathbf{Q}_{ll}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\left(\mathbf{y}+\mathbf{V}_A^{i+1}\xi^i\right)$
$\xi^{i+1}=\hat{\xi}^{i+1}-\underset{\sim}{0}$
$if\ \left\|\xi^{i+1}-\xi^i\right\|/\left\|\xi^i\right\|<\varepsilon$
$if\ \left\|\xi^{i+1}-\xi^i\right\|/\left\|\xi^i\right\|<\varepsilon$
$if\ \left\|\xi^{i+1}-\xi^i\right\|/\left\|\xi^i\right\|<\varepsilon$
$\hat{\xi}_{WTLS}:=\xi^{i+1}$
$\hat{\xi}_{WTLS}:=\xi^{i+1}$
$\hat{\xi}_{WTLS}:=\xi^{i+1}$

4.4 Solution using non-linear GHM method

Alternatively, the EIV model can be adjusted through another class of the adjustment algorithm with linearization, namely the iterative non-linear GHM. The mixed model, which combines fixed parameters and random parameters, was introduced as a general case of the LS adjustment by F. Helmert and therefore it is often called GHM. For the nonlinear case $\mathbf{f}\left(E(\mathbf{l}),\xi\right)=\mathbf{0}$ one often uses the LS type non-linear GHM. A false linearization and update of model matrices may lead to an incorrect convergence to the true solution. The first successful investigation for solving the non-linear GHM was addressed by Pope (1972), and the rigorous algorithm for the non-linear GHM was also proposed by Lenzmann and Lenzmann (2004). The LS estimator is used also for solving the TLS and WTLS problem, respectively. The false linearization to solve the TLS problem applying the GHM occurs in Kupferer (2005) (see Neitzel and Petrovic 2008). Using the non-linear GHM Neitzel and Petrovic (2008) established the identity with the classical TLS solution for fitting a straight line. Csanyi May (2008) utilized it in a comprehensive performance analysis of state-of-the-art airborne mobile mapping systems. Recently, based on the non-linear GHM Schaffrin and Snow (2010) put forward the regularized TLS solution in Tykhonov's sense to solve the circle fitting problem. Neitzel (2010) applied it to solve the 2D similarity transformation. However, the solution using the LS estimator is not explicitly algebraically formulated. (I.e. the detailed structures of the model (Jacobian) matrices and the inconsistency vector, which are represented by the parameters and the observations as well as residuals, are not presented and built into the final solution of the non-linear GHM method.)

From Chapter 2.3 we know that the WTLS problem can be also classified as a non-linear GHM. The model variable $\mathbf{v}$ and parameter ξ mentioned in (2.27) and (2.28) can be set to $\mathbf{v}=\mathbf{0}$, $\xi=\xi^0$ (initial value) in the first step and $\mathbf{v}^i=\hat{\mathbf{v}}^i-\underset{\sim}{\mathbf{0}}$, $\xi^i=\hat{\xi}^i-\underset{\sim}{\mathbf{0}}$ (estimated value stripping of the randomness) in the i+1'th steps of the iterative process, respectively. The residual vector is estimated and used to update the first model (Jacobian) matrix $\left.\frac{\partial\mathbf{f}}{\partial\xi^T}\right|_{\mathbf{l}^0+\mathbf{v}^i,\xi^i}$ (see following). Here, we linearize the model $\mathbf{f}(\mathbf{l}+\mathbf{v},\xi)=\mathbf{0}$ as (e.g., Schaffrin and Snow 2010)

$$\left.\frac{\partial\mathbf{f}}{\partial\xi^T}\right|_{\mathbf{l}^0+\mathbf{v}^i,\xi^i}\left(\xi-\xi^i\right)+\left.\frac{\partial\mathbf{f}}{\partial\mathbf{l}^T}\right|_{\mathbf{l}^0+\mathbf{v}^i,\xi^i}\left(\mathbf{v}-\mathbf{v}^i\right)+\mathbf{f}\left(\mathbf{l}+\mathbf{v}^i,\xi^i\right)=\mathbf{0} \tag{4.32}$$

neglecting the terms of the higher order. Note that approximate values used for the position of linearization are not random. Thus, $\mathbf{l}^0=\mathbf{l}-\underset{\sim}{\mathbf{0}}$.

Eq (4.32) can be rewritten as the linear GHM as (e.g., Schaffrin and Snow 2010)

$$\left.\frac{\partial\mathbf{f}}{\partial\xi^T}\right|_{\mathbf{l}^0+\mathbf{v}^i,\xi^i}d\xi^{i+1}+\left.\frac{\partial\mathbf{f}}{\partial\mathbf{l}^T}\right|_{\mathbf{l}^0+\mathbf{v}^i,\xi^i}\mathbf{v}+\mathbf{f}\left(\mathbf{l},\xi^i\right)=\mathbf{0}, \tag{4.33}$$

with $d\xi^{i+1}=\xi-\xi^i$ through combining the two terms

$$\begin{aligned}&\mathbf{f}\left(\mathbf{l}+\mathbf{v}^i,\xi^i\right)-\left.\frac{\partial\mathbf{f}}{\partial\mathbf{l}^T}\right|_{\mathbf{l}^0+\mathbf{v}^i,\xi^i}\mathbf{v}^i\\&=\left(\mathbf{A}+\mathbf{V}_{\mathrm{A}}^i\right)\xi^i-\mathbf{y}-\mathbf{v}_{\mathrm{y}}^i-\left[\left(\xi^i\right)^T\otimes\mathbf{I}_n,-\mathbf{I}_n\right]\mathbf{v}^i\\&=\mathbf{A}\xi^i-\mathbf{y}=\mathbf{f}\left(\mathbf{l},\xi^i\right)\end{aligned} \tag{4.34}$$

The detailed structure of both Jacobian matrices and the inconsistency vector are explicitly expressed here as the algebraic formulation of the observations and parameters as well as the residuals

$$\left.\frac{\partial \mathbf{f}}{\partial \xi^T}\right|_{\mathbf{l}^0+\mathbf{v}^i,\xi^i} = \mathbf{A}+\mathbf{V}_{\mathbf{A}}^i$$

$$\left.\frac{\partial \mathbf{f}}{\partial \mathbf{l}^T}\right|_{\mathbf{l}^0+\mathbf{v}^i,\xi^i} = \left[\left(\xi^i\right)^T \otimes \mathbf{I}_n, -\mathbf{I}_n\right] = \mathbf{B}^{i+1}, \tag{4.35}$$

$$\mathbf{f}\left(\mathbf{l},\xi^i\right) = \mathbf{A}\xi^i - \mathbf{y} = \mathbf{w}^{i+1}$$

where $\mathbf{w}^i$ is the inconsistency vector in the i'th iteration step.

By inserting the model matrices and inconsistency vector into the model and the final solution, the alternative representation of the EIV model as linearized GHM reads

$$\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}^i\right)d\xi^{i+1} + \left[\left(\xi^i\right)^T \otimes \mathbf{I}_n, -\mathbf{I}_n\right]\begin{bmatrix}\mathbf{V}_{\mathbf{A}}^{i+1}\\ \mathbf{V}_{\mathbf{y}}^{i+1}\end{bmatrix} + \mathbf{A}\xi^i - \mathbf{y} = \mathbf{0}, \tag{4.36}$$

and the WTLS solution can be derived as follows

$$\begin{aligned}
\hat{\xi}^{i+1} &= \xi^i + d\hat{\xi}^{i+1} \\
&= \xi^i + \left(\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}^i\right)^T\left(\mathbf{B}^{i+1}\mathbf{Q}_{\mathbf{ll}}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}^i\right)\right)^{-1}\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}^i\right)^T\left(\mathbf{B}^{i+1}\mathbf{Q}_{\mathbf{ll}}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\left(-\mathbf{w}^{i+1}\right) \\
&= \left(\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}^i\right)^T\left(\mathbf{B}^{i+1}\mathbf{Q}_{\mathbf{ll}}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}^i\right)\right)^{-1}\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}^i\right)^T\left(\mathbf{B}^{i+1}\mathbf{Q}_{\mathbf{ll}}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi^i\right)+\xi^i \\
&= \left(\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}^i\right)^T\left(\mathbf{B}^{i+1}\mathbf{Q}_{\mathbf{ll}}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}^i\right)\right)^{-1}\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}^i\right)^T\left(\mathbf{B}^{i+1}\mathbf{Q}_{\mathbf{ll}}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi^i+\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}^i\right)\xi^i\right) \\
&= \left(\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}^i\right)^T\left(\mathbf{B}^{i+1}\mathbf{Q}_{\mathbf{ll}}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}^i\right)\right)^{-1}\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}^i\right)^T\left(\mathbf{B}^{i+1}\mathbf{Q}_{\mathbf{ll}}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\left(\mathbf{y}+\mathbf{V}_{\mathbf{A}}^i\xi^i\right)
\end{aligned} \quad ,(4.37)$$

where $\mathbf{V}_{\mathbf{A}}^0 = \mathbf{0}$, and $\mathbf{V}_{\mathbf{A}}^i$ can be reconstructed through

$$\hat{\mathbf{v}}_{\mathbf{A}}^i = \left[\mathbf{Q}_{\mathbf{AA}} \quad \mathbf{Q}_{\mathbf{Ay}}\right]\left(\mathbf{B}^{i+1}\right)^T\left(\mathbf{B}^{i+1}\mathbf{Q}_{\mathbf{ll}}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\left(-\mathbf{w}^{i+1}-\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}^i\right)d\xi^{i+1}\right), \tag{4.38}$$

Note that the residuals and parameters are updated with $\mathbf{v}^i = \hat{\mathbf{v}}^i - \underset{\sim}{\mathbf{0}}$, $\xi^{i+1} = \xi^i + d\hat{\xi}^{i+1} - \underset{\sim}{\mathbf{0}}$. i.e. the residuals are not added on the previous residuals predicted. In contrast to the update of residuals, parameters are accumulated at the last stage. Schaffrin and Felus (2008) mentioned that the non-linear GHM's inverse computation is required at every iteration because of much larger matrices. However, the solution (4.37) solves this difficulty, as we have reduced the matrix size in Eq (4.37). The process stops if the parameter or the extended residual vector does not change in the order of magnitude. The numerically obtained result is defined as the solution for the WTLS problem: $\hat{\xi}_{WTLS} := \xi^{i+1}$.

4.5 Solution based on the Newton type

In the previous part the solutions proposed are obtained with the help of the auxiliary Lagrange multipliers. However, it is broadly acknowledged that the method of Lagrange multipliers or the non-linear GHM (implicitly using Lagrange multipliers) yield only necessary conditions for optimality in the constrained problems. Although the sufficient condition is fulfilled for the residual vector by $\frac{1}{2}\frac{\partial \Phi}{\partial \mathbf{v}\mathbf{v}^T} = \mathbf{P}$, little importance has been attached to the study about the sufficient condition of WTLS problem for the parameter vector until now. For the fully weighted case we can define the TLS problem alternatively as the minimum of a sum of weighted squared residuals, namely

$\min_{\xi}(\mathbf{y}-\mathbf{A}\xi)^T(\mathbf{B}\mathbf{Q}_{\text{ll}}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)$. The target function can be also classified as quasi indirect error adjustment (Wolf 1968, Eq (2153, 10)). The objective function is similar to the objective function of the weighted LS case since only the weighted matrices are different. In this case the objective function is not subject to any constraints and the variables such as $\mathbf{V}_{\text{A}}$ and $\mathbf{v}_{\text{y}}$ as well as the auxiliary vector of the Lagrange multipliers disappear. The equivalence of both definitions of the WTLS problem has been already proved in Eq (4.19) on the estimation level.

The first and second derivatives of the objective function mentioned above w.r.t. the parameters are analytically presented. This represents the necessity and sufficiency conditions of the WTLS solution. On the basis of the analytical expression of the gradient and the Hessian matrix derived, the Newton algorithm is designed for the WTLS problem to bring some advantages versus the Gauss-Newton algorithm, which is widely employed to solve the WTLS problem, respectively. At the end of this part, the minimization problem is interpreted as the non-linear LS problem from the aspect of the numeric analysis. Based on the study the WTLS problem can be not only treated by the classical Newton or Gauss-Newton algorithm, but also by the other modified algorithms from numerical analysis.

The necessity condition for the local minimizer is given by

$$\frac{\partial f(\xi)}{\partial \xi}=\mathbf{0}, \tag{4.39}$$

where $f(\xi)=(\mathbf{y}-\mathbf{A}\xi)^T(\mathbf{B}\mathbf{Q}_{\text{ll}}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)$, where $\mathbf{B}=\left[\xi^T\otimes\mathbf{I}_n,-\mathbf{I}_n\right]$.

It is obvious that the function $(\mathbf{y}-\mathbf{A}\xi)^T(\mathbf{B}\mathbf{Q}_{\text{ll}}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)$ is scalar. The matrix analysis property (differentiation of a scalar function w.r.t. a vector) in Grafarend and Schaffrin (1993) is applied as follows

$$\frac{\partial f(\xi)}{\partial \xi}=\begin{bmatrix}\dfrac{\partial f(\xi)}{\partial \xi_1}\\ \ldots \\ \dfrac{\partial f(\xi)}{\partial \xi_k}\\ \ldots \\ \dfrac{\partial f(\xi)}{\partial \xi_u}\end{bmatrix}, \tag{4.40}$$

where $\dfrac{\partial f(\xi)}{\partial \xi_k}=\dfrac{\partial(\mathbf{y}-\mathbf{A}\xi)^T(\mathbf{B}\mathbf{Q}_{\text{ll}}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)}{\partial \xi_k}$ with $\xi=[\xi_1,\ldots,\xi_u]^T$ and $1\le k\le u$.

The first derivative w.r.t. the parameter vector can be extended in three parts according to the product rule (the well-known Leibniz's Law) as follows

$$\begin{aligned}&\frac{\partial(\mathbf{y}-\mathbf{A}\xi)^T(\mathbf{B}\mathbf{Q}_{\text{ll}}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)}{\partial \xi_k}\\ &=\frac{\partial(\mathbf{y}-\mathbf{A}\xi)^T}{\partial \xi_k}(\mathbf{B}\mathbf{Q}_{\text{ll}}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)+(\mathbf{y}-\mathbf{A}\xi)^T\frac{\partial(\mathbf{B}\mathbf{Q}_{\text{ll}}\mathbf{B}^T)^{-1}}{\partial \xi_k}(\mathbf{y}-\mathbf{A}\xi)\ldots\\ &\quad+(\mathbf{y}-\mathbf{A}\xi)^T(\mathbf{B}\mathbf{Q}_{\text{ll}}\mathbf{B}^T)^{-1}\frac{\partial(\mathbf{y}-\mathbf{A}\xi)}{\partial \xi_k}\end{aligned} \tag{4.41}$$

Since $(\mathbf{y}-\mathbf{A}\xi)^T(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)$ is scalar, it is not difficult to obtain

$$\begin{aligned}
&\frac{\partial(\mathbf{y}-\mathbf{A}\xi)^T(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)}{\partial\xi_k} \\
&=2\frac{\partial(\mathbf{y}-\mathbf{A}\xi)^T}{\partial\xi_k}(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)+(\mathbf{y}-\mathbf{A}\xi)^T\frac{\partial(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}}{\partial\xi_k}(\mathbf{y}-\mathbf{A}\xi), \qquad (4.42)\\
&=-2\mathbf{a}_k^T(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)+(\mathbf{y}-\mathbf{A}\xi)^T\frac{\partial(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}}{\partial\xi_k}(\mathbf{y}-\mathbf{A}\xi)
\end{aligned}$$

where $\mathbf{A}=\left[\mathbf{a}_{1,\dots}\mathbf{a}_u\right]$ (i.e. $\mathbf{a}_k$ is the k'th column of the matrix $\mathbf{A}$).

Then, the second part can be solved with $\frac{\partial\mathbf{A}^{-1}}{\partial\xi_k}=-\mathbf{A}^{-1}\frac{\partial\mathbf{A}}{\partial\xi_k}\mathbf{A}^{-1}$ (e.g., Grafarend and Schaffrin 1993) as

$$\begin{aligned}
&(\mathbf{y}-\mathbf{A}\xi)^T\frac{\partial(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}}{\partial\xi_k}(\mathbf{y}-\mathbf{A}\xi)=-(\mathbf{y}-\mathbf{A}\xi)^T(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}\frac{\partial(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)}{\partial\xi_k}(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)\\
&=-(\mathbf{y}-\mathbf{A}\xi)^T(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{Q}_k^T(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)-(\mathbf{y}-\mathbf{A}\xi)^T(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}\mathbf{Q}_k\mathbf{B}^T(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi). \qquad (4.43)\\
&=-2(\mathbf{y}-\mathbf{A}\xi)^T(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{Q}_k^T(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)
\end{aligned}$$

The last step of Eq (4.43) can use the following property: A scalar is given as $s=\mathbf{a}^T\mathbf{C}\mathbf{a}$ (exemplarily square matrix $\mathbf{C}$ and vector $\mathbf{a}$). Then, $\mathbf{a}^T\mathbf{C}\mathbf{a}=s=s^T=\mathbf{a}^T\mathbf{C}^T\mathbf{a}$. If a scalar is expressed as $s=\mathbf{a}^T\mathbf{C}\mathbf{b}$, $s\neq\mathbf{a}^T\mathbf{C}^T\mathbf{b}$ in general (unless the matrix $\mathbf{C}$ is symmetric).

We combine the Eqs (4.41) (4.42) and (4.43), and finally have the gradient (the first derivative w.r.t. the parameter vector)

$$\mathbf{g}=\frac{\partial f(\xi)}{\partial\xi}=-2(\mathbf{A}+\mathbf{A}^*)^T(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi), \qquad (4.44)$$

with $\mathbf{A}^*=\left[\mathbf{Q}_1\mathbf{B}^T(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi),\ \dots\ ,\mathbf{Q}_u\mathbf{B}^T(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)\right]$ through

$$\frac{\partial f(\xi)}{\partial\xi_k}=-2\mathbf{a}_k^T(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)-2(\mathbf{y}-\mathbf{A}\xi)^T(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{Q}_k^T(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi), \qquad (4.45)$$

Note that $\mathbf{Q}_k\hat{\mathbf{B}}^T(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)$ is the k'th column in the matrix $\mathbf{A}^*$.

Based on the analytical formulation of the gradient (4.44), obtaining the stationary point should be fulfilled the following condition

$$(\mathbf{A}+\mathbf{A}^*)^T(\hat{\mathbf{B}}\mathbf{Q}_{ll}\hat{\mathbf{B}}^T)^{-1}(\mathbf{y}-\mathbf{A}\hat{\xi})=\mathbf{0}. \qquad (4.46)$$

which is identical to the normal equation (4.23) solved by the Lagrange multipliers, since the matrix $\mathbf{A}^*$ is actually identical with $Ivec_{n\times u}\left(\left[\mathbf{Q}_{\mathbf{AA}}\quad\mathbf{Q}_{\mathbf{Ay}}\right]\hat{\mathbf{B}}^T(\hat{\mathbf{B}}\mathbf{Q}_{ll}\hat{\mathbf{B}}^T)^{-1}(\mathbf{y}-\mathbf{A}\hat{\xi})\right)$, namely $\hat{\mathbf{V}}_{\mathbf{A}}$.

Certainly, the normal equation can be identically gained from a different starting point. As well known by geodesists, the gradient per se is also an essential quantity to design the algorithms for non-linear adjustment problems. Based on it, the descent direction can be calculated, and one of the oldest iterative descent methods (the steepest descent method) for solving a minimization problem is established (see. e.g., Teunissen 1990). In the method $-\frac{\partial f(\xi)}{\partial\xi}$ is as the direction vector, and the positive scalar can be chosen according to the different line search strategies. The advantages and disadvantages are significant, which can also be found in Teunissen (1990).

After finding the necessary condition the second derivative representing the sufficiency condition for the local minimizer can be illustrated as

$$\frac{\partial^2 f(\xi)}{\partial\xi\partial\xi^T} = \frac{\partial^2 (\mathbf{y}-\mathbf{A}\xi)^T \left(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T\right)^{-1} (\mathbf{y}-\mathbf{A}\xi)}{\partial\xi\partial\xi^T} = \left[h_{kj}\right], \tag{4.47}$$

where $h_{kj} = \dfrac{\partial^2 (\mathbf{y}-\mathbf{A}\xi)^T \left(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T\right)^{-1} (\mathbf{y}-\mathbf{A}\xi)}{\partial\xi_k\partial\xi_j}$ with $1 \le j \le u$.

The Hessian matrix is calculated in Appendix. The result is here given as:

$$\mathbf{H} = 2\left(\mathbf{A}+\mathbf{A}^*+\mathbf{A}^{**}\right)^T \left(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T\right)^{-1} \left(\mathbf{A}+\mathbf{A}^*+\mathbf{A}^{**}\right) - 2\left[\varpi_{kj}\right]. \tag{4.48}$$

with $\varpi_{kj} = (\mathbf{y}-\mathbf{A}\xi)^T \left(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T\right)^{-1} \mathbf{Q}_{kj} \left(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T\right)^{-1} (\mathbf{y}-\mathbf{A}\xi)$

$$\mathbf{A}^{**} = \left[\mathbf{B}\mathbf{Q}_1^T \left(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T\right)^{-1} (\mathbf{y}-\mathbf{A}\xi), \quad \dots \quad ,\mathbf{B}\mathbf{Q}_u^T \left(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T\right)^{-1} (\mathbf{y}-\mathbf{A}\xi)\right].$$

Here, the analytical formulation of the second derivative of the objective function w.r.t. parameter vector is given, based on which the Newton algorithm can be designed for the non-linear problem as $d\xi = -\mathbf{H}^{-1}\mathbf{g}$. Being as a Newton's type, the algorithm works efficiently in the final stage of the iteration. If the Hessian matrix at the position of the solution is positive definite (the sufficient condition for the minimizer), in the region around the convergence value we can get quadratic convergence rate which is normally not the case with the Gauss-Newton method (see e.g., Frandsen et al. 2004). The profit about the convergence rate in comparison with the Gauss-Newton algorithm can be also found for solving the non-linear GMM in the geodetic literature, e.g., Lenzmann and Lenzmann (2007).

Except requiring analytical second order derivatives of the objective function the original Newton method has three further disadvantages: 1. It is not globally convergent for many problems, 2. It may converge towards a maximum or saddle point, 3. The system of linear equations to be solved in each iteration may be ill-conditioned or singular (see Frandsen et al. 2004). In contrary, some modified versions of the Newton method such as Levenberg-Marquardt type damped Newton method can perform successfully in general and avoid all the disadvantages above if the second order derivatives are analytical available. The Levenberg-Marquardt type damped Newton method (Levenberg 1944, Marquardt 1963) can be described as $d\xi = -(\mathbf{H}+\mu\mathbf{I})^{-1}\mathbf{g}$. By proper adjustment of the damping parameter μ the method combines the good qualities of the steepest descent method in the global part of the iteration process with the fast ultimate convergence of the Newton method. In Levenberg-Marquardt type the parameter μ is updated in each iteration (Frandsen et al. 2004). If the parameter is small, the method fall back to the standard Newton method. Of course, all modified Newton type methods (or said the combination of methods) based on the analytical formulated gradiant and Hessian matrix can be applied to solve the WTLS problem with their own strength.

4.6 WTLS solution of the standard Gauss-Newton type

Due to the positive definiteness of the matrix $\left(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T\right)^{-1}$, we rewrite the objective function as

$$f(\xi) = \mathbf{f}^T\mathbf{f} = (\mathbf{y}-\mathbf{A}\xi)^T \mathbf{C}_{1/2}^T\mathbf{C}_{1/2} (\mathbf{y}-\mathbf{A}\xi). \tag{4.49}$$

where $\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}=\mathbf{C}_{1/2}^T\mathbf{C}_{1/2}$ and $\mathbf{f}=\mathbf{C}_{1/2}\left(\mathbf{y}-\mathbf{A}\boldsymbol{\xi}\right)$. $\mathbf{C}_{1/2}$ can be derived by e.g., Cholesky decomposition or eigenvalue decomposition.

From Eqs (4.42) and (4.43) we can easily see that $\dfrac{\partial f\left(\boldsymbol{\xi}\right)}{\partial \xi_k}$ as follows

$$\frac{\partial f\left(\boldsymbol{\xi}\right)}{\partial \xi_k}=-2\left(\left(\mathbf{y}-\mathbf{A}\boldsymbol{\xi}\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{BQ}_i^T+\mathbf{a}_k{}^T\right)\mathbf{C}_{1/2}^T\mathbf{C}_{1/2}\left(\mathbf{y}-\mathbf{A}\boldsymbol{\xi}\right), \tag{4.50}$$

which also equals to

$$\frac{\partial \mathbf{f}^T\mathbf{f}}{\partial \xi_k}=\frac{\partial \mathbf{f}^T}{\partial \xi_k}\mathbf{f}+\mathbf{f}^T\frac{\partial \mathbf{f}}{\partial \xi_k}=2\frac{\partial \mathbf{f}^T}{\partial \xi_k}\mathbf{f}=2\frac{\partial \mathbf{f}^T}{\partial \xi_k}\mathbf{C}_{1/2}\left(\mathbf{y}-\mathbf{A}\boldsymbol{\xi}\right). \tag{4.51}$$

We combine Eqs (4.50) and (4.51), and derive

$$\frac{\partial \mathbf{f}^T}{\partial \xi_k}=-\left(\left(\mathbf{y}-\mathbf{A}\boldsymbol{\xi}\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{BQ}_k^T+\mathbf{a}_k{}^T\right)\mathbf{C}_{1/2}^T \tag{4.52}$$

which yields the transposed Jacobian matrix

$$\mathbf{J}^T\left(\boldsymbol{\xi}\right)=\frac{\partial \mathbf{f}^T}{\partial \boldsymbol{\xi}}=-\left(\mathbf{A}+\mathbf{A}^*\right)^T\mathbf{C}_{1/2}^T. \tag{4.53}$$

The classical Gauss-Newton method can be applied here to solve the WTLS problem as follows

$$\begin{aligned}d\hat{\boldsymbol{\xi}}&=-\left(\mathbf{J}^T\left(\hat{\boldsymbol{\xi}}\right)\mathbf{J}\left(\hat{\boldsymbol{\xi}}\right)\right)^{-1}\mathbf{J}^T\left(\hat{\boldsymbol{\xi}}\right)\mathbf{f}\\&=\left(\left(\mathbf{A}+\hat{\mathbf{V}}_{\mathbf{A}}\right)^T\left(\hat{\mathbf{B}}\mathbf{Q}_{\text{ll}}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{A}+\hat{\mathbf{V}}_{\mathbf{A}}\right)\right)^{-1}\left(\mathbf{A}+\hat{\mathbf{V}}_{\mathbf{A}}\right)^T\left(\hat{\mathbf{B}}\mathbf{Q}_{\text{ll}}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\hat{\boldsymbol{\xi}}\right)\end{aligned} \tag{4.54}$$

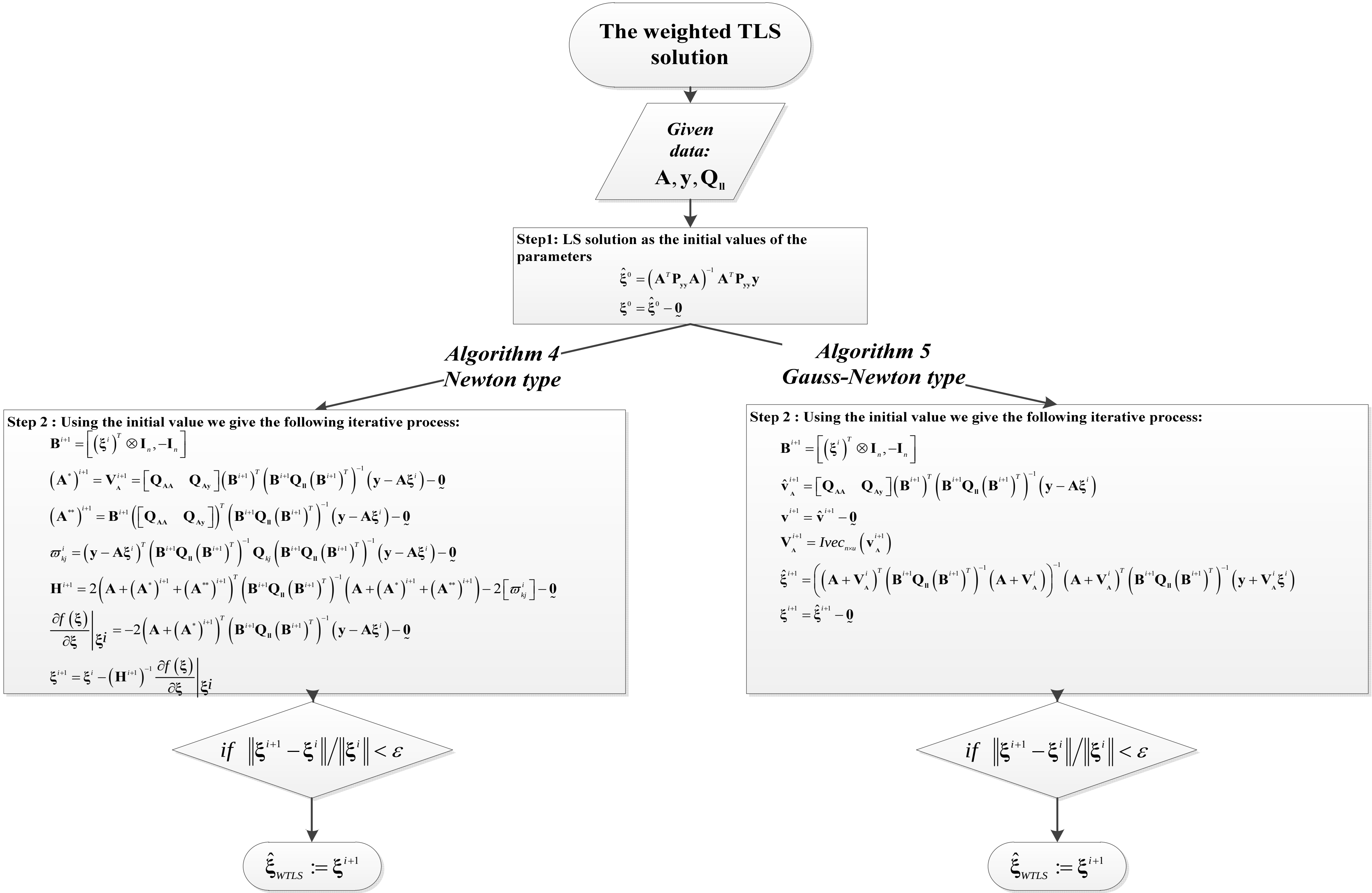
leading to

$$\begin{aligned}\hat{\boldsymbol{\xi}}&:=\hat{\boldsymbol{\xi}}+d\hat{\boldsymbol{\xi}}\\&=\left(\left(\mathbf{A}+\hat{\mathbf{V}}_{\mathbf{A}}\right)^T\left(\hat{\mathbf{B}}\mathbf{Q}_{\text{ll}}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{A}+\hat{\mathbf{V}}_{\mathbf{A}}\right)\right)^{-1}\left(\mathbf{A}+\hat{\mathbf{V}}_{\mathbf{A}}\right)^T\left(\hat{\mathbf{B}}\mathbf{Q}_{\text{ll}}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{y}+\hat{\mathbf{V}}_{\mathbf{A}}\hat{\boldsymbol{\xi}}\right),\end{aligned} \tag{4.55}$$

which is identical to our solution (4.26) and (4.37).

As mentioned above, the WTLS problem is interpreted as a non-linear LS problem instead of a normal unconstrained optimization problem. In this case it is not necessary to design an extra iteration procedure for the solution (4.55), because the third and fourth algorithms using the original Gauss-Newton method have already provided the iteration processes for the solution itself. The original Gauss-Newton method can be modified by, e.g., the well-known damped Gauss-Newton method, namely Levenberg-Marquardt method, which is more suitable than the Gauss-Newton algorithm if the current iterate is far from the desired solution.

The computational advantages of TLS problem (also in weighted case) are still largely unknown in the statistical community (Markovsky and Van Huffel 2007). However, with this contribution we have a more thorough understanding of the WTLS problem since the problem interpreted as the unconstrained optimization problem as well as the non-linear LS problem can be settled by numerous approaches from the numerical analysis aspect. It must be noted that the non-linear LS problem can be classified as the quasi indirect error adjustment problem (see Wolf p. 105), because the rank and the row number of the matrix $\hat{\mathbf{B}}$ are identical. If the row number larger than the rank, the problem is attributed to the general case of the problem of the mixed model adjustment.


The weighted TLS solution
Given data:
$\mathbf{A}, \mathbf{y}, \mathbf{Q}_{ll}$
Step1: LS solution as the initial values of the parameters
$\hat{\xi}^0 = \left(\mathbf{A}^T \mathbf{P}_{yy} \mathbf{A}\right)^{-1} \mathbf{A}^T \mathbf{P}_{yy} \mathbf{y}$
$\xi^0 = \hat{\xi}^0 - \underset{\sim}{\mathbf{0}}$
Algorithm 4
Newton type
Algorithm 5
Gauss-Newton type
Step 2 : Using the initial value we give the following iterative process:
$\mathbf{B}^{i+1} = \left[\left(\xi^i\right)^T \otimes \mathbf{I}_n, -\mathbf{I}_n\right]$
$\left(\mathbf{A}^*\right)^{i+1} = \mathbf{V}_{\mathbf{A}}^{i+1} = \left[\mathbf{Q}_{\mathbf{AA}} \quad \mathbf{Q}_{\mathbf{Ay}}\right]\left(\mathbf{B}^{i+1}\right)^T \left(\mathbf{B}^{i+1}\mathbf{Q}_{ll}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\left(\mathbf{y} - \mathbf{A}\xi^i\right) - \underset{\sim}{\mathbf{0}}$
$\left(\mathbf{A}^{**}\right)^{i+1} = \mathbf{B}^{i+1}\left(\left[\mathbf{Q}_{\mathbf{AA}} \quad \mathbf{Q}_{\mathbf{Ay}}\right]\right)^T \left(\mathbf{B}^{i+1}\mathbf{Q}_{ll}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\left(\mathbf{y} - \mathbf{A}\xi^i\right) - \underset{\sim}{\mathbf{0}}$
$\varpi_{kj}^i = \left(\mathbf{y} - \mathbf{A}\xi^i\right)^T \left(\mathbf{B}^{i+1}\mathbf{Q}_{ll}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1} \mathbf{Q}_{kj} \left(\mathbf{B}^{i+1}\mathbf{Q}_{ll}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\left(\mathbf{y} - \mathbf{A}\xi^i\right) - \underset{\sim}{\mathbf{0}}$
$\mathbf{H}^{i+1} = 2\left(\mathbf{A} + \left(\mathbf{A}^*\right)^{i+1} + \left(\mathbf{A}^{**}\right)^{i+1}\right)^T \left(\mathbf{B}^{i+1}\mathbf{Q}_{ll}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\left(\mathbf{A} + \left(\mathbf{A}^*\right)^{i+1} + \left(\mathbf{A}^{**}\right)^{i+1}\right) - 2\left[\varpi_{kj}^i\right] - \underset{\sim}{\mathbf{0}}$
$\left.\frac{\partial f(\xi)}{\partial \xi}\right|_{\xi^i} = -2\left(\mathbf{A} + \left(\mathbf{A}^*\right)^{i+1}\right)^T \left(\mathbf{B}^{i+1}\mathbf{Q}_{ll}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\left(\mathbf{y} - \mathbf{A}\xi^i\right) - \underset{\sim}{\mathbf{0}}$
$\xi^{i+1} = \xi^i - \left(\mathbf{H}^{i+1}\right)^{-1} \left.\frac{\partial f(\xi)}{\partial \xi}\right|_{\xi^i}$
Step 2 : Using the initial value we give the following iterative process:
$\mathbf{B}^{i+1} = \left[\left(\xi^i\right)^T \otimes \mathbf{I}_n, -\mathbf{I}_n\right]$
$\hat{\mathbf{v}}_{\mathbf{A}}^{i+1} = \left[\mathbf{Q}_{\mathbf{AA}} \quad \mathbf{Q}_{\mathbf{Ay}}\right]\left(\mathbf{B}^{i+1}\right)^T \left(\mathbf{B}^{i+1}\mathbf{Q}_{ll}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\left(\mathbf{y} - \mathbf{A}\xi^i\right)$
$\mathbf{v}^{i+1} = \hat{\mathbf{v}}^{i+1} - \underset{\sim}{\mathbf{0}}$
$\mathbf{V}_{\mathbf{A}}^{i+1} = Ivec_{n \times u}\left(\mathbf{v}_{\mathbf{A}}^{i+1}\right)$
$\hat{\xi}^{i+1} = \left(\left(\mathbf{A} + \mathbf{V}_{\mathbf{A}}^i\right)^T \left(\mathbf{B}^{i+1}\mathbf{Q}_{ll}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\left(\mathbf{A} + \mathbf{V}_{\mathbf{A}}^i\right)\right)^{-1}\left(\mathbf{A} + \mathbf{V}_{\mathbf{A}}^i\right)^T \left(\mathbf{B}^{i+1}\mathbf{Q}_{ll}\left(\mathbf{B}^{i+1}\right)^T\right)^{-1}\left(\mathbf{y} + \mathbf{V}_{\mathbf{A}}^i \xi^i\right)$
$\xi^{i+1} = \hat{\xi}^{i+1} - \underset{\sim}{\mathbf{0}}$
$if \quad \left\|\xi^{i+1} - \xi^i\right\| / \left\|\xi^i\right\| < \varepsilon$
$if \quad \left\|\xi^{i+1} - \xi^i\right\| / \left\|\xi^i\right\| < \varepsilon$
$\hat{\xi}_{WTLS} := \xi^{i+1}$
$\hat{\xi}_{WTLS} := \xi^{i+1}$

4.7 Comparison between the general solutions and the existing solutions

Comparison with the solution of the non-linear GHM method proposed in Neitzel (2010) and Schaffrin and Snow (2010)

As mentioned in chapter 4.4, the EIV model is adjusted by the iterative GHM method proposed by e.g., Schaffrin and Snow (2010) and Neitzel (2010). Being a non-linear adjustment problem, the various solutions are presented based on the different principles. Lenzmann and Lenzmann (2007) pointed out that there are two different ways for solving problems of non-linear adjustment. In one way, one linearizes all non-linear functions and solves the minimization problem of least squares for this approximated linear model. In the other way, one takes the non-linear minimization problem into account without any previous linearizations. For the WTLS problem presented in this thesis it is obvious that our third algorithm (Eq (4.26)) has the identical analytical formulation as the non-linear GHM method. However, Schaffrin (2007) pointed out that a LS solution of the non-linear GHM may be formed by proper iterative linearization and can be tricky as was pointed out already by Pope (1972), followed by another discussion by many authors. On the other hand, the solution of linear normal equation for iteratively linearized GHM only defines one particular algorithm to solve the non-linear normal equations for the GHM. The argument has actually been brought forward by Lenzmann and Lenzmann (2007) for the non-linear GMM and is equally valid for the present context.

Although the Algorithm 3 and the iterative GHM method have the same analytical formulation of the estimated parameters, the iterative processes of the Algorithm 3 have a different design from non-linear GHM. The i'th step of the iterations can be expressed as follows

The iterative process of the iterative GHM algorithm can be described as follows: $\mathbf{B}^{i+1}=\left[\left(\xi^{i}\right)^{T}\otimes\mathbf{I}_{n},-\mathbf{I}_{n}\right]$,

$$\xi^{i+1}=\left(\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}^{i}\right)^{T}\left(\mathbf{B}^{i+1}\mathbf{Q}_{\mathbf{ll}}\left(\mathbf{B}^{i+1}\right)^{T}\right)^{-1}\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}^{i}\right)\right)^{-1}\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}^{i}\right)^{T}\left(\mathbf{B}^{i+1}\mathbf{Q}_{\mathbf{ll}}\left(\mathbf{B}^{i+1}\right)^{T}\right)^{-1}\left(\mathbf{y}+\mathbf{V}_{\mathbf{A}}^{i}\xi^{i}\right)$$

$$\mathbf{v}_{\mathbf{A}}^{i+1}=\begin{bmatrix}\mathbf{Q}_{\mathbf{AA}} & \mathbf{Q}_{\mathbf{Ay}}\end{bmatrix}\left(\mathbf{B}^{i+1}\right)^{T}\left(\mathbf{B}^{i+1}\mathbf{Q}_{\mathbf{ll}}\left(\mathbf{B}^{i+1}\right)^{T}\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi^{i}-\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}^{i}\right)d\xi^{i+1}\right)\text{ and }\mathbf{V}_{\mathbf{A}}^{i+1}=Ivec_{n\times u}\left(\mathbf{v}_{\mathbf{A}}^{i+1}\right)$$

The iterative process (the Algorithm 3) reads: $\mathbf{B}^{i+1}=\left[\left(\xi^{i}\right)^{T}\otimes\mathbf{I}_{n},-\mathbf{I}_{n}\right]$

$$\mathbf{v}_{\mathbf{A}}^{i+1}=\begin{bmatrix}\mathbf{Q}_{\mathbf{AA}} & \mathbf{Q}_{\mathbf{Ay}}\end{bmatrix}\left(\mathbf{B}^{i+1}\right)^{T}\left(\mathbf{B}^{i+1}\mathbf{Q}_{\mathbf{ll}}\left(\mathbf{B}^{i+1}\right)^{T}\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi^{i}\right)\text{ and }\mathbf{V}_{\mathbf{A}}^{i+1}=Ivec_{n\times u}\left(\mathbf{v}_{\mathbf{A}}^{i+1}\right)$$

$$\xi^{i+1}=\left(\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}^{i+1}\right)^{T}\left(\mathbf{B}^{i+1}\mathbf{Q}_{\mathbf{ll}}\left(\mathbf{B}^{i+1}\right)^{T}\right)^{-1}\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}^{i+1}\right)\right)^{-1}\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}^{i+1}\right)^{T}\left(\mathbf{B}^{i+1}\mathbf{Q}_{\mathbf{ll}}\left(\mathbf{B}^{i+1}\right)^{T}\right)^{-1}\left(\mathbf{y}+\mathbf{V}_{\mathbf{A}}^{i+1}\xi^{i}\right)$$

The residual vector $\mathbf{v}_{\mathbf{A}}^{i+1}$ estimated in the iteration has already been considered in the same iteration of the parameter vector estimated whereas in the non-linear GHM algorithm the parameter vector is always derived through the residual matrix $\mathbf{V}_{\mathbf{A}}^{i}$ of the last iteration. Furthermore, the expressions of the residual vector are different. One is based on the iterative approach without linearization (see Eq (4.18)), the other is based on the iterative linearized GHM method (see Eq (4.38)). The difference between using linearization or not, the order of updating parameter or residuals, the formulas of the residual vector can influence the convergence behavior e.g., the iteration times to the convergence point. However, both will converge to the same end, i.e. they have identical results. The different iteration times and identical result will be seen in the numerical example (Chapter 6).

Comparison with the fairly weighted TLS methods proposed in Schaffrin and Wieser (2008)

The fairly weighted TLS methods proposed in Schaffrin and Wieser (2008) belongs to the EW-TLS according to the structure of the vcm. In this EW-TLS case the cofactor matrix can be expressed as follows (see Schaffrin and Wieser 2008)

$$\mathbf{Q}_{\mathrm{ll}} = \begin{bmatrix} \underset{u\times u}{\mathbf{Q}_0} \otimes \underset{n\times n}{\mathbf{Q}_{\mathrm{xx}}} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{\mathrm{yy}} \end{bmatrix}. \tag{4.56}$$

Certainly, the cofactor matrix is not fully populated. In Schaffrin and Wieser (2008) the sufficiency condition of the minimization problem is discussed, which has not been widely mentioned in the other TLS literature. However, the second derivatives of the target function w.r.t. the residual vector cannot completely represent the sufficiency condition, because the objective function is related to the parameter vector, which is indispensable in the final solution.

Comparison with the iterative TLS solution proposed in Shen et al. (2010)

Shen et al. (2010) uses the Gauss-Newton strategy after the linearization to solve the TLS problem proposed originally by Pope (1974). The algorithm has the identical form with the iterative GHM algorithm (Note that the third algorithm proposed in the thesis has the same analytical expression of the estimated parameter vector but the different algorithm design from the iterative GHM method). However, the method proposed by Pope (1974) and the iterative GHM method presented by Pope (1972) are almost identical in principle, since the difference is only ascribed to forming the solution to the GHM or not.

Comparison with the existing (weighted) TLS solutions in general

In this thesis three different types of the normal matrix are presented according the thorough analysis of the normal equation system. The difference to the LS solution is also analytically given. If the vcm does not have the form of Kronecker product, the solution of the weighted TLS problem is not more given by the SVD approach (see chapter 4.2). The existing algorithms (e.g., Neitzel (2010), Schaffrin and Snow (2010), Shen et al. (2010)) follow the same pattern, where the Gauss-Newton algorithm is used to solve the weighted TLS problem, because the method class takes advantage of the special structure (sum of squares) of the objective function. In this thesis except obtaining the solutions of the standard or modified Gauss-Newton type, a more efficient Newton algorithm is analytically derived based on the quadratic approximation of the loss function. As a consequence, if the loss function is quadratic, Newton's method locates the minimum in one iteration step, whereas the method based on the linear approximation needs in general an infinite number of iteration steps (Teunissen 1990). The character of the convergence performance will be demonstrated in Chapter 6. Furthermore, the standard Newton algorithm may be modified by combining with other algorithms to exhibit their own advantages.

5 Extension of the weighted TLS problem

In some geodetic applications, the reality cannot be explicitly modeled using the standard EIV model. Hence, some problems extended from the standard WTLS problem should be discussed in this chapter.

The WTLS problem with fixing columns is taken into account, where the model matrix $\mathbf{A}$ with fixing columns in the WTLS problem can be separated into two parts, namely the deterministic and stochastic columns, respectively. This approach is frequently used for, e.g., the regression model, where the column corresponding to the intercept parameter is fixed, and the transformation model, where the column for the translations is deterministic. One approach has been developed by van Huffel and Vandewalle (1991, p. 84), who use the QR decomposition to eliminate the parameter corresponding to the fixed columns (see Chapter 3). Another approach is that one may shift all coordinates to the center of mass by multiplying both sides of the equation with the idempotent centering matrix $\mathbf{\Psi} = \mathbf{I}_n - (1/n)\mathbf{1}_n\mathbf{1}_n^T$, where $\mathbf{1}_n$ is the vector of ones (see e.g., Shen et al. (2006), Schaffrin and Felus (2008), Felus and Burtch (2009)). However, the idempotent matrix is valid only for the equally weighted case. Considering the general vcm, the parameters corresponding to the fixed columns should be eliminated based on the normal equation system. In the more general case fixing elements may be also considered, where some elements (not column-wise) in the model matrix $\mathbf{A}$ are known exactly in the WTLS problem. The problem is not investigated in the present literature. The possible approach for the problem is the iterative GHM method, where one picks up the stochastic elements of the model matrix $\mathbf{A}$ and the observation vector to be arranged in the extended residual vector and updated with the predicted values.

The constrained adjustment problem is also an important issue in geodesy. This procedure is encountered in numerous geodetic research areas, e.g., the free net adjustment (Koch 1999), in the adjustment of gravity measurements (Hwang et al., 2002), in the Very Long Baseline Interferometry analysis (Kutterer 2003) or in integrated navigation (Yang 2010). For the framework of the TLS problem, Yeredor (2006) claims that the TLS problem with constraints holds an essential position in various parameter estimation problems. For the equally weighted TLS problem with linear, fiducial and quadratic constraints we refer to Schaffrin and Felus (2005), Schaffrin (2006) and Schaffrin and Felus (2009), in which the solutions are obtained based on the SVD. However, the WTLS problem with the linear constraints, which cannot be obtained by matrix decompositions, is not yet well understood. To solve the constrained problems Lagrange multipliers are applied in this section to attain the correspondent solutions.

In the standard EIV model the parameters and the conventional observations are expressed as the vectors. In some geodetic applications, e.g., affine transformation, the parameters and the conventional observations are always expressed matrix-wise in the model. The TLS solution with identity (Schaffrin and Felus 2008, Kwon et al., 2009) and the multivariate TLS solution (Schaffrin and Wieser 2009) was successfully used for 2D affine transformation. However, the WTLS cannot be defined as multivariate TLS problem, in which the errors can be expressed as a Kronecker product of two positive definite matrices. In order to solve the 2D, 3D affine transformation in the heteroscedastic case, the WTLS solution with the parameter matrix is presented here.

Undoubtly, the iterative GHM method can solve a lot of non-linear LS problems due to its simplicity. In this procedure one linearizes the non-linear function in the GHM. However, in many cases one must generalize a strategy to solve some geodetic applications: 1. The solution for the non-linear problem with the quadratic form should be regarded as the nonlinear TLS problem with the linear constraints about the paramters (see Chapter 6 for detail). 2. Although the 7-parameter transformation encountered to convert the World Geodetic system 84 and local coordinate is widely discussed (e.g. Teunissen 1988, Grafarend and Awange 2003, Shen et al. 2006, Felus and Burtch

2009), the solution for the problem with the general vcm is still not available. For the problem one can explicitly use the iterative GHM method (using the partial derivative w.r.t. the rotation angles) or solve it as the non-linear TLS problem with the non-linear (here, quadratic due to the orthogonality of the rotation matrix) constraints about the parameters. If one or more baselines are fixed, the constraints about the observations estimated may be taken into account. 3. In geodetic networks the TLS solution for the non-linear problem is discussed in Reinking (2008), however no one puts forward the solution if the stochastic parameters are available. Thus, the non-linear TLS problem with the linear and non-linear constraints about the parameter or observations should be considered, even at the same time. In addition, the proper modeling and solution for the non-linear problems with the stochastic parameter should be given. These applications and the state of research will be demonstrated in the Chapter 6 in detail.

5.1 Fixing columns

In the Chapter 3 we have reviewed fixing columns of the design matrix in the EIV model, which is alternatively called the mixed LS-TLS problem. In this part fixing columns in the fully weighted case will be solved. The objective function and the model is expressed as follows

$$\begin{aligned} &\min \mathbf{v}^T\mathbf{P}\mathbf{v} \\ &\textit{subject to } \mathbf{y}+\mathbf{v}_{\mathbf{y}} = \left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}\right)\boldsymbol{\xi} = \mathbf{A}_1\boldsymbol{\xi}_1 + \left(\mathbf{A}_2+\mathbf{V}_{\mathbf{A}_2}\right)\boldsymbol{\xi}_2 \end{aligned} \tag{5.1}$$

where $\mathbf{v}=\left[vec^T\mathbf{V}_{\mathbf{A}_2} \quad \mathbf{v}_{\mathbf{y}}^T\right]^T$. $\mathbf{A}=\left[\mathbf{A}_1,\mathbf{A}_2\right]$, $\mathbf{V}_{\mathbf{A}}=\left[\mathbf{0},\mathbf{V}_2\right]$, $\boldsymbol{\xi}=\left[\boldsymbol{\xi}_1;\boldsymbol{\xi}_2\right]$. The $\boldsymbol{\xi}_2$ vector is the subparameter vector with $u-t$ elements corresponding to the sub matrix $\mathbf{A}_2$; $\boldsymbol{\xi}_1$ with t elements is corresponding to the deterministic part $\mathbf{A}_1$ of $\mathbf{A}$.

To obtain the result the Lagrange multiplier is applied as follows

$$\Phi\left(\mathbf{v},\boldsymbol{\lambda},\boldsymbol{\xi}\right) = \mathbf{v}^T\mathbf{P}\mathbf{v} + 2\boldsymbol{\lambda}^T\left(\mathbf{y}-\mathbf{A}\boldsymbol{\xi}-\mathbf{V}_{\mathbf{A}_2}\boldsymbol{\xi}_2+\mathbf{v}_{\mathbf{y}}\right). \tag{5.2}$$

The corresponding necessary conditions for the stationary point read

$$\frac{1}{2}\frac{\partial\Phi}{\partial\boldsymbol{\xi}}\Big|_{\hat{\boldsymbol{\xi}},\hat{\mathbf{v}},\hat{\boldsymbol{\lambda}}} = -\mathbf{A}^T\hat{\boldsymbol{\lambda}} - \left(\left[\mathbf{0} \quad \hat{\mathbf{V}}_{\mathbf{A}_2}\right]\right)^T\hat{\boldsymbol{\lambda}} = \mathbf{0}, \tag{5.3}$$

$$\frac{1}{2}\frac{\partial\Phi}{\partial\mathbf{v}}\Big|_{\hat{\boldsymbol{\xi}},\hat{\mathbf{v}},\hat{\boldsymbol{\lambda}}} = \mathbf{P}\hat{\mathbf{v}} - \left[\hat{\boldsymbol{\xi}}_2^T\otimes\mathbf{I}_n, -\mathbf{I}_n\right]^T\hat{\boldsymbol{\lambda}} = \mathbf{P}\hat{\mathbf{v}} - \hat{\mathbf{B}}^T\hat{\boldsymbol{\lambda}} = \mathbf{0}, \tag{5.4}$$

$$\frac{1}{2}\frac{\partial\Phi}{\partial\boldsymbol{\lambda}}\Big|_{\hat{\boldsymbol{\xi}},\hat{\mathbf{v}},\hat{\boldsymbol{\lambda}}} = \mathbf{y} - \mathbf{A}\hat{\boldsymbol{\xi}} - \hat{\mathbf{B}}\hat{\mathbf{v}} = \mathbf{0}, \tag{5.5}$$

where $\hat{\mathbf{B}}=\left[\hat{\boldsymbol{\xi}}_2^T\otimes\mathbf{I}_n, -\mathbf{I}_n\right]$ in this case.

After some simplifications the solution can be given as

$$\left(\begin{bmatrix}\mathbf{A}_1 \\ \mathbf{A}_2+\hat{\mathbf{V}}_{\mathbf{A}_2}\end{bmatrix}\left(\hat{\mathbf{B}}\mathbf{Q}_{\mathbf{ll}}\hat{\mathbf{B}}^T\right)^{-1}\left[\mathbf{A}_1 \quad \mathbf{A}_2+\hat{\mathbf{V}}_{\mathbf{A}_2}\right]\right)\hat{\boldsymbol{\xi}} = \begin{bmatrix}\mathbf{A}_1 \\ \mathbf{A}_2+\hat{\mathbf{V}}_{\mathbf{A}_2}\end{bmatrix}\left(\hat{\mathbf{B}}\mathbf{Q}_{\mathbf{ll}}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{y}+\left[\mathbf{0} \quad \hat{\mathbf{V}}_{\mathbf{A}_2}\right]\hat{\boldsymbol{\xi}}\right) \tag{5.6}$$

which is analogous to Eq (4.26).

The normal matrix can be alternatively written as

$$\left(\begin{bmatrix}\mathbf{A}_1\\ \mathbf{A}_2+\hat{\mathbf{V}}_{\mathbf{A}_2}\end{bmatrix}\left(\hat{\mathbf{B}}\mathbf{Q}_{\text{ll}}\hat{\mathbf{B}}^T\right)^{-1}\begin{bmatrix}\mathbf{A}_1 & \mathbf{A}_2+\hat{\mathbf{V}}_{\mathbf{A}_2}\end{bmatrix}\right)= \\ \begin{bmatrix}\mathbf{A}_1^T\left(\hat{\mathbf{B}}\mathbf{Q}_{\text{ll}}\hat{\mathbf{B}}^T\right)^{-1}\mathbf{A}_1 & \mathbf{A}_1^T\left(\hat{\mathbf{B}}\mathbf{Q}_{\text{ll}}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{A}_2+\hat{\mathbf{V}}_{\mathbf{A}_2}\right)\\ \left(\mathbf{A}_2^T+\hat{\mathbf{V}}_{\mathbf{A}_2}^T\right)\left(\hat{\mathbf{B}}\mathbf{Q}_{\text{ll}}\hat{\mathbf{B}}^T\right)^{-1}\mathbf{A}_1 & \left(\mathbf{A}_2^T+\hat{\mathbf{V}}_{\mathbf{A}_2}^T\right)\left(\hat{\mathbf{B}}\mathbf{Q}_{\text{ll}}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{A}_2+\hat{\mathbf{V}}_{\mathbf{A}_2}\right)\end{bmatrix} \tag{5.7}$$

Parameter elimination

By means of parameter elimination (e.g., Niemeier 2002 p. 286) the estimated ξ_2 reads

$$\left(\mathbf{A}_2^T+\mathbf{V}_{\mathbf{A}_2}^T\right)\left(\hat{\mathbf{B}}\mathbf{Q}_{\text{ll}}\hat{\mathbf{B}}^T\right)^{-1}\mathbf{\Psi}\left(\mathbf{A}_2+\hat{\mathbf{V}}_{\mathbf{A}_2}\right)\hat{\xi}_2=\left(\mathbf{A}_2^T+\mathbf{V}_{\mathbf{A}_2}^T\right)\left(\hat{\mathbf{B}}\mathbf{Q}_{\text{ll}}\hat{\mathbf{B}}^T\right)^{-1}\mathbf{\Psi}\left(\mathbf{y}+\begin{bmatrix}\mathbf{0} & \hat{\mathbf{V}}_{\mathbf{A}_2}\end{bmatrix}\hat{\xi}\right), \tag{5.8}$$

where $\mathbf{\Psi}=\mathbf{I}_n-\mathbf{A}_1\left(\mathbf{A}_1^T\left(\hat{\mathbf{B}}\mathbf{Q}_{\text{ll}}\hat{\mathbf{B}}^T\right)^{-1}\mathbf{A}_1\right)^{-1}\mathbf{A}_1^T\left(\hat{\mathbf{B}}\mathbf{Q}_{\text{ll}}\hat{\mathbf{B}}^T\right)^{-1}$.

Instead of the QR decomposition and the centering matrix utilized in the unweighted case, the idempotent matrix $\mathbf{\Psi}$ is analytically derived. It is valid for the fully populated vcm of observations. Since then, the subparameter vector $\hat{\xi}_1$ can easily be obtained after deriving $\hat{\xi}_2$.

5.2 Fixing elements

There is a more generalized case, namely fixing elements, which means that arbitrary elements in the model matrix instead of a whole column are stochastic. For fixing elements only in the model matrix the optimization problem can be expressed as

$$\min \mathbf{v}^T\mathbf{P}\mathbf{v} \\ \textit{subject to } \mathbf{y}+\mathbf{v}_\mathbf{y}=\left(\mathbf{A}+\mathbf{V}_\mathbf{A}\right)\xi=\mathbf{A}_1\xi+\left(\mathbf{A}_2+\mathbf{V}_\mathbf{A}\right)\xi \tag{5.9}$$

where $\mathbf{v}=\begin{bmatrix}\mathbf{v}_\mathbf{A}^\#\\ \mathbf{v}_\mathbf{y}\end{bmatrix}$. $\mathbf{v}_\mathbf{A}^\#$ denotes the vectorization of the stochastic part of the residual matrix $\mathbf{V}_\mathbf{A}$ (excepting the 0 element, which represents the deterministic element in the model matrix). $\mathbf{P}$ is the correspondent weight matrix of the vector $\mathbf{v}$.

An example is shown to illustrate the structure of these matrices. If we have the corrected design matrix

$$\mathbf{A}+\mathbf{V}_\mathbf{A}=\mathbf{A}_1+\mathbf{A}_2+\mathbf{V}_\mathbf{A}=\begin{bmatrix}0 & a_{12} & 0\\ 0 & a_{22} & 0\\ a_{31} & 0 & 0\\ a_{41} & a_{42} & a_{43}\end{bmatrix}+\begin{bmatrix}a_{11}^\# & 0 & a_{13}^\#\\ a_{21}^\# & 0 & a_{23}^\#\\ 0 & a_{32}^\# & a_{33}^\#\\ 0 & 0 & 0\end{bmatrix}+\begin{bmatrix}v_{a_{11}^\#} & 0 & v_{a_{13}^\#}\\ v_{a_{21}^\#} & 0 & v_{a_{23}^\#}\\ 0 & v_{a_{32}^\#} & v_{a_{33}^\#}\\ 0 & 0 & 0\end{bmatrix}, \tag{5.10}$$

the model matrix $\mathbf{B}$ is the partial derivative w.r.t. the vector

$$\mathbf{v}=\begin{bmatrix}v_{a_{11}^\#} & v_{a_{21}^\#} & v_{a_{32}^\#} & v_{a_{13}^\#} & v_{a_{23}^\#} & v_{a_{33}^\#} & \mathbf{v}_\mathbf{y}^T\end{bmatrix}^T, \tag{5.11}$$

should be organized as

$$\mathbf{B} = \begin{bmatrix} \xi_1 & 0 & 0 & \xi_3 & 0 & 0 & \\ 0 & \xi_1 & 0 & 0 & \xi_3 & 0 & \\ 0 & 0 & \xi_2 & 0 & 0 & \xi_3 & -\mathbf{I}_4 \\ 0 & 0 & 0 & 0 & 0 & 0 & \end{bmatrix}, \tag{5.12}$$

where $-\mathbf{I}_4$ is the derivative w.r.t. $\mathbf{v}_y$.

The solution of iterative GHM method (see Eq (4.37)) can be described as follows

$$\hat{\boldsymbol{\xi}} = \left(\left(\mathbf{A}^T + \hat{\mathbf{V}}_\mathbf{A}^T \right) \left(\hat{\mathbf{B}} \mathbf{Q}_{ll} \hat{\mathbf{B}}^T \right)^{-1} \left(\mathbf{A} + \hat{\mathbf{V}}_\mathbf{A} \right) \right)^{-1} \left(\mathbf{A}^T + \hat{\mathbf{V}}_\mathbf{A}^T \right) \left(\hat{\mathbf{B}} \mathbf{Q}_{ll} \hat{\mathbf{B}}^T \right)^{-1} \left(\mathbf{y} + \hat{\mathbf{V}}_\mathbf{A} \hat{\boldsymbol{\xi}} \right), \tag{5.13}$$

The stochastic part and deterministic part of the matrix $\hat{\mathbf{V}}_\mathbf{A}$ can be also retrieved through $\mathbf{v}_\mathbf{A}^{\#}$ and zeros, respectively.

5.3 Linear constrained WTLS problem

The definition for unweighted TLS problem with linear constraints has been addressed in Chapter 3 whereas the linear constrained WTLS case is given here:

$$\begin{gathered} \mathbf{y} + \mathbf{v}_y = \left(\mathbf{A} + \mathbf{V}_\mathbf{A} \right) \boldsymbol{\xi} \qquad \text{and} \qquad \boldsymbol{\kappa}_0 = \mathbf{K} \boldsymbol{\xi} \\ D(\mathbf{l}) = D\left(\begin{bmatrix} vec(\mathbf{A}) \\ \mathbf{y} \end{bmatrix} \right) = \sigma_0^2 \mathbf{Q}_{ll} \end{gathered} . \tag{5.14}$$

To obtain the solution the Lagrange multiplier is applied:

$$\Phi\left(\mathbf{v}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\xi} \right) = \mathbf{v}^T \mathbf{P} \mathbf{v} + 2\boldsymbol{\lambda}^T \left(\mathbf{y} - \mathbf{A}\boldsymbol{\xi} - \mathbf{V}_\mathbf{A} \boldsymbol{\xi} + \mathbf{v}_y \right) - 2\boldsymbol{\mu}^T \left(\boldsymbol{\kappa}_0 - \mathbf{K}\boldsymbol{\xi} \right). \tag{5.15}$$

The corresponding necessary conditions for the stationary point read

$$\frac{1}{2} \frac{\partial \Phi}{\partial \boldsymbol{\xi}} \bigg|_{\hat{\boldsymbol{\xi}}, \hat{\mathbf{v}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\mu}}} = -\mathbf{A}^T \hat{\boldsymbol{\lambda}} - \hat{\mathbf{V}}_\mathbf{A}^T \hat{\boldsymbol{\lambda}} + \mathbf{K}^T \hat{\boldsymbol{\mu}} = \mathbf{0} \tag{5.16}$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial \mathbf{v}} \bigg|_{\hat{\boldsymbol{\xi}}, \hat{\mathbf{v}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\mu}}} = \mathbf{P} \hat{\mathbf{v}} - \hat{\mathbf{B}}^T \hat{\boldsymbol{\lambda}} = \mathbf{0} \tag{5.17}$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial \boldsymbol{\lambda}} \bigg|_{\hat{\boldsymbol{\xi}}, \hat{\mathbf{v}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\mu}}} = \mathbf{y} - \mathbf{A} \hat{\boldsymbol{\xi}} - \hat{\mathbf{B}} \hat{\mathbf{v}} = \mathbf{0} \tag{5.18}$$

$$\frac{1}{2} \frac{\partial \Phi}{\partial \boldsymbol{\mu}} \bigg|_{\hat{\boldsymbol{\xi}}, \hat{\mathbf{v}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\mu}}} = \boldsymbol{\kappa}_0 - \mathbf{K} \hat{\boldsymbol{\xi}} = \mathbf{0} \tag{5.19}$$

where $\hat{\mathbf{B}} = \left[\hat{\boldsymbol{\xi}}^T \otimes \mathbf{I}_n, -\mathbf{I}_n \right]$.

The vector of the estimated multipliers is derived from (5.17) and (5.18) as follows

$$\hat{\boldsymbol{\lambda}} = \left(\hat{\mathbf{B}} \mathbf{Q}_{ll} \hat{\mathbf{B}}^T \right)^{-1} \left(\mathbf{y} - \mathbf{A} \hat{\boldsymbol{\xi}} \right), \tag{5.20}$$

which is same as the unconstrained case.

Re-inserting the multiplier $\hat{\boldsymbol{\lambda}}$ to Eq (5.16), the normal equation system is derived as

$$\begin{bmatrix} \hat{\mathbf{N}} & \mathbf{K}^T \\ \mathbf{K} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\xi}} \\ \hat{\boldsymbol{\mu}} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{n}} \\ \boldsymbol{\kappa}_0 \end{bmatrix}, \tag{5.21}$$

where $\hat{\mathbf{N}} = \left(\mathbf{A}^T + \hat{\mathbf{V}}_{\mathbf{A}}^T\right)\left(\hat{\mathbf{B}}\mathbf{Q}_{\mathbf{ll}}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{A} + \hat{\mathbf{V}}_{\mathbf{A}}\right)$ and $\hat{\mathbf{n}} = \left(\mathbf{A}^T + \hat{\mathbf{V}}_{\mathbf{A}}^T\right)\left(\hat{\mathbf{B}}\mathbf{Q}_{\mathbf{ll}}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{y} + \hat{\mathbf{V}}_{\mathbf{A}}\boldsymbol{\xi}\right)$.

Then, the estimated parameter vector for the linear constrained WTLS (LCWTLS) problem can be expressed with the help of matrix identities (Koch 1999 p. 33) as

$$\begin{bmatrix} \hat{\boldsymbol{\xi}} \\ \hat{\boldsymbol{\mu}} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{N}} & \mathbf{K}^T \\ \mathbf{K} & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{n}} \\ \boldsymbol{\kappa}_0 \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{N}}^{-1} + \hat{\mathbf{N}}^{-1}\mathbf{K}^T\left(-\mathbf{K}\hat{\mathbf{N}}^{-1}\mathbf{K}^T\right)^{-1}\mathbf{K}\hat{\mathbf{N}}^{-1} & -\hat{\mathbf{N}}^{-1}\mathbf{K}^T\left(-\mathbf{K}\hat{\mathbf{N}}^{-1}\mathbf{K}^T\right)^{-1} \\ -\left(-\mathbf{K}\hat{\mathbf{N}}^{-1}\mathbf{K}^T\right)^{-1}\mathbf{K}\hat{\mathbf{N}}^{-1} & \left(-\mathbf{K}\hat{\mathbf{N}}^{-1}\mathbf{K}^T\right)^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{n}} \\ \boldsymbol{\kappa}_0 \end{bmatrix}, \tag{5.22}$$

$$\Rightarrow \hat{\boldsymbol{\xi}}_{LCWTLS} := \hat{\mathbf{N}}^{-1}\left(\hat{\mathbf{n}} + \mathbf{K}^T\left(\mathbf{K}\hat{\mathbf{N}}^{-1}\mathbf{K}^T\right)^{-1}\left(\boldsymbol{\kappa}_0 - \mathbf{K}\hat{\mathbf{N}}^{-1}\hat{\mathbf{n}}\right)\right)$$

or alternatively as

$$\hat{\boldsymbol{\xi}}_{LCWTLS} := \hat{\boldsymbol{\xi}}_{WTLS} - \hat{\mathbf{N}}^{-1}\mathbf{K}^T\left(\mathbf{K}\hat{\mathbf{N}}^{-1}\mathbf{K}^T\right)^{-1}\left(\mathbf{K}\hat{\boldsymbol{\xi}}_{WTLS} - \boldsymbol{\kappa}_0\right). \tag{5.23}$$

Now, we get access to the algorithm for the linear constrained WTLS with the iteration based on Eq (5.22) or (5.23).

5.4 WTLS with parameter matrix

In some geodetic applications, e.g., affine transformation, one formulates the parameters and the conventional observations always as the matrix-wise instead of the vector-wise. The well-known model for 3D affine transformation with 12 parameters, which can be obtained from the extension of 2D affine transformation mentioned in chapter 3, can be written as follows

$$\begin{bmatrix} y_1 + v_{y_1} \\ y_2 + v_{y_2} \\ y_3 + v_{y_3} \end{bmatrix} = \begin{bmatrix} \xi_{11} & \xi_{21} & \xi_{31} \\ \xi_{12} & \xi_{22} & \xi_{32} \\ \xi_{13} & \xi_{23} & \xi_{33} \end{bmatrix} \begin{bmatrix} x_1 + v_{x_1} \\ x_2 + v_{x_2} \\ x_3 + v_{x_3} \end{bmatrix} + \begin{bmatrix} \xi_{41} \\ \xi_{42} \\ \xi_{43} \end{bmatrix}, \tag{5.24}$$

where the 12 mathematical parameters represent the 12 implicit or explicit geometrical parameters, namely 3 translations, 3 rotation angles, 3 shears and 3 scale factors. The coordinates for a particular point are x_1, x_2 and x_3 in one coordinate system and y_1, y_2 and y_3 in another coordinate system. Besides, v with the indices denotes the correspondent corrections.

With the help of Eq (5.24) we obtain the EIV model with fixing column for the whole system as follows

$$\begin{bmatrix} y_{11} + v_{y_{11}} & y_{12} + v_{y_{12}} & y_{13} + v_{y_{13}} \\ y_{21} + v_{y_{21}} & y_{22} + v_{y_{22}} & y_{23} + v_{y_{23}} \\ \cdots & \cdots & \cdots \\ y_{n1} + v_{y_{n1}} & y_{n2} + v_{y_{n2}} & y_{n3} + v_{y_{n3}} \end{bmatrix} = \begin{bmatrix} x_{11} + v_{x_{11}} & x_{12} + v_{x_{12}} & x_{13} + v_{x_{13}} & 1 \\ x_{21} + v_{x_{21}} & x_{22} + v_{x_{22}} & x_{23} + v_{x_{23}} & 1 \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} + v_{x_{n1}} & x_{n2} + v_{x_{n2}} & x_{n3} + v_{x_{n3}} & 1 \end{bmatrix} \begin{bmatrix} \xi_{11} & \xi_{12} & \xi_{13} \\ \xi_{21} & \xi_{22} & \xi_{23} \\ \xi_{31} & \xi_{32} & \xi_{33} \\ \xi_{41} & \xi_{42} & \xi_{43} \end{bmatrix}, \tag{5.25}$$

where x_{ij} and y_{ij} ($1 \le i \le n,\ 1 \le j \le 3$) are the coordinates in both coordinate system with the correspondent residual $v_{x_{ij}}$ and $v_{y_{ij}}$ while n is the number of the observations.

In this case one may solve the WTLS problem with the parameter matrix instead of the parameter vector. The functional and stochastic model of the WTLS problem with parameter matrix can be expressed in the following equations:

$$\mathbf{Y}+\mathbf{V}_{\mathbf{Y}}=\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}\right)\boldsymbol{\Xi} \tag{5.26}$$

$$\mathbf{l}=\begin{bmatrix} vec(\mathbf{A}) \\ vec(\mathbf{Y}) \end{bmatrix} \qquad \mathbf{v}=\begin{bmatrix} vec(\mathbf{V}_{\mathbf{A}}) \\ vec(\mathbf{V}_{\mathbf{Y}}) \end{bmatrix}$$
$$D(\mathbf{l})=\sigma_0^2\mathbf{Q}_{\mathbf{ll}}=\sigma_0^2\begin{bmatrix} \mathbf{Q}_{\mathbf{AA}} & \mathbf{Q}_{\mathbf{AY}} \\ \mathbf{Q}_{\mathbf{YA}} & \mathbf{Q}_{\mathbf{YY}} \end{bmatrix}=\mathbf{P}^{-1} \tag{5.27}$$

$\boldsymbol{\Xi}$ denotes the $u\times d$ parameter matrix. $\mathbf{Y}$ is the $n\times d$ observation matrix, and $\mathbf{V}_{\mathbf{Y}}$ is the $n\times d$ residual matrix referring to $\mathbf{Y}$.

To solve the system, the Lagrange function reads

$$\begin{aligned}\Phi\left(\mathbf{v},\boldsymbol{\Lambda},\boldsymbol{\Xi}\right)&=\mathbf{v}^T\mathbf{P}\mathbf{v}+2tr\left(\boldsymbol{\Lambda}^T\left(\mathbf{Y}+\mathbf{V}_{\mathbf{Y}}-\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}\right)\boldsymbol{\Xi}\right)\right)\\ &=\mathbf{v}^T\mathbf{P}\mathbf{v}+2vec^T\left(\boldsymbol{\Lambda}\right)vec\left(\mathbf{Y}+\mathbf{V}_{\mathbf{Y}}-\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}\right)\boldsymbol{\Xi}\right)\end{aligned}, \tag{5.28}$$

where $\boldsymbol{\Lambda}$ denotes the $n\times d$ matrix of Lagrange multipliers.

We then obtain the necessary Euler-Lagrange conditions:

$$\frac{1}{2}\frac{\partial\Phi}{\partial\boldsymbol{\Xi}}\Big|_{\hat{\boldsymbol{\Xi}},\hat{\mathbf{v}},\hat{\boldsymbol{\Lambda}}}=-\mathbf{A}^T\hat{\boldsymbol{\Lambda}}-\hat{\mathbf{V}}_{\mathbf{A}}^T\hat{\boldsymbol{\Lambda}}=-\left(\mathbf{I}_d\otimes\left(\mathbf{A}^T+\hat{\mathbf{V}}_{\mathbf{A}}^T\right)\right)vec\left(\hat{\boldsymbol{\Lambda}}\right)=\mathbf{0}\,. \tag{5.29}$$

$$\frac{1}{2}\frac{\partial\Phi}{\partial\mathbf{v}}\Big|_{\hat{\boldsymbol{\Xi}},\hat{\mathbf{v}},\hat{\boldsymbol{\Lambda}}}=\mathbf{P}\hat{\mathbf{v}}-\left[\hat{\boldsymbol{\Xi}}^T\otimes\mathbf{I}_n,-\mathbf{I}_d\otimes\mathbf{I}_n\right]^T vec\left(\hat{\boldsymbol{\Lambda}}\right)=\mathbf{P}\hat{\mathbf{v}}-\hat{\mathbf{B}}^T vec\left(\hat{\boldsymbol{\Lambda}}\right)=\mathbf{0}\,. \tag{5.30}$$

$$\frac{1}{2}\frac{\partial\Phi}{\partial\boldsymbol{\Lambda}}\Big|_{\hat{\boldsymbol{\Xi}},\hat{\mathbf{v}},\hat{\boldsymbol{\Lambda}}}=\left(\mathbf{Y}+\hat{\mathbf{V}}_{\mathbf{Y}}-\left(\mathbf{A}+\hat{\mathbf{V}}_{\mathbf{A}}\right)\hat{\boldsymbol{\Xi}}\right)=\mathbf{0}\Rightarrow vec\left(\mathbf{Y}-\mathbf{A}\hat{\boldsymbol{\Xi}}\right)-\hat{\mathbf{B}}\hat{\mathbf{v}}=\mathbf{0}\,, \tag{5.31}$$

with $\hat{\mathbf{B}}=\left[\hat{\boldsymbol{\Xi}}^T\otimes\mathbf{I}_n,-\mathbf{I}_d\otimes\mathbf{I}_n\right]$.

Inserting $\hat{\mathbf{v}}=-\mathbf{Q}_{\mathbf{ll}}\hat{\mathbf{B}}^T vec\left(\hat{\boldsymbol{\Lambda}}\right)$ obtained by Eq (5.30) into Eq (5.31) leads to

$$vec\left(\hat{\boldsymbol{\Lambda}}\right)=\left(\hat{\mathbf{B}}\mathbf{Q}_{\mathbf{ll}}\hat{\mathbf{B}}^T\right)^{-1}vec\left(\mathbf{Y}-\mathbf{A}\hat{\boldsymbol{\Xi}}\right), \tag{5.32}$$

The residual vector can be derived as

$$\hat{\mathbf{v}}=\mathbf{Q}_{\mathbf{ll}}\hat{\mathbf{B}}^T vec\left(\hat{\boldsymbol{\Lambda}}\right). \tag{5.33}$$

Then, we obtain the normal equation from Eqs (5.29) and (5.32) as

$$\left(\mathbf{I}_d\otimes\left(\mathbf{A}^T+\hat{\mathbf{V}}_{\mathbf{A}}^T\right)\right)\left(\hat{\mathbf{B}}\mathbf{Q}_{\mathbf{ll}}\hat{\mathbf{B}}^T\right)^{-1}vec\left(\mathbf{Y}-\mathbf{A}\hat{\boldsymbol{\Xi}}\right)=\mathbf{0} \tag{5.34}$$

or

$$\left(\mathbf{I}_d\otimes\left(\mathbf{A}^T+\hat{\mathbf{V}}_{\mathbf{A}}^T\right)\right)\left(\hat{\mathbf{B}}\mathbf{Q}_{\mathbf{ll}}\hat{\mathbf{B}}^T\right)^{-1}\left(vec\left(\mathbf{Y}\right)+\left(\mathbf{I}_d\otimes\hat{\mathbf{V}}_{\mathbf{A}}\right)vec\hat{\boldsymbol{\Xi}}-\left(\mathbf{I}_d\otimes\left(\mathbf{A}+\hat{\mathbf{V}}_{\mathbf{A}}\right)\right)vec\hat{\boldsymbol{\Xi}}\right)=\mathbf{0}\,. \tag{5.35}$$

with

$$vec\left(\mathbf{A}\hat{\boldsymbol{\Xi}}\mathbf{I}_d\right)=\left(\mathbf{I}_d\otimes\mathbf{A}\right)vec\hat{\boldsymbol{\Xi}}=\left(\mathbf{I}_d\otimes\left(\mathbf{A}+\hat{\mathbf{V}}_{\mathbf{A}}-\hat{\mathbf{V}}_{\mathbf{A}}\right)\right)vec\hat{\boldsymbol{\Xi}}=\left(\mathbf{I}_d\otimes\left(\mathbf{A}+\hat{\mathbf{V}}_{\mathbf{A}}\right)\right)vec\hat{\boldsymbol{\Xi}}-\left(\mathbf{I}_d\otimes\hat{\mathbf{V}}_{\mathbf{A}}\right)vec\hat{\boldsymbol{\Xi}}.$$

To solve the WTLS problem with parameter matrix we rearrange Eq (5.35) and write:

$$vec\left(\hat{\boldsymbol{\Xi}}\right)=\mathbf{N}^{-1}\left(\mathbf{I}_d\otimes\left(\mathbf{A}^T+\hat{\mathbf{V}}_{\mathbf{A}}^T\right)\right)\left(\hat{\mathbf{B}}\mathbf{Q}_{\mathbf{ll}}\hat{\mathbf{B}}^T\right)^{-1}\left(vec\left(\mathbf{Y}\right)+\left(\mathbf{I}_d\otimes\hat{\mathbf{V}}_{\mathbf{A}}\right)vec\hat{\boldsymbol{\Xi}}\right), \tag{5.36}$$

where $\mathbf{N}=\left(\left(\mathbf{I}_d\otimes\left(\mathbf{A}^T+\hat{\mathbf{V}}_\mathbf{A}^T\right)\right)\left(\hat{\mathbf{B}}\mathbf{Q}_{\mathbf{ll}}\hat{\mathbf{B}}^T\right)^{-1}\left(\mathbf{I}_d\otimes\left(\mathbf{A}+\hat{\mathbf{V}}_\mathbf{A}\right)\right)\right)$.

The WTLS problem with parameter matrix with fixing columns to solve the transformation problem can be analogously treated as shown in Chapter 5.1 and is left out here.

5.5 Rigorous solutions of the non-linear EIV model

In some case of handling the geodetic data, the linear EIV model presented in previous part is improper. For example, the method proposed by Drixler (1993) for estimation of the form parameters is widely used in geodesy. Of course, the model can be solved with the non-linear GHM method proposed in Pope (1972). However, in many cases the parameters and observations are constrained, e.g., some parameters are fixed if the quadratic form is a sphere. Thus, the Gauss-Newton algorithm should be generalized to treat the non-linear TLS problem with constraints. Thus, the general functional models can be organized as

$$\begin{aligned}&\mathbf{f}_1\left(\boldsymbol{\xi},\mathbf{l}+\mathbf{v}\right)=\mathbf{0}\\&\mathbf{f}_2\left(\boldsymbol{\xi}\right)=\mathbf{0}\\&\mathbf{f}_3\left(\mathbf{l}+\mathbf{v}\right)=\mathbf{0}\end{aligned}\tag{5.37}$$

The function $\mathbf{f}_1$ denotes the observation equations containing the parameters and observations; This is also called non-linear GHM. $\mathbf{f}_2$ denotes the linear or non-linear constraints about the parameters. $\mathbf{f}_3$ is the conditional function about the observations. The combination is complicated. However, the model is useful for some open geodetic applications (see Chapter 6 for detail).

The objective function expressed with Lagrange multipliers is given as follows

$$\Phi\left(\mathbf{v},\boldsymbol{\lambda},\boldsymbol{\xi}\right)=\mathbf{v}^T\mathbf{P}\mathbf{v}+2\boldsymbol{\lambda}_1^T\mathbf{f}_1\left(\boldsymbol{\xi},\mathbf{l}+\mathbf{v}\right)+2\boldsymbol{\lambda}_2^T\mathbf{f}_2\left(\boldsymbol{\xi}\right)+2\boldsymbol{\lambda}_3^T\mathbf{f}_3\left(\mathbf{l}+\mathbf{v}\right)\tag{5.38}$$

where $\boldsymbol{\lambda}=\begin{bmatrix}\boldsymbol{\lambda}_1^T & \boldsymbol{\lambda}_2^T & \boldsymbol{\lambda}_3^T\end{bmatrix}^T$.

In order to obtain the minimum of the objective function, the functions $\mathbf{f}_2\left(\boldsymbol{\xi}\right)$ and $\mathbf{f}_3\left(\mathbf{l}+\mathbf{v}\right)$ are rebuilt as two pseudo observation equations $\mathbf{f}_2\left(\boldsymbol{\xi}\right)=\mathbf{0}+\mathbf{v}_{pseudo}$ and $\mathbf{f}_3\left(\mathbf{l}+\mathbf{v}\right)=\mathbf{0}\cdot\boldsymbol{\xi}$, where the vector $\mathbf{v}_{pseudo}$ with small variances ensures the fulfillment of the constraints. The both functions are originally introduced as the constraints and conditional function of observations, and are now formulated as pseudo observation equations and integrated into the non-linear GHM. In order to obtain the minimum, we give the partial derivative of the objective function w.r.t the parameter vector and the residual vector as

$$\begin{aligned}\frac{1}{2}\frac{\partial\Phi}{\partial\boldsymbol{\xi}}\bigg|_{\hat{\boldsymbol{\xi}},\hat{\mathbf{v}},\hat{\boldsymbol{\lambda}},\hat{\mathbf{v}}_{pseudo}}&=\frac{\partial\mathbf{f}_1^T\left(\boldsymbol{\xi},\mathbf{l}+\mathbf{v}\right)}{\partial\boldsymbol{\xi}}\bigg|_{\hat{\boldsymbol{\xi}},\hat{\mathbf{v}},\hat{\boldsymbol{\lambda}},\hat{\mathbf{v}}_{pseudo}}\hat{\boldsymbol{\lambda}}_1+\frac{\mathbf{f}_2^T\left(\boldsymbol{\xi}\right)}{\partial\boldsymbol{\xi}}\bigg|_{\hat{\boldsymbol{\xi}},\hat{\mathbf{v}},\hat{\boldsymbol{\lambda}},\hat{\mathbf{v}}_{pseudo}}\hat{\boldsymbol{\lambda}}_2+\mathbf{0}\cdot\hat{\boldsymbol{\lambda}}_3\\&=\begin{bmatrix}\dfrac{\partial\mathbf{f}_1\left(\boldsymbol{\xi},\mathbf{l}+\mathbf{v}\right)}{\partial\boldsymbol{\xi}^T}\bigg|_{\hat{\boldsymbol{\xi}},\hat{\mathbf{v}},\hat{\boldsymbol{\lambda}},\hat{\mathbf{v}}_{pseudo}}\\\dfrac{\mathbf{f}_2\left(\boldsymbol{\xi}\right)}{\partial\boldsymbol{\xi}^T}\bigg|_{\hat{\boldsymbol{\xi}},\hat{\mathbf{v}},\hat{\boldsymbol{\lambda}},\hat{\mathbf{v}}_{pseudo}}\\\mathbf{0}\end{bmatrix}^T\hat{\boldsymbol{\lambda}}=\hat{\mathbf{J}}_1^T\hat{\boldsymbol{\lambda}}=\mathbf{0}\end{aligned}\tag{5.39}$$

$$\frac{1}{2}\frac{\partial\Phi}{\partial\begin{bmatrix}\mathbf{v}\\ \mathbf{v}_{pseudo}\end{bmatrix}}\Bigg|_{\hat{\xi},\hat{\mathbf{v}},\hat{\boldsymbol{\lambda}},\hat{\mathbf{v}}_{pseudo}}$$

$$=\begin{bmatrix}\mathbf{P} & \mathbf{0}\\ \mathbf{0} & \mathbf{P}_{pseudo}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{v}}\\ \hat{\mathbf{v}}_{pseudo}\end{bmatrix}+\begin{bmatrix}\dfrac{\partial\mathbf{f}_1^T(\boldsymbol{\xi},\mathbf{l}+\mathbf{v})}{\partial\mathbf{v}}\Bigg|_{\hat{\xi},\hat{\mathbf{v}},\hat{\boldsymbol{\lambda}},\hat{\mathbf{v}}_{pseudo}}\\ \mathbf{0}\end{bmatrix}\hat{\boldsymbol{\lambda}}_1+\begin{bmatrix}\mathbf{0}\\ -\mathbf{I}_{n_2}\end{bmatrix}\hat{\boldsymbol{\lambda}}_2+\begin{bmatrix}\dfrac{\partial\mathbf{f}_3^T(\mathbf{l}+\mathbf{v})}{\partial\mathbf{v}}\\ \mathbf{0}\end{bmatrix}\cdot\hat{\boldsymbol{\lambda}}_3$$

$$=\mathbf{P}_{total}\hat{\mathbf{v}}_{total}+\begin{bmatrix}\dfrac{\partial\mathbf{f}_1^T(\boldsymbol{\xi},\mathbf{l}+\mathbf{v})}{\partial\mathbf{v}}\Bigg|_{\hat{\xi},\hat{\mathbf{v}},\hat{\boldsymbol{\lambda}},\hat{\mathbf{v}}_{pseudo}} & \mathbf{0} & \dfrac{\partial\mathbf{f}_3^T(\mathbf{l}+\mathbf{v})}{\partial\mathbf{v}}\\ \mathbf{0} & -\mathbf{I}_{n_2} & \mathbf{0}\end{bmatrix}\hat{\boldsymbol{\lambda}}$$

$$=\mathbf{P}_{total}\hat{\mathbf{v}}_{total}+\hat{\mathbf{J}}_2^T\hat{\boldsymbol{\lambda}}=\mathbf{0}\Leftrightarrow\hat{\mathbf{v}}_{total}=\mathbf{P}_{total}^{-1}\hat{\mathbf{J}}_2^T\hat{\boldsymbol{\lambda}}\Leftrightarrow\left(\hat{\mathbf{J}}_2\mathbf{P}_{total}^{-1}\hat{\mathbf{J}}_2^T\right)^{-1}\hat{\mathbf{J}}_2\hat{\mathbf{v}}_{total}=\hat{\boldsymbol{\lambda}}$$

(5.40)

if we define $\mathbf{P}_{total}:=\begin{bmatrix}\mathbf{P} & \mathbf{0}\\ \mathbf{0} & \mathbf{P}_{pseudo}\end{bmatrix}$, $\mathbf{v}_{total}:=\begin{bmatrix}\mathbf{v}\\ \mathbf{v}_{pseudo}\end{bmatrix}$, n_2 is the number of the constraints function $\mathbf{f}_2$.

The normal equation system can be obtained by combining the Eq (5.39) and (5.40) as

$$\hat{\mathbf{J}}_1^T\left(\hat{\mathbf{J}}_2\mathbf{P}_{total}^{-1}\hat{\mathbf{J}}_2^T\right)^{-1}\hat{\mathbf{J}}_2\hat{\mathbf{v}}_{total}=\mathbf{0} \tag{5.41}$$

If we linearize the combined functional model $\left[\mathbf{f}_1^T(\boldsymbol{\xi},\mathbf{l}+\mathbf{v})\quad\mathbf{f}_2^T(\boldsymbol{\xi},\mathbf{0}+\mathbf{v}_{pseudo})\quad\mathbf{f}_3^T(\mathbf{0},\mathbf{l}+\mathbf{v})\right]^T=\mathbf{0}$ at $\boldsymbol{\xi}^0$ and $\mathbf{l}^0+\mathbf{v}^0\left(\mathbf{l}^0=\mathbf{l}-\underset{\sim}{\mathbf{0}}\right)$ neglecting the terms of the higher order, we have

$$\mathbf{J}_1^0 d\hat{\boldsymbol{\xi}}^1+\mathbf{J}_2^0\left(\hat{\mathbf{v}}_{total}-\begin{bmatrix}\mathbf{v}^0\\ \mathbf{v}_{pseudo}^0\end{bmatrix}\right)+\begin{bmatrix}\mathbf{f}_1\left(\boldsymbol{\xi}^0,\mathbf{l}^0+\mathbf{v}^0\right)\\ \mathbf{f}_2\left(\boldsymbol{\xi}^0,\mathbf{0}+\mathbf{v}_{pseudo}^0\right)\\ \mathbf{f}_3\left(\mathbf{0},\mathbf{l}^0+\mathbf{v}^0\right)\end{bmatrix}=\mathbf{0}, \tag{5.42}$$

leading to

$$\mathbf{J}_2^0\hat{\mathbf{v}}_{total}=-\left(\mathbf{J}_1^0 d\hat{\boldsymbol{\xi}}^1+\begin{bmatrix}\mathbf{f}_1\left(\boldsymbol{\xi}^0,\mathbf{l}^0+\mathbf{v}^0\right)\\ \mathbf{f}_2\left(\boldsymbol{\xi}^0,\mathbf{0}+\mathbf{v}_{pseudo}^0\right)\\ \mathbf{f}_3\left(\mathbf{0},\mathbf{l}^0+\mathbf{v}^0\right)\end{bmatrix}-\mathbf{J}_2^0\begin{bmatrix}\mathbf{v}^0\\ \mathbf{v}_{pseudo}^0\end{bmatrix}\right)=-\left(\mathbf{J}_1^0 d\hat{\boldsymbol{\xi}}^1+\mathbf{w}_{total}^0\right), \tag{5.43}$$

where the vector $\mathbf{w}_{total}^0=\begin{bmatrix}\mathbf{f}_1\left(\boldsymbol{\xi}^0,\mathbf{l}^0+\mathbf{v}^0\right)\\ \mathbf{f}_2\left(\boldsymbol{\xi}^0,\mathbf{0}+\mathbf{v}_{pseudo}^0\right)\\ \mathbf{f}_3\left(\mathbf{0},\mathbf{l}^0+\mathbf{v}^0\right)\end{bmatrix}-\mathbf{J}_2^0\begin{bmatrix}\mathbf{v}^0\\ \mathbf{v}_{pseudo}^0\end{bmatrix}$ is usually called the inconsistency vector in the sense of GHM.

Because $\mathbf{J}_2$ is the first partial derivative of the functional model w.r.t. the observations, $\begin{bmatrix} \mathbf{f}_1\left(\xi^0,\mathbf{l}^0\right) \\ \mathbf{f}_2\left(\xi^0,\mathbf{0}\right) \\ \mathbf{f}_3\left(\mathbf{0},\mathbf{l}^0\right) \end{bmatrix} = \begin{bmatrix} \mathbf{f}_1\left(\xi^0,\mathbf{l}^0+\mathbf{v}^0\right) \\ \mathbf{f}_2\left(\xi^0,\mathbf{0}+\mathbf{v}^0_{pseudo}\right) \\ \mathbf{f}_3\left(\mathbf{0},\mathbf{l}^0+\mathbf{v}^0\right) \end{bmatrix} - \mathbf{J}^0_2 \begin{bmatrix} \mathbf{v}^0 \\ \mathbf{v}^0_{pseudo} \end{bmatrix}$ if there are no terms of the quadratic and higher form of the observations.

Inserting the Eq (5.43) into Eq (5.41), the updated term of the estimated parameter $d\xi^1$ and the estimated parameter vector $\hat{\xi}^1$ can be obtained as

$$\begin{aligned} &\left(\mathbf{J}^0_1\right)^T \left(\mathbf{J}^0_2\mathbf{P}^{-1}_{total}\left(\mathbf{J}^0_2\right)^T\right)^{-1}\left(\mathbf{J}^0_1 d\hat{\xi}^1 + \mathbf{w}^0_{total}\right) = \mathbf{0} \\ &\Leftrightarrow d\hat{\xi}^1 = -\left(\left(\mathbf{J}^0_1\right)^T \left(\mathbf{J}^0_2\mathbf{P}^{-1}_{total}\left(\mathbf{J}^0_2\right)^T\right)^{-1}\mathbf{J}^0_1\right)^{-1}\left(\mathbf{J}^0_1\right)^T \left(\mathbf{J}^0_2\mathbf{P}^{-1}_{total}\left(\mathbf{J}^0_2\right)^T\right)^{-1}\mathbf{w}^0_{total} \end{aligned} \tag{5.44}$$

and

$$\hat{\xi}^1 = \xi^0 - \left(\left(\mathbf{J}^0_1\right)^T \left(\mathbf{J}^0_2\mathbf{P}^{-1}_{total}\left(\mathbf{J}^0_2\right)^T\right)^{-1}\mathbf{J}^0_1\right)^{-1}\left(\mathbf{J}^0_1\right)^T \left(\mathbf{J}^0_2\mathbf{P}^{-1}_{total}\left(\mathbf{J}^0_2\right)^T\right)^{-1}\mathbf{w}^0_{total} \tag{5.45}$$

respectively.

The residual vector is estimated as

$$\mathbf{v}^1_{total} = -\mathbf{P}^{-1}_{total}\left(\mathbf{J}^0_2\right)^T \left(\mathbf{J}^0_2\mathbf{P}^{-1}_{total}\left(\mathbf{J}^0_2\right)^T\right)^{-1}\left(\mathbf{w}^0_{total} + \mathbf{J}^0_1 d\hat{\xi}^1\right), \tag{5.46}$$

The estimated parameters and residuals should substitute the old parameters and update the corrected observations with $\mathbf{l}^0 + \mathbf{v}^i$ in the both Jacobian matrices and the inconsistency vector. The process is iteratively computed until the values of parameters do not change in the given magnitude.

Considering the stochastic parameter expressed as $\xi \sim N\left(\boldsymbol{\mu} \quad \boldsymbol{\Sigma}_{\xi\xi}\right)$ (e.g., from the previous adjustment), the objective function based on Eq (5.38) reads

$$\Phi\left(\mathbf{v},\boldsymbol{\lambda},\xi\right) = \left(\xi-\boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}_{\xi\xi}\left(\xi-\boldsymbol{\mu}\right) + \mathbf{v}^T\mathbf{P}\mathbf{v} + 2\boldsymbol{\lambda}^T_1\mathbf{f}_1\left(\xi,\mathbf{l}+\mathbf{v}\right) + 2\boldsymbol{\lambda}^T_2\mathbf{f}_2\left(\xi\right) + 2\boldsymbol{\lambda}^T_3\mathbf{f}_3\left(\mathbf{l}+\mathbf{v}\right) \tag{5.47}$$

With the partial derivative $\frac{1}{2}\frac{\partial\left(\xi-\boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{-1}_{\xi\xi}\left(\xi-\boldsymbol{\mu}\right)}{\partial\xi} = \boldsymbol{\Sigma}^{-1}_{\xi\xi}\left(\xi-\boldsymbol{\mu}\right) = \boldsymbol{\Sigma}^{-1}_{\xi\xi}\left(\xi^0 + d\xi - \boldsymbol{\mu}\right)$ and Eq (5.44) the solution taking the stochastic information observed into consideration can be derived as

$$d\hat{\xi}^1 = -\left(\left(\mathbf{J}^0_1\right)^T \left(\mathbf{J}^0_2\mathbf{P}^{-1}_{total}\left(\mathbf{J}^0_2\right)^T\right)^{-1}\mathbf{J}^0_1 + \boldsymbol{\Sigma}^{-1}_{\xi\xi}\right)^{-1}\left(\left(\mathbf{J}^0_1\right)^T \left(\mathbf{J}^0_2\mathbf{P}^{-1}_{total}\left(\mathbf{J}^0_2\right)^T\right)^{-1}\mathbf{w}^0_{total} + \boldsymbol{\Sigma}^{-1}_{\xi\xi}\left(\xi^0-\boldsymbol{\mu}\right)\right) \tag{5.48}$$

or

$$\begin{aligned} \hat{\xi}^1 &= d\hat{\xi}^1 + \xi^0 \\ &= -\left(\left(\mathbf{J}^0_1\right)^T \left(\mathbf{J}^0_2\mathbf{P}^{-1}_{total}\left(\mathbf{J}^0_2\right)^T\right)^{-1}\mathbf{J}^0_1 + \boldsymbol{\Sigma}^{-1}_{\xi\xi}\right)^{-1}\left(\left(\mathbf{J}^0_1\right)^T \left(\mathbf{J}^0_2\mathbf{P}^{-1}_{total}\left(\mathbf{J}^0_2\right)^T\right)^{-1}\mathbf{w}^0_{total} + \boldsymbol{\Sigma}^{-1}_{\xi\xi}\left(\xi^0-\boldsymbol{\mu}\right)\right) + \xi^0 \\ &= \left(\left(\mathbf{J}^0_1\right)^T \left(\mathbf{J}^0_2\mathbf{P}^{-1}_{total}\left(\mathbf{J}^0_2\right)^T\right)^{-1}\mathbf{J}^0_1 + \boldsymbol{\Sigma}^{-1}_{\xi\xi}\right)^{-1}\left(\left(\mathbf{J}^0_1\right)^T \left(\mathbf{J}^0_2\mathbf{P}^{-1}_{total}\left(\mathbf{J}^0_2\right)^T\right)^{-1}\mathbf{J}^0_1\xi^0 - \left(\mathbf{J}^0_1\right)^T \left(\mathbf{J}^0_2\mathbf{P}^{-1}_{total}\left(\mathbf{J}^0_2\right)^T\right)^{-1}\mathbf{w}^0_{total} + \boldsymbol{\Sigma}^{-1}_{\xi\xi}\boldsymbol{\mu}\right) \end{aligned} \tag{5.49}$$

If there is only the observation equation $\mathbf{f}_1\left(\xi,\mathbf{l}+\mathbf{v}\right)$, the integrated Gauss-Newton solution is identical to the non-linear GHM method proposed by Pope (1972). The formulas proposed in this part will solve some typical geodetic problem presented in Chapter 6.

6 Applications

In the chapter some typical geodetic applications are demonstrated via the solutions proposed in the thesis. The purpose of the examples is to show the applicability of the solutions in geodetic problems and to compare their performances with the one of the existing methods.

In the well-known orthogonal regression problem, our methods are investigated to present advantages according to the convergence behavior and the weight information. In this example the solutions proposed in the thesis provide identical results but different convergence times from the existing methods, and exhibit the applicability without any limitation of the weight information.

Although using the TLS technique to solve the similarity transformation problem is widely discussed recently, the 3D transformation model with different variances and correlations of the errors can be only adjusted within the EIV model by the method proposed in Chapter 5.5. The adjusted result indicates that the estimated scale of the transformation (from the source system to the target system) multiplying the estimated scale of the reverse transformation (from the target system to the source system) equals exactly to 1. The phenomenon denotes the correct symmetric treatment of two set of coordinates for identical point fields. When distances between points are fixed or not, the generalized Gauss Newton algorithm proposed in Chapter 5.5 can solve the transformation problem.

In the next example the quadratic form of point clouds are estimated by means of the WTLS technique. The method can be used for estimating the form parameters of the general quadratic equation when the form is exactly known (e.g., circle) or unknown.

Moreover, the free stationing with stochastic parameters is presented within geodetic networks with the weak datum. The result indicates that our solution is the weighted mean of the WTLS solution and the mean given by the prior stochastic information.

6.1 Orthogonal regression

One of the common applications in geodesy, where the observations in the model matrix $\mathbf{A}$ and the conventional observation vector $\mathbf{y}$ are stochastic, is the orthogonal regression. In this application one minimizes the sum of the squared orthogonal distances instead of the sum of the squared vertical distances from the data to fitting line. The application is frequently used for the surface reconstruction of point clouds in engineering geodesy, and also to find the relationship between some quantities observed. For example, Mann and Emanuel (2006) is concerned with finding statistical evidence for a natural climate cycle, which may be related to Atlantic hurricane activity.

The target of this example is as follows:

1. To validate that our solutions give the identical solution with existing methods under the condition that the vcm is not general.

2. To show the advantages of convergence behavior, especially the iteration times of various solutions.

3. To solve the problem with fully populated vcm: It is emphasized that our solutions can solve the WTLS, and results may significantly differ from the TLS solution.

4. To show the constrained WTLS problem: in this scenario it is also discussed that only our solution can solve the constrained WTLS problem.

As the first scenario, we use the data in Tab 6.1, to solve a simple orthogonal regression problem, where the slope ξ should be estimated. The model can be expressed as follows

$$\mathbf{y}+\mathbf{v}_y=(\mathbf{x}+\mathbf{v}_x)\xi, \tag{6.1}$$

where the intercept is omitted in the model.

The data of this example are given in the following Table. The data are originally given in Neri et al (1989) and also presented in Schaffrin and Wieser (2008) and Shen et al (2010). However, the values of the 4^{th} column for y is changed to be proper for the model (6.1).

point number	x	weight of x	y	weight of y
1	0.0	1000	14.5	1.0
2	0.9	1000	15.1	1.8
3	1.8	500	34.7	4.0
4	2.6	800	37.4	8.0
5	3.3	200	40.4	20
6	4.4	80	58.1	20
7	5.2	60	62.9	70
8	6.1	20	69.9	70
9	6.5	1.8	78.9	100
10	7.4	1.0	86.4	500

Table 6.1 Observation data vector (x, y) and corresponding weights

A comparison of our solutions to the weighted LS and TLS solution is shown in the following Table.

parameter	TLS	WLS	Algorithms 1 until 5
ξ	12.3902755822	11.7614570786	14.0066360261

Table 6.2 Results of orthogonal regression with the data of Table 6.1; TLS solution with vcm, weighted LS using only individual weights of y, WTLS solutions with the uncorrelated weighted information

The WTLS solutions are derived with the algorithms developed in the thesis. The Algorithms 1 to 5 give identical results within the magnitude of 10^{-10}. The differences between the WTLS solutions and the WLS solution, or WTLS solutions and TLS solution are clearly shown in Tab 6.2. This indicates that the WTLS solution can be significantly different to the traditional LS and TLS results, if one take the weights of the uncertainty of the model matrix into account and calculates with the WTLS solutions.

Algorithm	Algorithm 1	Algorithm 2	Algorithm 3	Algorithm 4	Algorithm 5	Non-linear GHM	Schaffrin and Wieser 2008
Times	11	10	9	4	16	13	11

Table 6.3 Number of iteration of different solutions using the WLS solution as the initial values

In the Chapter 4 we have intensively discussed the convergence behavior of the different solutions. Here, we present the convergence speed represented with the iteration times in the following table, where the WLS solution is set as the initial value for the calculation process, and the WTLS solutions converge within the given magnitude (10^{-10}).

The Algorithms 1, 2 and 3 have almost the same iteration times of the algorithm presented in Schaffrin and Wieser (2008), whereas Algorithm 4 (Newton approach) needs less iterations. Though our Algorithms 3 and 5 have the identical analytical formulation as the non-linear GHM algorithm, the iteration times of the non-linear GHM are more than Algorithm 3 and less than Algorithm 5 due to the original algorithm design. The time for one iteration of the methods presented in Tab 6.3 is in the same level. The reason can be that the algorithms except Algorithm 4 have the similar design (the inversion of the same dimension). Furthermore, Algorithm 4 needs only one inversion (without the inversion of the normal matrix) though it requires more matrix multiplication.

Being an one-dimensional problem, it is able to demonstrate the values of the objective function (see Eq (4.19)) according to parameter values. The values of the objective function are plotted in the following left-up figure where our WTLS solution is exactly minimal. In the right-up figure we present the estimated line using the WTLS technique introduced in the thesis. The two figures below indicate the trend of the second and first derivative of the objective function at the region, which is close to the WTLS solution. The objective function is (local) convex at the region. The convexity guarantees the sufficient condition of the Lagrange conditions, because the second derivative of the objective function is positive. Furthermore, the first derivative is 0 at the WTLS solution, which means identically that the WTLS result corresponds to the minimum of the objective function (stationary point and the positive definite Hessian matrix).

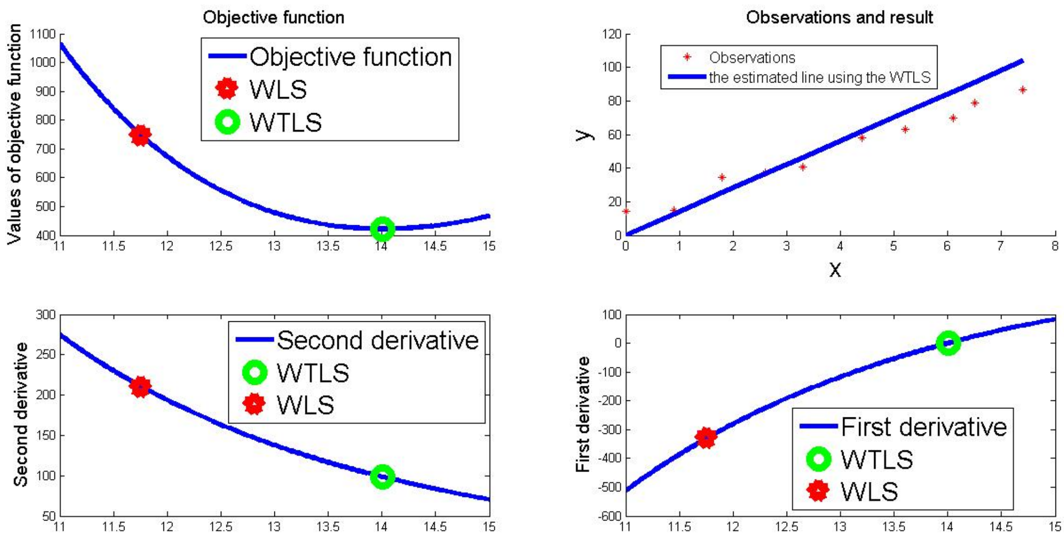


Figure 6.1 The objective function, the observation and the fitting line, the second derivative and the first derivative of the objective function w.r.t. parameter vector

In previous part we have shown that our solutions yield the identical result as Schaffrin and Wieser (2008) in this numerical example. However, the EW-TLS solution (e.g., Schaffrin and Wieser 2008) cannot solve the fully weighted TLS problem. The case of correlated observation is always encountered in geodesy. In order to exhibit the advantage of solutions designed in the thesis, we give the correlation coefficients between the i^{th} observation of x and i^{th} observation of y as 0.1, 0.5, or 0.9 (i =1…10). At the same time, weights are still hold (see Tab 6.1).

The results with various correlation coefficients are presented in Tab 6.4. The bigger correlation coefficient leads to a more significant difference to the WTLS solution without correlation. It means that the fully populated vcm could affect the WTLS result. Furthermore, the difference between the TLS solution and the WTLS solution could be significantly larger than the difference between the TLS and the LS solution (see Tab 6.2 and 6.4).

Parameter	WTLS with correlation coefficient 0.1	WTLS with correlation coefficient 0.5	WTLS with correlation coefficient 0.9	WTLS without correlations
ξ	14.07208090823	14.4438768236	15.3638711544	14.0066360261

Table 6.4 The parameter estimates obtained by using the different vcm, which have a different correlation coefficient between the observation the i^{th} observation of x and i^{th} observation of y

Fixing column

In order to demonstrate the solution of the WTLS problem with fixing columns, the slope ξ_1 and the intercept ξ_2 are estimated and hence the model can be expressed as follows:

$$\mathbf{y} + \mathbf{v}_y = \left[(\mathbf{x} + \mathbf{v}_x), \mathbf{e} \right] \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}, \tag{6.2}$$

Thus, the model becomes more complete in comparison with the model of (6.1).

point number	x	weight of x	Y	weight of y
1	0.0	1,000	5.9	1.0
2	0.9	1,000	5.4	1.8
3	1.8	500	4.4	4.0
4	2.6	800	4.6	8.0
5	3.3	200	3.5	20
6	4.4	80	3.7	20
7	5.2	60	2.8	70
8	6.1	20	2.8	70
9	6.5	1.8	2.4	100
10	7.4	1.0	1.5	500

Table 6.5 Observation data vector (x, y) and corresponding weights, taken from Neri et al. (1989)

Parameter	Unweighted TLS	Solution of Schaffrin and Wieser (2008)	Solutions designed in the thesis
$\hat{\xi}_1$	-0.545561197	-0.480533407	-0.480533407
$\hat{\xi}_2$	5.784043775	5.479910224	5.479910224

Table 6.6 The parameter estimates using the data of Table 6.5

The data of this example can be seen in Tab 6.5. A comparison of our solution to the solution of Schaffrin and Wieser (2008) is shown in Table 6.6. In fact, our solutions of parameters correspond with the solution of Schaffrin and Wieser (2008). After the elimination of the parameter ξ_2 the model can also be transformed as an one-dimensional problem like in Eq (6.1). The comparison of the solutions proposed in the theses with Schaffrin and Wieser (2008) are intensively discussed in Chapter 4. For example, the convergence behavior of the Newton algorithm has the advantage that we can get quadratic convergence rate which is normally not the case with the Gauss-Newton method.

Constrained WTLS

The example in this section demonstrates the use of the WTLS algorithm with linear constraints. The data of the example is taken from Schaffrin and Felus (2009), which easily shows the difference between the algorithm of Schaffrin and Felus (2009) and ours. The data are given as follows

$$\mathbf{A}=\begin{bmatrix}-0.5 & 1 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\\ 1 & 0 & 1\end{bmatrix}\quad \mathbf{y}=\begin{bmatrix}6\\3\\4\\10\end{bmatrix}\quad \text{and}\quad \begin{matrix}\mathbf{K}=\begin{bmatrix}-2 & 0 & 3\end{bmatrix}\\ \kappa_0=16\end{matrix}, \tag{6.3}$$

For the model and the notation refer to Chapter 5.3.

Using the solution proposed in the thesis (see Chapter 5.3) the result of the parameters, squared residuals and fulfillment of constraints are presented in the following table and compared with the result of Schaffrin and Felus (2009). In order to make the comparison to Schaffrin and Felus (2009), the weights for all observations are equal.

	Algorithm proposed in this thesis	Algorithm proposed in Schaffrin and Felus	Result of Schaffrin and Felus
$\hat{\xi}_1$	2.36823	2.36823	2.36819
$\hat{\xi}_2$	5.69850	5.69850	5.69844
$\hat{\xi}_3$	6.91215	6.91215	6.91213
$\mathbf{v}^T\mathbf{v}$	0.21284	0.21284	0.21284

Table 6.7 Results of the parameters and sum of the squared residuals using the constrained WTLS approach

Obviously, first two results do not have the difference. The first result is obtained by the algorithm proposed in the thesis, and the second one is derived by using the algorithm proposed in Schaffrin and Felus (2009). Both algorithms convergence at exactly identical positions which represent the estimate of the parameters at the magnitude of 10^{-5}. The result presented in Schaffrin and Felus (2009) (in 4th column of Tab 6.7) has slight difference to the first two results. Although the sums of squared residuals are identical, the estimates must be same for treating the identical data. However, the differences should be investigated in the future.

The strategy of Schaffrin and Felus (2009) is based on the SVD method and can only solve the TLS problem with a fairly vcm. In geodesy and related science domains, the vcm may be expressed as a Toeplitz matrix (c.f. Xu et al 2007). In this example the vcm of the model matrix **A** is given as a positive definite Toeplitz matrix with the structure:

$$\boldsymbol{\Sigma}_{AA} = \begin{bmatrix} \Gamma_0 & \Gamma_1 & \dots & \Gamma_9 & 0 & 0 \\ \Gamma_1 & \Gamma_0 & & & \Gamma_9 & 0 \\ \dots & \Gamma_1 & \dots & & & \Gamma_9 \\ \Gamma_9 & \dots & \dots & \Gamma_0 & & \dots \\ 0 & \Gamma_9 & & & \Gamma_0 & \Gamma_1 \\ 0 & 0 & \Gamma_9 & \dots & \Gamma_1 & \Gamma_0 \end{bmatrix}, \tag{6.4}$$

where $\Gamma_0 = 1, \Gamma_1 = 0.9, \Gamma_2 = 0.8 \dots \Gamma_8 = 0.2, \Gamma_9 = 0.1$. In Matlab code, this matrix can be written as toeplitz([1:-0.1:0,0]). The correlation coefficients of the vcm are between 0 and 1.

The results are presented as

	TLS	WTLS	LCTLS	LCWTLS
$\hat{\xi}_1$	4.68316	3.52734	2.36823	5.25272
$\hat{\xi}_2$	6.24535	33.75578	5.69850	9.38222
$\hat{\xi}_3$	5.13041	34.74160	6.91215	8.83515
$\mathbf{v}^T\boldsymbol{\Sigma}^{-1}\mathbf{v}$	0.18400	0.26490	0.21284	0.84946

Table 6.8 Results of the parameters and sum of the squared residuals using TLS, WTLS, the constrained WTLS approach with identity and the Toeplitz vcm structure

Tab 6.8 indicates that if one considers the correlation information the results can be significantly changed, especially between the TLS and the WTLS solutions. Though the parameters estimated from the TLS and WTLS cases are totally different, the sums of squared (weighted) residuals are at the same level. The phenomenon may be explained by the objective function $\min_{\xi}(\mathbf{y}-\mathbf{A}\xi)^T(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)$. Some estimates of the WTLS solution can lead to a large value of $(\mathbf{y}-\mathbf{A}\xi)^T(\mathbf{y}-\mathbf{A}\xi)$. However, the WTLS solution has the minimum of the function $(\mathbf{y}-\mathbf{A}\xi)^T(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)$, which can be significantly reduced by $(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}$. The correlated case cannot be solved with the existing methods, e.g., Schaffrin and Felus (2009), because the vcm of the model matrix cannot be expressed as a Kronecker structure. The LCTLS solution differs from the LCWTLS solution, particularly the weighted sum of squared residuals.

6.2 3D similarity transformation

The 3-D similarity transformation is frequently used to transform spatial data from a source coordinate system to a target coordinate system (e.g., WGS84 coordinates to a local datum) in geodesy and the related science domains. The main task of the problem is the estimation of the 7 transformation parameters (scale factor, three rotation angles, and three translations). If the 7 parameters are obtained, an optimal fitting of the data sets is achieved.

It is well-known that the mathematical model of the similarity transformation can be simplified into a linear system if the rotation angles are small enough (see e.g., Yang 1999). However, the solution of the nonlinear system is required if the angles are not small in general. Many LS algorithms are employed to adjust the GMM representing the problem. Two key algorithms are procrustes analysis (e.g., Grafarend and Awange 2003) and the unit quaternion-based approach (e.g., Shen et al. 2006). The standard GMM assumes that the elements of the coefficient matrix are error-free. However, this is not always the case because the elements of the coefficient matrix can be the quantities observed in the 3D similarity problem.

A successful investigation considering the randomness of the coefficient matrix on the 2-D similarity transformation is firstly given by Teunissen (1985, 1988) who obtained a closed form solution based on the singular values decomposition. Bleich and Illner (1989) proposed the linearized GHM method to solve transformation problems. For the similarity transformation they used the constraint that the scale factors are identical in x, y and z directions. To solve the constrained problem pseudo observation equations are applied. From the result obtained by processing the data given in their paper, the method proposed by Bleich and Illner (1989) does not give the optimal solution. The false convergence may be referred to pitfalls discussed in Pope (1972). Felus and Schaffrin (2005) use the Cadzow algorithm for solving the planar similarity problem, which has been modified by Schaffrin et al (2009). The structured TLS solution proposed can only estimate rotation and scale parameters. Felus and Burtch (2009) derive a novel algorithm, which is based on the equivalence between the solutions for the rotation matrix of the GMM and the EIV model in the fairly weighted case. Using the estimated rotation matrix the scale factor is obtained by the meaningful solution of a quadratic equation. Neitzel (2010) uses the iterative Gauss-Newton approach originally proposed by Pope (1972) to estimate the implicit parameters (the scale multiplying the sine and cosine of the angle instead of the scale and angle) for the 2-D similarity transformation in the heteroscedastic case. The transformation model in 2-D can be arranged to a standard EIV model presented in Eqs (7) to (10) in Neitzel (2010). However, the method cannot be explicitly extended for the 3-D similarity transformation problem, as the elements of the rotation matrix for the 3-D problem are totally different. Thus, it is necessary to put forward an algorithm to adjust the 3-D similarity transformation model in the general weighted case.

The 3-D similarity transformation model

The functional model of the similarity transformation in the 3-D space is considered as follows

$$\begin{bmatrix} x_t + v_{x_t} \\ y_t + v_{y_t} \\ z_t + v_{z_t} \end{bmatrix} = \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} + \mu \mathbf{M}^T \begin{bmatrix} x_s + v_{x_s} \\ y_s + v_{y_s} \\ z_s + v_{z_s} \end{bmatrix} \tag{6.5}$$

$\Delta x, \Delta y, \Delta z$ are the three translation parameters, μ is the scale factor. The vector $\begin{bmatrix} x_t & y_t & z_t \end{bmatrix}^T$ represents the coordinates in the target system whereas $\begin{bmatrix} x_s & y_s & z_s \end{bmatrix}^T$ denotes the coordinates in the source system. v_{x_t} , v_{y_t} , v_{z_t} are the corrections of unavoidable errors for the correspondent observations in the target coordinate system, and v_{x_s} , v_{y_s} , v_{z_s} are the corrections of unavoidable errors for the corresponding observations in the source coordinate system. The structure of the rotation matrix $\mathbf{M}$ is here introduced as

$$\begin{aligned} \mathbf{M} &= \mathbf{M_3}(\alpha_3)\mathbf{M_2}(\alpha_2)\mathbf{M_1}(\alpha_1) \\ &= \begin{bmatrix} \cos\alpha_3 & \sin\alpha_3 & 0 \\ -\sin\alpha_3 & \cos\alpha_3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\alpha_2 & 0 & -\sin\alpha_2 \\ 0 & 1 & 0 \\ \sin\alpha_2 & 0 & \cos\alpha_2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha_1 & \sin\alpha_1 \\ 0 & -\sin\alpha_1 & \cos\alpha_1 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}, \end{aligned} \tag{6.6}$$

α_1, α_2 and α_3 are the rotation angles about the x, y and z axes, respectively. $\mathbf{M_1}(\alpha_1)$, $\mathbf{M_2}(\alpha_2)$ and $\mathbf{M_3}(\alpha_3)$ are three skew-symmetric matrices. $\mathbf{M}$ is an orthonormal matrix. a, b and c with indices denote the 9 elements of the rotation matrix. $\mathbf{M}^T$ is the rotation matrix for the single point.

Considering the whole observations, the system is given as

$$\mathbf{Y}+\mathbf{V}_{\mathbf{Y}}=\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}\right)\mu\mathbf{M}+\mathbf{1}_n\otimes\begin{bmatrix}\Delta x & \Delta y & \Delta z\end{bmatrix} \tag{6.7}$$

with

$$\begin{aligned}\mathbf{Y}+\mathbf{V}_{\mathbf{Y}}&=\begin{bmatrix}\vdots & \vdots & \vdots\\ (x_t)_i+(v_{x_t})_i & (y_t)_i+(v_{y_t})_i & (z_t)_i+(v_{z_t})_i\\ \vdots & \vdots & \vdots\end{bmatrix}\\ \mathbf{A}+\mathbf{V}_{\mathbf{A}}&=\begin{bmatrix}\vdots & \vdots & \vdots\\ (x_s)_i+(v_{x_s})_i & (y_s)_i+(v_{y_s})_i & (z_s)_i+(v_{z_s})_i\\ \vdots & \vdots & \vdots\end{bmatrix}\end{aligned} \tag{6.8}$$

where the lower index i denote the ith observation group. $\mathbf{1}_n$ is the vector of ones with dimension $n\times1$.

Thus, the functional model connecting the parameters and observations can be expressed as

$$\text{Model I}: \mathbf{f}\left(\xi,\mathbf{l}+\mathbf{v}\right)=\mathbf{0} \tag{6.9}$$

where $\xi=\begin{bmatrix}\alpha_1 & \alpha_2 & \alpha_3 & \mu & \Delta x & \Delta y & \Delta z\end{bmatrix}$, $\mathbf{l}=\begin{bmatrix}vec(\mathbf{A})\\ vec(\mathbf{Y})\end{bmatrix}$, $\mathbf{v}=\begin{bmatrix}vec(\mathbf{V}_{\mathbf{A}})\\ vec(\mathbf{V}_{\mathbf{Y}})\end{bmatrix}$.

or alternatively formulated as

$$\text{Model II}: \mathbf{f}_1\left(\xi_C,\mathbf{l}+\mathbf{v}\right)=\mathbf{0},\ \mathbf{f}_2\left(\xi_C\right)=\mathbf{0}, \tag{6.10}$$

where $\xi_C=\begin{bmatrix}a_1 & a_2 & a_3 & b_1 & b_2 & b_3 & c_1 & c_2 & c_3 & \mu & \Delta x & \Delta y & \Delta z\end{bmatrix}$.

$\mathbf{f}_1\left(\xi_C,\mathbf{l}+\mathbf{v}\right)=\mathbf{0}$ is the function using all nine elements of the rotation matrix instead of the rotation angles as the parameters and $\mathbf{f}_2\left(\xi_C\right)=\mathbf{0}$ is the constrained function for the orthogonality of the rotation matrix (see Eq (6.19)).

Using Model I

For the transformation problem, the Model I is given as

$$\mathbf{f}\left(\xi,\mathbf{l}+\mathbf{v}\right)=vec\left(\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}\right)\mu\mathbf{M}\right)+\begin{bmatrix}\Delta x & \Delta y & \Delta z\end{bmatrix}^T\otimes\mathbf{1}_n-vec\left(\mathbf{Y}+\mathbf{V}_{\mathbf{Y}}\right)=\mathbf{0}, \tag{6.11}$$

where $\mathbf{M}=\mathbf{M}_3(\alpha_3)\mathbf{M}_2(\alpha_2)\mathbf{M}_1(\alpha_1)$.

According to the discussion of the Chapter 5.5, the partial derivative w.r.t. the parameters should be calculated as

$$\frac{\partial\mathbf{f}\left(\xi,\mathbf{l}+\mathbf{v}\right)}{\partial\alpha_i}=vec\left(\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}\right)\mu\mathbf{M}'_{\alpha i}\right), \tag{6.12}$$

$$\frac{\partial\mathbf{f}\left(\xi,\mathbf{l}+\mathbf{v}\right)}{\partial\mu}=vec\left(\left(\mathbf{A}+\mathbf{V}_{\mathbf{A}}\right)\mathbf{M}\right), \tag{6.13}$$

$$\frac{\partial\mathbf{f}\left(\xi,\mathbf{l}+\mathbf{v}\right)}{\partial\begin{bmatrix}\Delta x & \Delta y & \Delta z\end{bmatrix}}=\mathbf{I}_3\otimes\mathbf{1}_n, \tag{6.14}$$

with

$$\mathbf{M}_{\alpha 1}^{d} = \mathbf{M}_3(\alpha_3)\mathbf{M}_2(\alpha_2)\mathbf{M}_1^{d}(\alpha_1)$$
$$\mathbf{M}_{\alpha 2}^{d} = \mathbf{M}_3(\alpha_3)\mathbf{M}_2^{d}(\alpha_2)\mathbf{M}_1(\alpha_1), \tag{6.15}$$
$$\mathbf{M}_{\alpha 3}^{d} = \mathbf{M}_3^{d}(\alpha_3)\mathbf{M}_2(\alpha_2)\mathbf{M}_1(\alpha_1)$$

where $\mathbf{M}_{\alpha 1}^{d}(\alpha_1) = \frac{d\mathbf{M}_1(\alpha_1)}{d\alpha_1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -\sin\alpha_1 & \cos\alpha_1 \\ 0 & -\cos\alpha_1 & -\sin\alpha_1 \end{bmatrix}$, $\mathbf{M}_{\alpha 2}^{d}(\alpha_2) = \begin{bmatrix} -\sin\alpha_2 & 0 & -\cos\alpha_2 \\ 0 & 0 & 0 \\ \cos\alpha_2 & 0 & -\sin\alpha_2 \end{bmatrix}$ and

$$\mathbf{M}_{\alpha 3}^{d}(\alpha_3) = \begin{bmatrix} -\sin\alpha_3 & \cos\alpha_3 & 0 \\ -\cos\alpha_3 & -\sin\alpha_3 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus, the Jabobian matrix $\mathbf{J}_1$ is obtained by Eqs (6.12), (6.13) and (6.14) as

$$\mathbf{J}_1 = \frac{\partial \mathbf{f}(\xi, \mathbf{l}+\mathbf{v})}{\partial \xi^T}$$
$$= \left[vec\left((\mathbf{A}+\mathbf{V}_\mathbf{A})\mu\mathbf{M}_{\alpha 1}^{d}\right) \quad vec\left((\mathbf{A}+\mathbf{V}_\mathbf{A})\mu\mathbf{M}_{\alpha 2}^{d}\right) \quad vec\left((\mathbf{A}+\mathbf{V}_\mathbf{A})\mu\mathbf{M}_{\alpha 3}^{d}\right) \quad vec\left((\mathbf{A}+\mathbf{V}_\mathbf{A})\mathbf{M}\right) \quad \mathbf{I}_3 \otimes \mathbf{1}_n \right] \tag{6.16}$$

The Jacobian matrix $\mathbf{J}_2$ representing the partial derivative w.r.t. the observations can be also derived as

$$\frac{\partial \mathbf{f}(\xi, \mathbf{l}+\mathbf{v})}{\partial \mathbf{l}^T} = \left[(\mu\mathbf{M})^T \otimes \mathbf{I}_n \quad -\mathbf{I}_3 \otimes \mathbf{I}_n \right], \tag{6.17}$$

due to $vec\left(\mathbf{I}_n (\mathbf{A}+\mathbf{V}_\mathbf{A})\mu\mathbf{M}\right) = \left((\mu\mathbf{M})^T \otimes \mathbf{I}_n\right) vec(\mathbf{A}+\mathbf{V}_\mathbf{A})$.

If the initial values for the parameters and residuals are given, the function can be linearized. In this case the vector $\mathbf{w}^0 = \mathbf{f}(\xi^0, \mathbf{l})$, since there is no quadratic and higher order term of the observations in the functional model (see Chapter 5.5). The solution can be obtained by iterative Gauss-Newton algorithm (see Eqs (5.45) and (5.46)).

Using Model II

Analogously, Model II reads

$$\mathbf{f}_1(\xi_C, \mathbf{l}+\mathbf{v}) = vec\left((\mathbf{A}+\mathbf{V}_\mathbf{A})\mu\mathbf{M}\right) + [\Delta x \quad \Delta y \quad \Delta z]^T \otimes \mathbf{1}_n - vec(\mathbf{Y}+\mathbf{V}_\mathbf{Y}) = \mathbf{0}, \tag{6.18}$$

where $\mathbf{M} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}$, with the constraints $\mathbf{f}_2(\xi_C)$ represented by

$$\begin{aligned} &a_1^2 + a_2^2 + a_3^2 - 1 = 0 \\ &b_1^2 + b_2^2 + b_3^2 - 1 = 0 \\ &c_1^2 + c_2^2 + c_3^2 - 1 = 0 \\ &a_1 a_2 + b_1 b_2 + c_1 c_2 = 0 \\ &a_1 a_3 + b_1 b_3 + c_1 c_3 = 0 \\ &a_2 a_3 + b_2 b_3 + c_2 c_3 = 0 \end{aligned} \tag{6.19}$$

which fulfill the conditions of the orthonormal matrix. Note that based on the 6 constraints both functional models (Eq (6.9)and Eq (6.10)) have identical degrees of freedom.

Here, the constraints are arranged as the pseudo observation equations $\mathbf{f}_2(\xi_C) \approx \mathbf{0} + \mathbf{v}_{pseudo}$ (see Chapter 5.5). In this case, the Jacobian matrix $\mathbf{J}_{1C}$ is given as

$$\mathbf{J}_{1C} = \begin{bmatrix} \dfrac{\partial \mathbf{f}_1(\xi_C, \mathbf{l}+\mathbf{v})}{\partial \xi_C^T} \\ \dfrac{\mathbf{f}_2(\xi_C)}{\partial \xi_C^T} \end{bmatrix} \tag{6.20}$$

with

$$\begin{aligned} &\frac{\partial \mathbf{f}_1(\xi_C, \mathbf{l}+\mathbf{v})}{\partial \xi_C^T} = \\ &\begin{bmatrix} vec\left((\mathbf{A}+\mathbf{V_A})\mu\mathbf{M}_{a1}^d\right) & \ldots & vec\left((\mathbf{A}+\mathbf{V_A})\mu\mathbf{M}_{c3}^d\right) & vec\left((\mathbf{A}+\mathbf{V_A})\mathbf{M}\right) & \mathbf{I}_3 \otimes \mathbf{1}_n \end{bmatrix} \end{aligned} \tag{6.21}$$

and

$$\begin{aligned} &\frac{\mathbf{f}_2(\xi_C)}{\partial \xi_C^T} = \\ &\begin{bmatrix} 2a_1 & 2a_2 & 2a_3 & 0 & 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 2b_1 & 2b_2 & 2b_3 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & 0 & 2c_1 & 2c_2 & 2c_3 & \mathbf{0}_{6\times 4} \\ a_2 & a_1 & 0 & b_2 & b_1 & 0 & c_2 & c_1 & 0 & \\ a_3 & 0 & a_1 & b_3 & 0 & b_1 & c_3 & 0 & c_1 & \\ 0 & a_3 & a_2 & 0 & b_3 & b_2 & 0 & c_3 & c_2 & \end{bmatrix} \end{aligned} \tag{6.22}$$

where $\mathbf{M}_{a1}^d = \dfrac{d\mathbf{M}}{da_1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \ldots, \mathbf{M}_{c3}^d = \dfrac{d\mathbf{M}}{dc_3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

The Jacobian matrix $\mathbf{J}_{2C}$ representing the partial derivative w.r.t. the observations and pseudo observations can be given as

$$\mathbf{J}_{2C} = \frac{\partial \begin{bmatrix} \mathbf{f}_1(\xi_C, \mathbf{l}+\mathbf{v}) \\ \mathbf{f}_2(\xi_C) \end{bmatrix}}{\partial \mathbf{v}_{total}^T} = \begin{bmatrix} (\mu\mathbf{M})^T \otimes \mathbf{I}_n & -\mathbf{I}_3 \otimes \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_6 \end{bmatrix}, \tag{6.23}$$

where $\mathbf{v}_{total} = \begin{bmatrix} \mathbf{v} \\ \mathbf{v}_{pseudo} \end{bmatrix}$.

In this functional model there is no quadratic and higher order term of observations. Thus, the vector $\mathbf{w}_{total} = \begin{bmatrix} \mathbf{f}_1(\xi^i, \mathbf{l}) \\ \mathbf{f}_2(\xi^i_C) \end{bmatrix}$. The corresponding weight matrix for the vector $\mathbf{v}_{total}$ is $\mathbf{P}_{total} = \begin{pmatrix} \mathbf{Q}_{\mathbf{ll}}^{-1} & \mathbf{0} \\ \mathbf{0} & \delta\mathbf{I}_6 \end{pmatrix}$, where δ is a given constant variance level at which the six constrained equations should be. By iterative Gauss-Newton algorithm (updating the Jabobian matrices $\mathbf{J}^i_{1C}$, $\mathbf{J}^i_{2C}$ and vector $\mathbf{w}^i_{total}$) the solution can be obtained (see Eqs (5.45) and (5.46)).

Experiments

The experiments are demonstrated to present our proposed methods. The data in the first example originates from Felus and Burtch (2009). This example is based on an actual experiment of fitting two surfaces that are surveyed in two different reference systems (datum's). Four control points are identified and recorded in the two coordinate systems. The data in the source and target system are represented in the matrices $\mathbf{A}$ and $\mathbf{Y}$, respectively.

$$\mathbf{A} = \begin{bmatrix} 30 & 40 & 10 \\ 100 & 40 & 10 \\ 100 & 130 & 10 \\ 30 & 130 & 10 \end{bmatrix} \text{ and } \mathbf{Y} = \begin{bmatrix} 290 & 150 & 15 \\ 420 & 80 & 2 \\ 540 & 200 & 20 \\ 390 & 300 & 5 \end{bmatrix} \tag{6.24}$$

In order to solve this adjustment problem, Felus and Burtch (2009) proposed a new algorithm which employs a closed-form Procrustes method to obtain the rotation matrix, while the scale parameter is derived later from the solution of a quadratic equation. However, the method based on the matrix decomposition cannot solve the problem with a general weight matrix. Thus, we organize the demonstration as follows:

1. These points are assumed to be measured using the same method and have the same statistical properties, namely $\mathbf{P} = \mathbf{I}_{24}$, and it is assumed that the weight matrix is expressed as a Kronecker product $\mathbf{P} = \mathbf{I}_6 \otimes diag(1, 4, 6.25, 16)$.

2. The results using the fully populated weight matrix are given.

3. The distance between two points is fixed

4. The method is also valid for the real data.

The algorithms based on the Model I and II are implemented for solving the nonlinear 3D similarity transformation. The standard assumptions (no translations, no rotations, scale equals to 1) are given for the initial value of the parameters whereas the initial values of the residuals are zeros. The transformation parameters are estimated with the identity weight matrix ($\mathbf{P} = \mathbf{I}_{24}$) and the fairly general weight matrix ($\mathbf{P} = \mathbf{I}_6 \otimes diag(1, 4, 6.25, 16)$). Tab 6.9 displays that the solutions based on Model I and II yield the identical result to Felus and Burtch (2009) (The estimated elements in the rotation matrix are transformed to the angles for the Model II). Note that Felus and Burtch (2009) presents $\mathbf{v}^T\mathbf{v}$ instead of $\mathbf{v}^T\mathbf{\Sigma}^{-1}\mathbf{v}$. We do not discuss the difference between TLS solution and LS solution in general since this is widely done in the geodetic literature e.g., Schaffrin and Felus (2008). But the argument 'even small differences between LS and TLS solution may result in significant differences which can exceed the required 5cm accuracy' (Schaffrin and Felus 2008) is also valid for the difference between the TLS and WTLS solution.

For Model 2 the weights for the six constrained equations are 10^{12} and the constraints for the fulfilling the orthogonality are given in Tab 6.10, which shows that the conditions of the constraints are numerical successfully fulfilled. For the example, the residuals are not shown since they are also identical to those presented by Felus and Burtch (2009).

Parameters	The weight matrix	Algorithm using the Model I	Algorithm using the Model II	Algorithm (Felus and Burtch 2009)
Translation Δx	$\mathbf{P}=\mathbf{I}_{24}$	193.01696	193.01696	193.01696
	$\mathbf{P}=\mathbf{I}_6\otimes diag(1,4,6.25,16)$	188.97714	188.97714	188.97714
Translation Δy	$\mathbf{P}=\mathbf{I}_{24}$	117.40274	117.40274	117.40274
	$\mathbf{P}=\mathbf{I}_6\otimes diag(1,4,6.25,16)$	101.51720	101.51720	101.51720
Translation Δz	$\mathbf{P}=\mathbf{I}_{24}$	-15.40738	-15.40738	-15.40738
	$\mathbf{P}=\mathbf{I}_6\otimes diag(1,4,6.25,16)$	-33.38008	-33.38008	-33.38008
Scale μ	$\mathbf{P}=\mathbf{I}_{24}$	2.1216362	2.1216362	2.1216362
	$\mathbf{P}=\mathbf{I}_6\otimes diag(1,4,6.25,16)$	2.1761269	2.1761269	2.1761269
Rotation α_1	$\mathbf{P}=\mathbf{I}_{24}$	-0°54′45.3561″	-0°54′45.3561″	-0°54′45.3561″
	$\mathbf{P}=\mathbf{I}_6\otimes diag(1,4,6.25,16)$	-0°30′51.4666″	-0°30′51.4666″	-0°30′51.4666″
Rotation α_2	$\mathbf{P}=\mathbf{I}_{24}$	0°57′47.4029″	0°57′47.4029″	0°57′47.4029″
	$\mathbf{P}=\mathbf{I}_6\otimes diag(1,4,6.25,16)$	4°31′21.1255″	4°31′21.1255″	4°31′21.1255″
Rotation α_3	$\mathbf{P}=\mathbf{I}_{24}$	35°49′30.6166″	35°49′30.6166″	35°49′30.6166″
	$\mathbf{P}=\mathbf{I}_6\otimes diag(1,4,6.25,16)$	33°32′19.5111″	33°32′19.5111″	33°32′19.5111″
$\mathbf{v}^T\mathbf{\Sigma}^{-1}\mathbf{v}$	$\mathbf{P}=\mathbf{I}_{24}$	236.89	236.89	236.89
	$\mathbf{P}=\mathbf{I}_6\otimes diag(1,4,6.25,16)$	1359.20	1359.20	1359.20

Table 6.9 The Comparison of the estimated transformation parameters to Felus and Burtch (2009)

Constraints	**Values**	**Constraints**	**Values**
$\hat{a}_1^2+\hat{a}_2^2+\hat{a}_3^2-1$	-3.42789e-09	$\hat{a}_1\hat{a}_2+\hat{b}_1\hat{b}_2+\hat{c}_1\hat{c}_2$	-3.98909e-09
$\hat{b}_1^2+\hat{b}_2^2+\hat{b}_3^2-1$	3.46923e-09	$\hat{a}_1\hat{a}_3+\hat{b}_1\hat{b}_3+\hat{c}_1\hat{c}_3$	9.45261e-11
$\hat{c}_1^2+\hat{c}_2^2+\hat{c}_3^2-1$	-4.13325e-11	$\hat{a}_2\hat{a}_3+\hat{b}_2\hat{b}_3+\hat{c}_2\hat{c}_3$	-2.38012e-10

Table 6.10 The fulfillment of the pseudo equations

In order to highlight the advantage of our method with respect to the general weight information, the weight matrix with the Toeplitz structure is given

$$\mathbf{P}=\begin{bmatrix} 1 & 0.95 & 0.9 & & 0 & 0 & 0 \\ 0.95 & 1 & 0.95 & \vdots & 0 & 0 & 0 \\ 0.9 & 0.95 & 1 & & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & & 1 & 0.95 & 0.9 \\ 0 & 0 & 0 & \vdots & 0.95 & 1 & 0.95 \\ 0 & 0 & 0 & & 0.9 & 0.95 & 1 \end{bmatrix}, \tag{6.25}$$

which is generated by the Matlab function P =toeplitz([1:-0.05:0,zeros(1,3)]). Note that the weight matrix is positive definite, and correlation coefficients of the corresponding cofactor matrix are between -0.7010 and 1.

The approach proposed in Felus and Burtch (2009) is based on matrix decomposition and it cannot solve the problem with such a weight matrix. Furthermore, in Schaffrin and Felus (2008) the serious limitations of the method based on the SVD are discussed (e.g., the accuracy of the method based on SVD may be a concern for applications that require high precision, because the SVD operation is only guaranteed to have a few correct digits). Certainly, the fully populated weight matrix can be frequently encountered in error adjustment. The algorithms based on Model I and II are tested for the problem of the weight matrix. Here, we do not only transform coordinates of the source system to coordinates of the target system but also from the target system to the source system (reverse transformation). The results based on Model I and II are identical for transformation and reverse transformation and presented in the following table

Parameters	Algorithms proposed in the thesis Source system to Target system	Parameters Reverse transformation	Algorithms proposed in the thesis Target system to source system
Translation Δx	199.69690	Translation Δx_r	-42.82873
Translation Δy	120.25074	Translation Δy_r	-102.84880
Translation Δz	-23.04864	Translation Δz_r	16.56097
Scale μ	2.0796914	Scale μ_r	0.4808406
Rotation α_1	-0°25′50.3770″	Rotation α_{1r}	2°7′54.6758″
Rotation α_2	3°1′14.2924″	Rotation α_{2r}	-2°10′59.9964″
Rotation α_3	36°10′41.0091″	Rotation α_{3r}	-36°13′48.1482″
$\mathbf{v}^T\mathbf{Pv}$	8.39	$\left(\mathbf{v}^T\mathbf{Pv}\right)_r$	8.39

Table 6.11 The result of the transformation parameter and sum of squared residuals for the 3D similarity transformation and the reverse transformation

The two important quantities to check the correctness of the solutions are the sum of the residuals squared and the scale μ multiplying the scale μ_r of the reverse transformation. Teunissen (1998) claims that the scale multiplying the scale of the reverse transformation in 2D similarity transformation equals to 1 if the model is considered as an EIV model with an identity weight matrix. Tab 6.11 indicates that the scale of the transformation multiplying the scale of the reverse transformation equals exactly to 1 (2.0796914*0.4808406=1). Of course, the identity $\mu \cdot \mu_r = 1$ is not fulfilled if one adjusts the 3D similarity transformation within the GMM. The sums of the squared residuals for the both transformations are identical, since the optimal fitting for both transformations are achieved. The sum of the squared residuals (8.39, see Tab 6.11) is significantly smaller than one in the equal case (236.89, see Tab 6.9) since the residuals are strongly correlated. For example: $\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0.95 \\ 0.95 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0.1 \ll 2 = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

The translation parameters for both transformations can be controlled by the formula $-\begin{bmatrix} \Delta x_r & \Delta y_r & \Delta z_r \end{bmatrix} \mathbf{M}_r^T = \begin{bmatrix} \Delta x & \Delta y & \Delta z \end{bmatrix} \cdot \mu_r$, where $\mathbf{M}_r$ is the rotation matrix of the reverse transformation. If the rotation angles are not small, the translation parameters for both transformations should be not proportional. The relationship between the two rotation matrices is $\mathbf{M}_r^T = \mathbf{M}$. The corresponding rotation angels (e.g., α_1 to α_{1r}) do not have the same absolute values, because the rotation matrix is not skew symmetric.

Considering the fixing of distance between two points in the transformation problem is also useful. In geodesy, the distance between two GPS antennas mounted in aircraft for gravimetry is known and fixed. In an integrated vehicle navigation system, the distances among various sensors are invariant when in motion (Yang et al. 2010). Therefore they can be used as constraints within the estimated observations $\mathbf{f}_3(\mathbf{l}+\mathbf{v}) = \mathbf{0}$. Although the 3D similarity transformation with constraints within the GMM is discussed in Shen et al. (2006) with the quaternion approach, the transformation with constraints within the EIV model is given neither for the equal weight case nor for the unequal weight case. In order to demonstrate the approach proposed in this thesis, we assume the baseline-length between the 3[th] point and 4[th] point in the source system (third and fourth rows of matrix $\mathbf{A}$ in Eq (6.24)) is fixed to its original length 70. The algorithm based on Model II is implemented, since we want to present the solution of the complete combination of $\mathbf{f}_1(\boldsymbol{\xi}_C, \mathbf{l}+\mathbf{v}) = \mathbf{0}$, $\mathbf{f}_2(\boldsymbol{\xi}_C) = \mathbf{0}$ and $\mathbf{f}_3(\mathbf{l}+\mathbf{v}) = \mathbf{0}$. Thus, both Jacobian matrices are

$$\mathbf{J}_{1CC} = \begin{bmatrix} \mathbf{J}_{1C} \\ \mathbf{0}_{1\times13} \end{bmatrix}, \tag{6.26}$$

and

$$\mathbf{J}_{2CC} = \begin{bmatrix} & & & & \mathbf{J}_{2C} & & & \\ \mathbf{0}_{1\times6} & \frac{2a_{31}}{s} & \frac{2a_{32}}{s} & \frac{2a_{33}}{s} & \frac{-2a_{41}}{s} & \frac{-2a_{42}}{s} & \frac{-2a_{43}}{s} & \mathbf{0}_{1\times12} \end{bmatrix}, \tag{6.27}$$

where $s = \sqrt{(a_{31}-a_{41})^2 + (a_{32}-a_{42})^2 + (a_{33}-a_{43})^2}$.

The parameters are estimated in the equal weighted case, and result is not completely given here. The sum of squared residuals is 563.63, which is larger than 236.89 (see Tab 6.9) due to the constraints. The estimated baseline length minus 70 is 1.42e-014, which indicates that the constraint is numerically fulfilled. Meanwhile, the six constraints for the orthogonality of the rotation matrix are fulfilled in magnitude of about 1e-008.

Solutions based on Model I and II are tested also on a real example involving a geodetic datum conversion. The data are originally provided by the Coordinate System Analysis Team of the office of GEOINT Sciences at the National Geospatial-Intelligence Agency and presented in the Felus and Burtch (2009). The data set includes the rectangular coordinates of six control points in Tunisia. The rectangular coordinates were converted from measured geographic coordinates (latitude, longitude, and ellipsoidal heights) in WGS84 datum and in the local CARTHAGE datum (CGE- a Clarke 1880 ellipsoid).

The algorithms proposed, for the real data, are based on the Model I and II and are implemented. The results derived using the method proposed in the thesis are totally identical to the results presented in Felus and Burtch (2009) in equally weight case. When processing the real data the inversion of the matrix $\left(\mathbf{J}_1^T\left(\mathbf{J}_2\mathbf{P}^{-1}\mathbf{J}_2^T\right)^{-1}\mathbf{J}_1\right)$ or $\left(\mathbf{J}_{1C}^T\left(\mathbf{J}_{2C}\mathbf{P}_{total}^{-1}\mathbf{J}_{2C}^T\right)^{-1}\mathbf{J}_{1C}\right)$ may be close to singularity. In this case we can use the idempotent matrices proposed in Eq (5.8) multiplying the data matrix to significantly reduce the conditional number of the matrix required to be inverted.

6.3 Quadratic form analysis

In engineering geodesy there are some methods for fitting a surface to a point cloud, such as, Delaunay triangulation, Non-Uniform rational B-Splines and the estimation of a quadratic form (c.f. Eling 2010). In the thesis of Eling a dam was scanned using a terrestrial laser scanner of Geodetic Institute of Hannover in order to detect the deformations. In order to segment the point cloud, the scanned points of the dam was modeled by a best fitted ellipsoid with 3 translations and only one rotation axes from the original coordinate system. As in a typical deformation scenario the shape of the objects changes due to bending and flexing, a thorough but information preserving parameter reduction is needed by a set of a few characteristic parameters. This is often possible by describing the objects or parts of them using quadratic forms (Hesse and Kutterer 2006). Moreover, it is well-known that one important task of processing scanned data is registration. The parameters obtained by registration can transform the different point clouds in an identical coordinate system. For laser scan it is not possible to specifically measure the discrete points. Therefore, the artificial target is usually used to obtain the position of a single point. The type of the reference targets is described in Table 2.2 in Eling (2010). One popular target is the pass sphere, which can be scanned and then be adjusted to obtain the center of the ball for the late registration of point clouds.

The algorithm to fit the quadratic form has been proposed by Drixler (1993), and also used in Hesse and Kutterer (2006). In Niemeier (2002) and Kupferer (2005) the methods for fitting the special quadratic form 'circle' are discussed. Unfortunately, some linearization pitfalls are not avoided in the approaches proposed. The exact solution for the circle fitting can be found in Schaffrin and Snow (2010) in geodesy, who use the iterative GHM method to give the solution for the fitting problem. Here, methods of Drixler (1993) and Schaffrin and Snow (2010) are briefly presented with fitting circle as the special form of the 2D quadratic form. Then, our solution proposed in Chapter 5.5 is demonstrated to compare the methods in the case of the known quadratic form. At the end, the parameters can be also obtained by our method if the quadratic form is not exactly known. It is also presented that our method is an optimal method in the sense of the minimal sum of squared residuals.

Method of Drixler (1993) for circle fitting

The quadratic form for the 2D is given in e.g., Drixler (1993) as

$$\begin{bmatrix} x_i & y_i \end{bmatrix}\begin{bmatrix} a_1 & a_3 \\ a_3 & a_2 \end{bmatrix}\begin{bmatrix} x_i \\ y_i \end{bmatrix}+\begin{bmatrix} x_i & y_i \end{bmatrix}\begin{bmatrix} a_4 \\ a_5 \end{bmatrix}+a_6=0, \tag{6.28}$$

where $\begin{bmatrix} x_i & y_i \end{bmatrix}$ represents the observations of i[th] point. If the quadratic form is confirmed as a circle, the estimation of the quadratic parameters is obtained in Eq (5.2.17) of Drixler (1993) as follows

$$\begin{bmatrix} \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \end{bmatrix} = -\left(\mathbf{A}_2^T \mathbf{A}_2\right)^{-1} \mathbf{A}_2^T \mathbf{A}_1 , \tag{6.29}$$

where $\mathbf{A}_1 = -\begin{bmatrix} \vdots \\ x_i^2 + y_i^2 \\ \vdots \end{bmatrix}$ and $\mathbf{A}_2 = -\begin{bmatrix} \mathbf{x} & \mathbf{y} & \mathbf{1}_n \end{bmatrix}$ with $\mathbf{x} = \begin{bmatrix} \vdots \\ x_i \\ \vdots \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} \vdots \\ y_i \\ \vdots \end{bmatrix}$. The first 3 parameters are fixed as 1, 1, 0 for the circle fitting (see Drixler 1993).

Method of Schaffrin and Snow (2010) for circle fitting

For the circle fitting Schaffrin and Snow (2010) give the original functional model of a circle as

$$f_i(\xi, \mathbf{x}, \mathbf{y}) = (x_i - x_m)^2 + (y_i - y_m)^2 - r^2 \approx 0 , \tag{6.30}$$

where $\xi = \begin{bmatrix} x_m, y_m, r \end{bmatrix}^T$. In this thesis Eq (6.28) is called as the general function or equation of the quadratic form (using the parameter a_1 to a_6) whereas the equations such as Eq (6.30) are termed as the original equation of the quadratic form (using parameter ξ).

Being a non-linear system, the model can be linearized as follows

$$f\left(\xi^0, \mathbf{x}, \mathbf{y}\right) + \mathbf{A} d\xi + \mathbf{B}\mathbf{v} = \mathbf{0} , \tag{6.31}$$

where

$$\begin{aligned} \mathbf{A} &= -\left[2\left(\mathbf{x} - x_m^0 \mathbf{1}_n\right), 2\left(\mathbf{y} - y_m^0 \mathbf{1}_n\right), 2 r_m^0 \mathbf{1}_n \right] \\ \mathbf{B} &= \left[2\left(\mathbf{x} - x_m^0 \mathbf{1}_n\right), 2\left(\mathbf{y} - y_m^0 \mathbf{1}_n\right) \right] \end{aligned} \tag{6.32}$$

In the next iteration, the approximations $\mathbf{v}^1$ and ξ^1 are used after stripping the solutions $\hat{\mathbf{v}}^1$ and $\hat{\xi}^1$ of their random character (see Chapter 4.4). Then, the linearized model can be expressed in the GHM as

$$\mathbf{w}^1 + \mathbf{A}^1 d\xi + \mathbf{B}^1 \mathbf{v} = \mathbf{0} , \tag{6.33}$$

where

$$\begin{aligned} \mathbf{A}^1 &= -\left[2\left(\mathbf{x}^1 - x_m^1 \mathbf{1}_n\right), 2\left(\mathbf{y}^1 - y_m^1 \mathbf{1}_n\right), 2 r_m^1 \mathbf{1}_n \right] \\ \mathbf{B}^1 &= \left[2\left(\mathbf{x}^1 - x_m^1 \mathbf{1}_n\right), 2\left(\mathbf{y}^1 - y_m^1 \mathbf{1}_n\right) \right] \\ \xi^1 &= \left[x_m^1, y_m^1, r_m^1 \right] \\ f\left(\xi^1, \mathbf{x}^1, \mathbf{y}^1\right) &- \mathbf{B}^1 \mathbf{v}^1 = \mathbf{w}^1 \end{aligned} \tag{6.34}$$

and $\mathbf{x}^1 = \mathbf{x} + \mathbf{v}_x^1$, $\mathbf{y}^1 = \mathbf{y} + \mathbf{v}_y^1$.

Method proposed in this thesis for circle fitting

Some parameters are fixed if one transforms the original functional model (6.30) to the general quadratic form (6.28), which leads to the constrained problem. Thus, we formulate the circle fitting problem as

$$\begin{aligned} & a_1 x_i^2 + a_2 y_i^2 + a_3 x_i y_i + a_4 x_i + a_5 y_i + a_6 = 0 \\ & \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \end{aligned}, \tag{6.35}$$

According to the discussion of Chapter 5.5 the both Jacobian matrices for the first iteration are

$$\mathbf{J}_1^0 = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_i x_i & x_i y_i & y_i y_i & x_i & y_i & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}, \tag{6.36}$$

and

$$\mathbf{J}_2^0 = \begin{bmatrix} diag\left(a_1\mathbf{x} + a_3\mathbf{y} + a_4\right), diag\left(a_2\mathbf{y} + 2a_3\mathbf{x} + a_5\right) & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 \end{bmatrix}, \tag{6.37}$$

where the constraints $\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ are regarded as pseudo observation equations. The last 3 rows of both Jacobian matrices denote the pseudo observation equations representing the constraints in the estimation process.

Meanwhile, the vector $\mathbf{w}^0$ is given as

$$\mathbf{w}^0 = \begin{bmatrix} f\left(\xi^0, \mathbf{x}^0, \mathbf{y}^0\right) \\ a_1^0 - 1 \\ a_2^0 - 1 \\ a_3^0 \end{bmatrix} - \mathbf{J}_2^0 \mathbf{v}_{total}^0, \tag{6.38}$$

where $\mathbf{v}_{total}^0 = \begin{bmatrix} \mathbf{v}_\mathbf{x} \\ \mathbf{v}_\mathbf{y} \\ \mathbf{v}_{pseudo} \end{bmatrix} = \mathbf{0}$, $\mathbf{x}^0 = \mathbf{x}$ and $\mathbf{y}^0 = \mathbf{y}$.

In the next iteration the Jacobian matrices and the vector $\mathbf{w}^0$ should be updated with the estimated parameters and residuals after stripping their randomness

$$\mathbf{J}_1^1 = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_i^1 x_i^1 & y_i^1 y_i^1 & x_i^1 y_i^1 & x_i^1 & y_i^1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}, \tag{6.39}$$

$$\mathbf{J}_2^1 = \begin{bmatrix} diag\left(a_1^1\mathbf{x}^1 + a_3^1\mathbf{y}^1 + a_4^1\right), diag\left(a_2^1\mathbf{y}^1 + 2a_3^1\mathbf{x} \ + a_5^1\right) & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_3 \end{bmatrix}, \tag{6.40}$$

and

$$\mathbf{w}^1 = \begin{bmatrix} f\left(\xi^1, \mathbf{x}^1, \mathbf{y}^1\right) \\ a_1^1 - 1 \\ a_2^1 - 1 \\ a_3^1 \end{bmatrix} - \mathbf{J}_2^1\mathbf{v}^1, \tag{6.41}$$

where $\mathbf{x}^1 = \mathbf{x} + \mathbf{v}_x$ and $\mathbf{y}^1 = \mathbf{y} \ + \mathbf{v}_y$.

Comparison of results

The data demonstrated came from Gander et al. (1994) who fit a circle to six data points using a parametric form of the circle.

Points number	X coordinates	Y coordinates
1	1	7
2	2	6
3	3	8
4	7	7
5	9	5
6	3	7

Table 6.12 The data for circle fitting from Gander et al. (1994)

With this data we present the results of the three methods discussed as

	Method of Drixler 1993	Method of Schaffrin and Snow 2010	Method proposed in the thesis
x_m	4.742331	4.739782	4.739782
y_m	3.835123	2.983533	2.983533
r	4.108762	4.714226	4.714226
$\mathbf{v}^T\mathbf{P}\mathbf{v}$	1.398289	1.227599	1.227599

Table 6.13 The result of parameter and sum of squared residuals for the circle fitting example

The approach of Drixler (1993) without any iteration cannot obtain the optimal solution, which can be seen by comparing the sum of the squared residuals. As the method of Drixler (1993) is relatively simple, it can be used for the approximation of the exact solution.

Tab 6.13 indicates also that our result is exactly identical to the TLS solution proposed in Schaffrin and Snow (2010) for the circle fitting (we transform the parameters a_1 until a_6 to the parameters given in Tab 6.13). Although the method proposed by Schaffrin and Snow (2010) and our method are Gauss-Newton type, the iterative processes are totally different. The advantage of our method is that the linearization of the general equation (6.28) of the quadratic form is very simple (also for the 3D case). Schaffrin and Snow (2010) proposed the method for fitting a circle. If one generalizes the method in another quadratic form, the linearization for the original function representing the quadratic form (e.g., (6.30) for the circle) w.r.t. the parameters and observations may be very difficult. Furthermore, one needs to linearize the original function for any special case. For example, in 3D case except the sphere, three orientation parameters should be considered for the quadratic form. Here, the original function of the rotated ellipsoid is given as

$$\begin{gathered}\frac{(x_o - x_m)^2}{a^2} + \frac{(y_o - y_m)^2}{b^2} + \frac{(z_o - z_m)^2}{c^2} = 1 \\ \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix}_i = \mathbf{R}_{3\times 3} \begin{bmatrix} x \\ y \\ z \end{bmatrix}_i \end{gathered} , \tag{6.42}$$

where $\begin{bmatrix} x \\ y \\ z \end{bmatrix}_i$ are the original observations. $\begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix}_i$ are observations transformed by the rotation matrix $\mathbf{R}_{3\times 3}$, which contains 3 parameters (3 rotation angles).

If one wants to linearize the model (6.42) explicitly w.r.t. the parameters (e.g., translations, rotations) and the observations, the task is rather complicated. However, with the method proposed in this thesis the estimation is relative simple, because the linearization of the quadratic function w.r.t. parameters ($a_1 ... a_{10}$ in 3D case) and the observations is identical. In this case, the constraints can be given as e.g., $a_{10} = 1$, with which the degrees of freedom (10 paramters-1 constraint) using the equation of the general quadratic form are identical to the degrees of freedom of Eq (6.42) (9 parameters: 3 rotations, 3 translations and 3 regular ellipsoid parameters). Based on the discussion above, the TLS method proposed in the thesis provides the exact TLS solution and can be easily generalized for the estimation of any quadratic forms.

The estimation method for the unknown form

If the actual type of the quadratic form (e.g., circle, parabola) is unknown, Drixler (1993) porposed a method based on the eigenvalue decomposition in order to obtain the quadratic parameters. In the estimation process the quadratic form of the observations is approximated as the linear term (i.e. $\left(x_i + v_{x_i}\right)^2$ where $v_{x_i} \sim \left(0, \sigma^2 q_{x_i}\right)$ is replaced by $\tau_i = x_i^2$ with the variance $2\sigma^2 q_{x_i} x_i$). This approximation cannot give the optimal solution and it is suitable for the initial values of our iterative estimation process.

In order to demonstrate the difference between our results and the results of Drixler (1993), an ellipsoid with 400 points is generated with the function $\frac{x^2}{4^2}+\frac{y^2}{3^2}+\frac{z^2}{1^2}=1$; noise is added with the normal distribution $N\left(0,\ 0.05^2\right)$ in all three spatial directions x, y and z. The parameter vector is firstly obtained by the eigenvalue decomposition method of Drixler (1993). Then, the constraint is given that a_{10} equals to the value of a_{10} obtained by the eigenvalue decomposition method. Based on the constraint the iterative process can be implemented. Hence, the constrained model is given as

$$\begin{aligned} &a_1 x_i^2 + a_2 y_i^2 + a_3 z_i^2 + a_4 x_i y_i + a_5 x_i z_i + a_6 x_i z_i + a_7 x_i + a_8 y_i + a_9 z_i + a_{10} = 0 \\ &a_{10} = \left(\hat{a}_{10}\right)_{eig} \end{aligned} \tag{6.43}$$

According to the discussion in Chapter 5.5 the two Jacobian matrices for the first iteration are

$$\mathbf{J}_1^0 = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_i x_i & y_i y_i & z_i z_i & x_i y_i & x_i z_i & y_i z_i & x_i & y_i & z_i & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{6.44}$$

and

$$\mathbf{J}_2^0 = \begin{bmatrix} diag\left(2a_1\mathbf{x}+a_4\mathbf{y}+a_5\mathbf{z}+a_7\right), diag\left(2a_2\mathbf{y}+a_4\mathbf{x}+a_6\mathbf{z}+a_8\right), diag\left(2a_3\mathbf{z}+a_5\mathbf{x}+a_6\mathbf{y}+a_9\right) & \mathbf{0}_{n\times 1} \\ \mathbf{0} & -1 \end{bmatrix} \tag{6.45}$$

where n is the number of points.

Meanwhile, the vector $\mathbf{w}^0$ is given as

$$\mathbf{w}^0 = \begin{bmatrix} f\left(\xi^0, \mathbf{x}^0, \mathbf{y}^0\right) \\ a_{10}^0 - \left(\hat{a}_{10}\right)_{eig} \end{bmatrix} - \mathbf{J}_2^0 \mathbf{v}_{total}^0, \tag{6.46}$$

The iterative process is implemented using Eqs (5.45) and (5.46) and updating both Jacobian matrices as well as the vector $\mathbf{w}^i$ until convergence.

The results of the parameters of one generation are exemplarily given as

$$\hat{\xi}_{eig} = \begin{bmatrix} -0.063420 \\ -0.110654 \\ -0.991682 \\ 0.001049 \\ -0.000001 \\ 0.017248 \\ -0.001488 \\ 0.002845 \\ 0.010736 \\ 1.015013 \end{bmatrix} \quad \text{and} \quad \hat{\xi} = \begin{bmatrix} -0.063576 \\ -0.110377 \\ -0.995654 \\ 0.000294 \\ -0.001208 \\ 0.014478 \\ -0.002646 \\ 0.002515 \\ 0.011647 \\ 1.015013 \end{bmatrix}, \tag{6.47}$$

The first three parameters represent the parameters of the original function of the ellipsoid. The fourth until the sixth parameters indicate that the rotation is not significant since the estimates are close to 0. The same case is also for the parameters a_7, a_8, a_9, which means that there are almost no translations of the center of ellipsoid. If the parameters are estimated, one can use the method proposed in Kutterer and Schön (1999) to confirm the form.

The generation of the ellipsoid is repeated 20 times (the function $\frac{x^2}{4^2}+\frac{y^2}{3^2}+\frac{z^2}{1^2}=1$ is fixed and only the noises are generated according to the normal distribution with standard deviation 0.05). The sums of squared residual of the eigenvalue decomposition method and our iterative method are computed and presented in Fig 6.2.

Our iterative method gives the optimal solution represented by minimal sum of the squared residuals, though the difference to the method based on the eigenvalue decomposition is small. Meanwhile, the method based on the eigenvalue cannot solve the problem without any limitation of vcm. Furthermore, the estimated sum of squared residuals is smaller than the sum of squared generated errors, because we estimate the general form of the quadratic equation instead of the exact ellipsoid (i.e. more parameters are used).

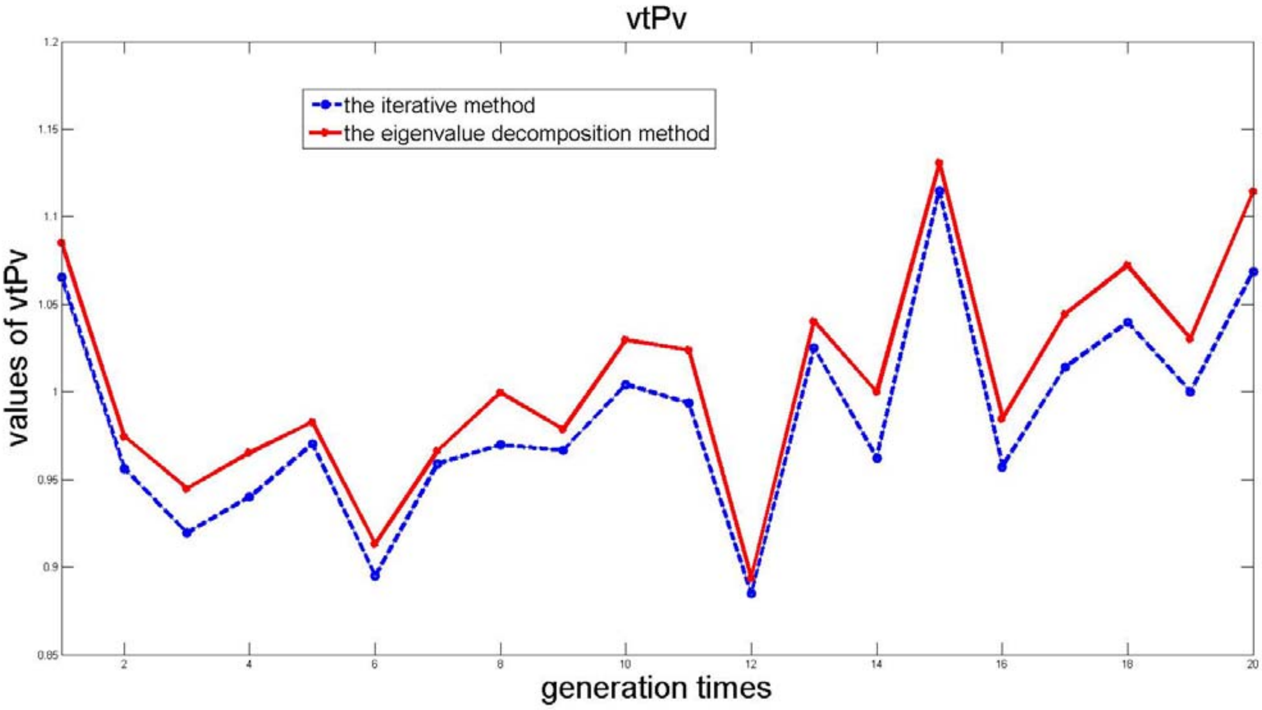


Figure 6.2 The sum of squared residuals obtained by the iterative method and eigenvalue method for 20 time generations of the quadratic form estimation

6.4 TLS solution for geodetic networks

In order to establish the datum of a local geodetic network, one approach may be the geodetic network adjustment with the weak datum, which takes the stochastic prior information of the coordinates of the connection stations into account. In this case, such connection stations cannot not be regarded as fixed (error free) station. The corresponding linear model has been described in, e.g., Niemeier (2002)

$$E\left(\begin{bmatrix}\mathbf{l}\\ \mathbf{l}_C\end{bmatrix}\right)=\begin{bmatrix}\mathbf{A}_N & \mathbf{A}_C\\ \mathbf{0} & \mathbf{I}_C\end{bmatrix}\begin{bmatrix}\boldsymbol{\xi}_N\\ \boldsymbol{\xi}_C\end{bmatrix}, \tag{6.48}$$

where $\boldsymbol{\xi}_N$ and $\boldsymbol{\xi}_C$ are the regular parameters and the parameters for the coordinates of the stations carrying the stochastic information, respectively. $\mathbf{A}_N$ and $\mathbf{A}_C$ are the design matrix for the linear relationship between the observations and regular parameters $\boldsymbol{\xi}_N$ or between the observations and parameters $\boldsymbol{\xi}_C$. $\mathbf{l}_C$ contain coordinates of the connection stations possessing the stochastic information, which may be obtained by a pre-processing.

Recently, Reinking (2008) proposed the use of TLS technique in order to treat the problem of the stationing through the stochastic datum points (connection stations). The approximated coordinates of the 3 connection stations in the weak datum are given in the following table, where the measured distances to the point N required to be determined are also presented

Points	X	Y	Distance to point N
1	528.76	440.27	85.350
2	697.31	518.85	145.503
3	650.23	288.64	124.397

Table 6.14 The coordinates of 3 known network points and corresponding distances measured to the point required to be stationed

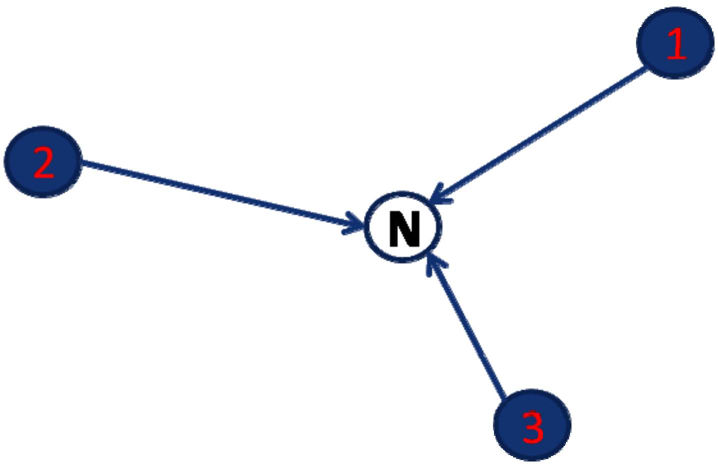


Figure 6.3 The geodetic networks

The situation representing the coordinates and distance in Table 6.14 is illustrated in the above figure. The coordinates and the measured distance are stochastic in this case. In this contribution, the stochastic parameters are also taken into account within the EIV model. The prior stochastic information of the coordinates of the point N is here assumed as a normal distribution with $\boldsymbol{\xi} = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} \sim \left(\boldsymbol{\xi}_{sto}, \boldsymbol{\Sigma}_{\xi\xi} \right) = \left(\begin{bmatrix} 606.5 \\ 405.1 \end{bmatrix}, \sigma_0^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)$, where $[\xi_1 \quad \xi_2]$ are the coordinates of the point to be determined.

In order to solve the problem, the nonlinear functional relationship for each distance measured is described as

$$f_i = \sqrt{\left(E(y_i) - \xi_2\right)^2 + \left(E(x_i) - \xi_1\right)^2} - E(s_i) = 0\,, \tag{6.49}$$

where x_i and y_i are the 2D coordinates of the ith connection point. s_i is the measured distance between the point to be stationed and the ith control point.

The observations and their vcm are organized in vectors as follows

$$\mathbf{l}^T = \begin{bmatrix} \mathbf{x}^T & \mathbf{y}^T & \mathbf{s}^T \end{bmatrix}$$

$$D(\mathbf{l}) = \sigma_0^2 \mathbf{Q}_{\mathbf{ll}} = \sigma_0^2 \begin{bmatrix} \mathbf{I}_3 & 0.7\cdot\mathbf{I}_3 & \mathbf{0} \\ 0.7\cdot\mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \tag{6.50}$$

According to the functional relationship (6.49) the partial derivative w.r.t. the parameter vector and observation vector is obtained

$$\mathbf{A}^0 = \frac{\partial \mathbf{f}}{\partial \boldsymbol{\xi}^T}\Big|\boldsymbol{\xi}^0, \mathbf{l}^0 = \begin{bmatrix} -\left(x_1^0 - \xi_1^0\right)/d_1^0 & -\left(y_1^0 - \xi_2^0\right)/d_1^0 \\ -\left(x_2^0 - \xi_1^0\right)/d_2^0 & -\left(y_2^0 - \xi_2^0\right)/d_2^0 \\ -\left(x_3^0 - \xi_1^0\right)/d_3^0 & -\left(y_3^0 - \xi_2^0\right)/d_3^0 \end{bmatrix}$$

$$\mathbf{B}^0 = \frac{\partial \mathbf{f}}{\partial \mathbf{l}^T}\Big|\boldsymbol{\xi}^0, \mathbf{l}^0 = \begin{bmatrix} \left(x_1^0 - \xi_1^0\right)/d_1^0 & 0 & 0 & \left(y_1^0 - \xi_2^0\right)/d_1^0 & 0 & 0 & \\ 0 & \left(x_2^0 - \xi_1^0\right)/d_2^0 & 0 & 0 & \left(y_2^0 - \xi_2^0\right)/d_2^0 & 0 & -\mathbf{I}_3 \\ 0 & 0 & \left(x_3^0 - \xi_1^0\right)/d_3^0 & 0 & 0 & \left(y_3^0 - \xi_2^0\right)/d_3^0 & \end{bmatrix} \tag{6.51}$$

where $d_i^0 = \sqrt{\left(y_i^0 - \xi_2^0\right)^2 + \left(x_i^0 - \xi_1^0\right)^2}$.

Furthermore, the inconsistency vector for the first iteration can be written as

$$\mathbf{w}^0 = \begin{bmatrix} \sqrt{\left(y_1^0 - \xi_2^0\right)^2 + \left(x_1^0 - \xi_1^0\right)^2} - s_1^0 \\ \sqrt{\left(y_2^0 - \xi_2^0\right)^2 + \left(x_2^0 - \xi_1^0\right)^2} - s_2^0 \\ \sqrt{\left(y_3^0 - \xi_2^0\right)^2 + \left(x_3^0 - \xi_1^0\right)^2} - s_3^0 \end{bmatrix}, \tag{6.52}$$

According to Eq (5.49) the solution of the first iteration is obtained as follows

$$\hat{\boldsymbol{\xi}}^1 = \left(\left(\mathbf{A}^0\right)^T \left(\mathbf{B}^0\mathbf{Q}_{\mathbf{ll}}\left(\mathbf{B}^0\right)^T\right)^{-1} \mathbf{A}^0 + \boldsymbol{\Sigma}_{\xi\xi}^{-1}\right)^{-1} \left(-\left(\mathbf{A}^0\right)^T \left(\mathbf{B}^0\mathbf{Q}_{\mathbf{ll}}\left(\mathbf{B}^0\right)^T\right)^{-1} \mathbf{w} + \left(\mathbf{A}^0\right)^T \left(\mathbf{B}^0\mathbf{Q}_{\mathbf{ll}}\left(\mathbf{B}^0\right)^T\right)^{-1} \mathbf{A}^0\boldsymbol{\xi}^0 + \boldsymbol{\Sigma}_{\xi\xi}^{-1}\boldsymbol{\xi}_{sto}\right) \tag{6.53}$$

After some iterations (updating both Jacobian matrices and the inconsistency vector by the estimated parameters and residuals) the result converges at the solution presented as follows

	X	**Y**
Mean of stochastic parameters (prior known)	606.5000	405.1000
Estimate of parameters without stochastic prior information	606.5417	405.1197
Estimate of parameters with stochastic prior information	606.5161	405.1056

Table 6.15 The result of estimated coordinates of new point without or with stochastic prior information

In Tab 6.15 the first row represents the mean given by the prior stochastic information about parameters. In the second row the parameters estimated by the weighted TLS solution are given, which are identical to the solution presented by Reinking (2008). Considering the stochastic parameters within the EIV model the solution (6.53) is calculated and exhibited in the last row of Tab 6.15. The solution could be the weighted mean of the expectation of the given stochastic parameter and the weighted TLS solution. Hence, the estimates of the parameters of the weighted TLS solution considering the stochastic parameters lie between both other values.

6.5 Summary of applications

By means of the theoretical development some geodetic applications are demonstrated. The example of the orthogonal regression indicates that our solutions provide identical results with other present methods in the not fully populated weights case. They present their advantages in the aspect of the convergence behavior as well as the limitation of stochastic information. Furthermore, it is exhibited that the WTLS solution can be significantly different to the TLS solution; the difference is even larger than the difference between the LS and TLS solution. In the second example the 3D similarity transformation problem is rigorously solved in the different weight cases. Our method results the identical estimates to the existing method in the particular weight case and can be used with any vcm (e.g., Toeplitz structure). The estimation is not only implemented in the transformation from the source coordinate system to the target system but also from the target system to the source system. By comparison of results it is presented that the product of both scales is exactly 1 and the sums of the squared residuals are identical. In addition, the Gauss-Newton type estimation can also be used if the distance between two points is fixed in the transformation problem. In the next example parameters of quadratic form are estimated in the condition that the quadratic form is exact known or unknown. The last application the WTLS technique under the consideration of the stochastic parameter is used in geodetic networks. In this example the datum points and the point to be determined are stochastic simultaneously. The result indicates that our solution is the weighted mean of the WTLS solution and the mean given by the prior stochastic information.

7 Conclusions and outlook

In this thesis, various solutions of the WTLS problem are intensively investigated. From the methodological aspect, a class of the traditional geodetic methods, for example Lagrange multiplier and the non-linear GHM, is applied to solve the fully weighted TLS problem with and without linearization, respectively. Another class of methods is from the community of numerical analysis, such as the Newton approach and the non-linear unconstrained LS method. From the algorithms points of view, the Gauss-Newton algorithm and Newton algorithm are presented. They can overcome the computational difficulty of the WTLS problem if they are modified. The necessary and sufficiency conditions are obtained by the first and second derivative of the target function w.r.t. the parameters, which guarantee the convergence of the algorithm to the (local) minimum. Therefore, the general solutions for the WTLS problem can be put forward based on the complete and reliable analysis.

Furthermore, some extensions of the WTLS problem are presented. First, the WTLS problem with fixing columns, which cannot be solved by the matrix decomposition and the centering matrix, is introduced. The solution is given by the parameter elimination based on the normal equation system. The idempotent matrix for the parameter elimination is identical with the centering matrix if the observations are equally weighted. The WTLS problem with fixing elements, is also presented and solved by the iterative GHM. Subsequently, the linear constrained WTLS problem is demonstrated, and is solved by Lagrange multipliers. In order to adjust the geodetic transformation model, the WTLS solution for the parameters formulated in matrix-wise way is given. In addition, the integrated Gauss-Newton solution of the general non-linear EIV model is proposed.

Therefore, some geodetic problems are solved by the methods proposed in this thesis. The examples have shown the correctness as well as the applicability of the methods via the comparison with existing methods. In the first example the convergence behavior of different algorithms is shown when the parameters of the lines are estimated. Moreover, the WTLS solution is used in the similarity transformation, quadratic form analysis and geodetic networks.

The presented results give a good basis for further work; some studies can be systematically investigated in the future:

- From a mathematical aspect the TLS problem with inequality constraints may be investigated on the basis of the first and second derivatives given analytically in the thesis.
- Solving the problem with the Bayesian statistic will be investigated in the future, based on which the solution may be put forward under the assumptions that other probability distributions of observations instead of the traditional normal distribution are available.
- It may be also a problem that outliers occur in observations within the EIV model. It may be treated by using the generalized maximum likelihood objective function $\rho\left(\mathbf{v}^2\right)$ instead of the function $\mathbf{v}^2$.
- In order to give a rigorous estimation about the variance factor of the unit weight, we refer to repro-BIQUE in Schaffrin (1983) and BIQUUE (Best Invariant Quadratic Uniformly Unbiased Estimate) in Grafarend and Schaffrin (1993 p. 315-319). As stated above, the rigorous estimation of the vcm of the estimated parameters as well as variance components will be also the research topic in further work.

References

Adcock, R. (1877) Note on the method of least squares. Analyst 4: 183-184.

Akyilmaz, O. (2007) Total least squares solution of coordinate transformation. Survey Review, 39 (303): 68-80.

Amiri-simkooei, A. (2007) Least-squares variance component estimation: theory and GPS applications. Ph.D.thesis, Delft University of Technology, Publication on Geodesy 64, Netherlands Geodetic Commission, Delft.

Bleich, P. & Illner, M (1989) Strenge Lösung der räumlichen Koordinatentransformation durch iterative Berechnung. AVN 96 (4): 133-144

Björck, A., Heggernes, P., & Matstoms, P. (2000) Methods for large scale total least-squares problems. SIAM Journal on Matrix Analysis and Applications 22(2): 413-429

Csanyi May, N. (2008) A rigorous approach to comprehensive performance analysis of state-of.the-art airborne mobile mapping systems. Ph.D.thesis, School of the Ohio State University.

De Moor, B. (1993) Structured total least squares and L2 approximation problems. Linear Algebra and its Applications.188-189: 163-207.

Drixler, E. (1993) Analyse der Form und Lage von Objekten im Raum (Bd. Reihe C 409). München: Deutsche Geodätische Kommission.

Eckart, C. & Young, G. (1936) The approximation of one matrix by another of lower rank. Psychometrika, 1 (3): 211-218

Eling, D. (2010) Terrestrisches Laserscanning für die Bauwerksüberwachung. Dissertation, Leibniz Universität Hannover.

Felus, F. (2004) Application of Total Least Squares for Spatial Point Process Analysis. Journal of Surveying Engineering, 130 (3):126-133.

Felus, F., & Burtch, R. (2009) On symmetrical three-dimensional datum conversion. *GPS Solutions , 13* (1):65-74.

Felus, Y., & Schaffrin, B. (2005) Performing similarity transformations using the error-in-variable model. ASPRS Annual conference 'Geospatial Goes Global: From Your Neighborhood to the Whole Planet'.

Frandsen, P., Jonasson, K., Nielsen, H., & Tingleff, O. (2004) Unconstrained optimization. Lecture note. Denmark: Technical University of Denmark.

Gleser, L. (1981) Estimation in a multivariate 'errors in variables' regression model: large sample results. Annals of Statistics 9 (1): 24-44.

Golub, G., & Van Loan, C. (1980) An analysis of the Total least –squares problem. SIAM Journal on Numerical Analysis, 17 (6): 883-893.

Grafarend, E., & Awange, J. (2003) Nonlinear analysis of the three-dimensional datum transformation. Journal of Geodesy , 77 (1-2): 66-76.

Grafarend, E., & Schaffrin, B. (1993) Ausgleichungsrechnung in linearen Modellen. Mannheim: BI-Wissenschaftsverlag.

Grander, W., Golub, GH. & Strebel, R. (1994) Least-squares fitting of circles and ellipses, BIT Numerical Mathematics 34: 558-578.

Harville, D. (1977) Maximum Likelihood approaches to variance component estimation and to related problems. J. Amer. Statist. Assoc. , 358, S. 320-338.

Hesse, C., & Kutterer, H. (2006) Automated Form Recognition of Laser Scanned Deformable Objects. In: Sanso, F., Gil, A. (Eds.) (2007): Geodetic Deformation Monitoring: From Geophysical to Engineering Roles. IAG Symposia, 131, Springer.

Hwang, C., Wang, C., & Lee, L. (2002) Adjustment of relative gravity measurements using weighted and datum free constraints. Comput.Geosci., 28: 1005-1015.

Koch, K. (1999) Parameter estimation and hypothesis testing in linear models. 2nd edn. Springer.

Kupferer, S. (2005) Anwendung der Total-Least-Squares-Technik bei geodätischen Problemstellungen. Dissertation, Schriftreihe des Studiengangs Geodäsie der Universität Karlsruhe.

Kutterer, H., & Schön, S. (1999) Statistische Analyse quadratischer Formen der Determinantenansatz. AVN 10: 322-330

Kwon, JH., Lee, J., Schaffrin, B., Choi, YS., & Lee, I. (2009) New Affine Transformation Parameters for the Horizontal Network of Seoul/Korea by Multivariate TLS-Adjustment. Survey review.

Lenzmann, L., & Lenzmann, E. (2004) Rigorous adjustment of the nonlinear Gauss Helmert Model. AVN. 111: 68-73.

Lenzmann, L., & Lenzmann, E. (2007) Zur Lösung des nichtlinearen Gauss-Markov-Modells. ZfV , 132: 108-120.

Levenberg, M. (1944) A method for the solution of certain problems in Least Squares. Quart. Appl. Math. 2: 164-168

Mann, ME., & Emanuel, KA. (2006) Atlantic hurricane trends linked to climate change. EOS 87-24:233-241

Markovsky, I., & Van Huffel, S. (2005) High-performance numerical algorithms and software for structured total least squares. J. Comput. Appl. Math, 180 (2), S. 311-331.

Markovsky, I., & Van Huffel, S. (2007) Overview of total least-squares methods. Signal Processing, 87 (10): 2283-2302.

Markovsky, I., Rastello, M., Premoli, A., Kukush, A., & van Huffel, S. (2006) The element-wise weighted total least-squares problem. Computational Statistics & Data Analysis, 50: 181-209.

Marquardt, D. (1963) An algorithm for Least Squares Estimation on Nonlinear Parameters. SIAM J. Appl. MATH. 11: 431-441

Neitzel, F., & Petrovic, S. (2008) Total Least Squares (TLS) im Kontext der Ausgleichung nach kleinsten Quadraten am Beispiel der Ausgleichenden Geraden. ZfV 133: 141-148.

Neitzel, F. (2010) Generalization of total least-squares on example of unweighted and weighted 2D similarity transformation. Journal of Geodesy 84:751-762

Neri, F. Saitta, G. & Chiofalo S (1989) An accurate and straightforward approach to line regression analysis of error-affected experimental data. J Phys Ser E: Sci Instr 22:215:217

Pope, A. (1972) Some pitfalls to be avoided in the iterative adjustment of nonlinear problems. Proc 38th Ann Meet Am Soc Phot, Waschington, DC , pp 449-473.

Pope, A. (1974) Two approaches to nonlinear least squares adjustments. Can Surv 28(5): 663-669

Reinking, J. (2008) Total least squares? Zeitschrift für Vermessungswesen , 133, S. 384-389.

Schaffrin, B. (1983) Varianz-Kovarianz-Komponenten-Schätzung bei der Ausgleichung heterogener Wiederholungsmessungen. München: Deutsche Geodätische Kommission 282.

Schaffrin, B. (1997) Reliability measures for correlated observations. Journal of Surveying Engineering, 123 (3):126-137.

Schaffrin, B., & Felus, Y. (2005) On total Least-Squares adjustment with constraints. Sanso F.(Ed.), A Window on the future of Geodesy. International Association of Geodesy Symposia, 128: 175-180.

Schaffrin, B. (2006) A note on Constrained Total Least-Squares estimation. Linear Algebra and its Applications, 417: 245-258.

Schaffrin, B. (2007) Connecting the Dots: The Straight-Line Case revisited. ZfV 132: 385-394

Schaffrin, B., & Felus, Y. (2008) On the multivariate total least-squares approach to empirical co-ordinate transformations. Three algorithms. Journal of Geodesy, 82: 373-383.

Schaffrin, B., & Wieser, A. (2008) On weighted total least-squares adjustment for linear regression. Journal of Geodesy, 82: 415-421.

Schaffrin, B., & Felus, Y. (2009) An algorithmic approach to the total least-squares problem with linear and quadratic constraints. Studia Geophysica et Geodaetica, 53 (1): 1-16.

Schaffrin, B., & Wieser, A. (2009) Empirical Affine Reference Frame Transformations by Weighted Multivariate TLS Adjustment. In: Drewes H (Ed) Geodetic Reference Frames, IAG Symposium Munich, Germany , 9-14 October 2006; Series: International Association of Geodesy Symposia , Vol. 134, Springer .

Schaffrin, B., Neitzel, F., & Uzum, S. (2009) Empirical similarity transformation via TLS-adjustment: exact solution vs. Cadzow’s approximation. International Geomatics Forum, 28-30 May 2009, Qingdao, People’s Republic of China

Schaffrin, B., & Snow, K. (2010) Total Least-Squares regularization of Tykhonov type and an ancient racetrack in Corinth. Linear Algebra and its Applications, 432: 2061-2076.

Schuermans, M., Markovsky, I., & Van Huffel, S. (2007) An adapted version of the element-wise weighted total least squares method for applications in chemometrics. Chemometrics and intelligent laboratory systems, 85: 40-46.

Shan, J. (1989) A fast recursive method for repeated computation of the reliability matrix QvvP. Photogrammetria, 43: 337-346.

Shen, Y., Chen, Y., & Zheng, D. (2006) A quanternion-based geodetic datum transformation. Journal of Geodesy, 80 (5):. 233-239.

Shen, Y., Li, BF., & Chen, Y. (2010) An iterative solution of weighted total least-squares adjustment. Journal of Geodesy, 85: 229-238

Teunissen, PJG. (1985) The geometry of geodetic inverse linear mapping and non-linear adjustment. Hetherlands Geodetic Commission, Publications on Geodesy 8(1), Delft, 177pp.

Teunissen, PJG. (1988) The non-linear 2D symmetric Helmert transformation: an exact non-linear least-squares solution. Manuscripta Geodetica, 62 (1):1-16.

Teunissen, PJG. (1990) Nonlinear least squares. Manuscr Geod 15:137-150

Teunissen, PJG., & Amiri-Simkooei, A. (2008) Least-squares variance component estimation. Journal of Geodesy , 82: 65-82.

Van Huffel, S., & Vandewalle, J. (1991) The Total least –squares problem. Computational aspects and analysis, Society for Industrial and Applied Mathematics, Philadelphia.

Vennebusch M, Nothnagel, A., & Kutterer, H. (2009) Singular value decomposition and cluster analysis as regression diagnostics tools for geodetic applications. Journal of Geodesy, 83: 877-891 .

Niemeier, W. (2002) Ausgleichungsrechnung. Berlin New York: Walter de Gruyter.

Wicki, F. (1998) Robuste Schätzverfahren für die Parameterschätzung in geodätischen Netzen. Dissertation ETH Nr. 12894.

Williams, S. (2003) The effect of coloured noise on the uncertainties of rates estimated from geodetic time series. Journal of Geodesy, 76: 483-494.

Wolf, H. (1968) Ausgleichungsrechnung nach der Methode der kleinsten Quadrat. Dümmler

Wolf, H. (1978) Das geodätische Gauss-Helmert-Modell und seine Eigenschaften. ZfV 103:41-43

Xu, G. (2007) GPS: theory, algorithms, and applications. Berlin: Springer.

Xu. PL., Liu, YM., Shen, YZ. & Fukuda Y. (2007) Estimability analysis of variance and covariance components. Journal of Geodesy, 81: 593-602.

Yang, Y. (1999) Robust estimation of geodetic datum transformation. Journal of Geodesy, 73: 268-274.

Yang, Y., Gao, W., & Zhang, X. (2010) Robust Kalman filtering with constraints: a case study for integrated navigation. Journal of Geodesy, 84: 373-381

Yeredor, A. (2006) On the role of constraints in system identification. Fourth International Workshop on Total Leat-Squares and Errors-in-Variables Modeling (S. 46-48). Leuven/Belgium: Van Huffel S. and Markovsky I.

List of Figures and Tables

Figures

Tables

Appendix

The second derivative of the WTLS objective function w.r.t. the parameter vector

According to the first derivative (4.45) the second derivative can be organized in two parts as follows

$$\frac{\partial^2 (\mathbf{y}-\mathbf{A}\xi)^T (\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)}{\partial\xi_k \partial\xi_j}$$

$$= -2\frac{\partial \mathbf{a}_k^T (\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)}{\partial\xi_j}\ldots$$

$$-2\frac{\partial (\mathbf{y}-\mathbf{A}\xi)^T (\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{Q}_k^T (\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)}{\partial\xi_j}$$

The first and second part of the equation above can be analytically formulated by

$$-2\frac{\partial \mathbf{a}_k^T (\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)}{\partial\xi_j}$$

$$= -2\mathbf{a}_k^T \frac{\partial (\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}}{\partial\xi_j}(\mathbf{y}-\mathbf{A}\xi) - 2\mathbf{a}_k^T (\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}\frac{\partial(\mathbf{y}-\mathbf{A}\xi)}{\partial\xi_j}$$

$$= 2\mathbf{a}_k^T (\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}\mathbf{Q}_j\mathbf{B}^T (\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi) + 2\mathbf{a}_k^T (\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{Q}_j^T (\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi) + 2\mathbf{a}_k^T (\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}\mathbf{a}_j$$

and

$$-2\frac{\partial (\mathbf{y}-\mathbf{A}\xi)^T (\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{Q}_k^T (\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)}{\partial\xi_j}$$

$$= -2\frac{\partial (\mathbf{y}-\mathbf{A}\xi)^T}{\partial\xi_j}(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{Q}_k^T (\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)\ldots$$

$$-2(\mathbf{y}-\mathbf{A}\xi)^T (\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{Q}_k^T (\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}\frac{\partial(\mathbf{y}-\mathbf{A}\xi)}{\partial\xi_j}\ldots$$

$$-2(\mathbf{y}-\mathbf{A}\xi)^T \frac{\partial (\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}}{\partial\xi_j}\mathbf{B}\mathbf{Q}_k^T (\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)\ldots$$

$$-2(\mathbf{y}-\mathbf{A}\xi)^T (\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{Q}_k^T \frac{\partial (\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}}{\partial\xi_j}(\mathbf{y}-\mathbf{A}\xi)\ldots$$

$$-2(\mathbf{y}-\mathbf{A}\xi)^T (\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}\frac{\partial \mathbf{B}\mathbf{Q}_k^T}{\partial\xi_j}(\mathbf{B}\mathbf{Q}_{ll}\mathbf{B}^T)^{-1}(\mathbf{y}-\mathbf{A}\xi)$$

$$
\begin{aligned}
&= 2\mathbf{a}_j^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{BQ}_k^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi\right)\\
&\quad+2\left(\mathbf{y}-\mathbf{A}\xi\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{BQ}_k^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{a}_j\ldots\\
&\quad+2\left(\mathbf{y}-\mathbf{A}\xi\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{Q}_j\mathbf{B}^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{BQ}_k^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi\right)\ldots\\
&\quad+2\left(\mathbf{y}-\mathbf{A}\xi\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{BQ}_j^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{BQ}_k^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi\right)\ldots\\
&\quad+2\left(\mathbf{y}-\mathbf{A}\xi\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{BQ}_k^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{Q}_j\mathbf{B}^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi\right)\ldots\\
&\quad+2\left(\mathbf{y}-\mathbf{A}\xi\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{BQ}_k^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{BQ}_j^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi\right)\ldots\\
&\quad-2\left(\mathbf{y}-\mathbf{A}\xi\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{Q}_{kj}^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi\right)\\
&= 2\left(\mathbf{y}-\mathbf{A}\xi\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{Q}_k\mathbf{B}^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{a}_j\\
&\quad+2\left(\mathbf{y}-\mathbf{A}\xi\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{BQ}_k^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{a}_j\ldots\\
&\quad+2\left(\mathbf{y}-\mathbf{A}\xi\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{Q}_k\mathbf{B}^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{BQ}_j^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi\right)\ldots\\
&\quad+2\left(\mathbf{y}-\mathbf{A}\xi\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{Q}_k\mathbf{B}^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{Q}_j\mathbf{B}^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi\right)\ldots\\
&\quad+2\left(\mathbf{y}-\mathbf{A}\xi\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{BQ}_k^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{Q}_j\mathbf{B}^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi\right)\ldots\\
&\quad+2\left(\mathbf{y}-\mathbf{A}\xi\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{BQ}_k^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{BQ}_j^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi\right)\ldots\\
&\quad-2\left(\mathbf{y}-\mathbf{A}\xi\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{Q}_{kj}^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi\right)
\end{aligned}
$$

In the last step the property introduced in Eq (4.43) and the property of the transposed scalar are used. Thus, combining the solved equations of first and second part, the Hessian matrix is given as follows

$$
\begin{aligned}
\mathbf{H} &= \frac{\partial^2\left(\mathbf{y}-\mathbf{A}\xi\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi\right)}{\partial\xi\partial\xi^T}\\
&= 2\mathbf{A}^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{A}^* + 2\mathbf{A}^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{A}^{**} + 2\mathbf{A}^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{A}\ldots\\
&+2\left(\mathbf{A}^{**}\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{A} + 2\left(\mathbf{A}^*\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{A} + 2\left(\mathbf{A}^{**}\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{A}^{**}\ldots\\
&+2\left(\mathbf{A}^{**}\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{A}^* + 2\left(\mathbf{A}^*\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{A}^* + 2\left(\mathbf{A}^*\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{A}^{**} - 2\left[\varpi_{kj}\right]\\
&= 2\left(\mathbf{A}+\mathbf{A}^*+\mathbf{A}^{**}\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\left(\mathbf{A}+\mathbf{A}^*+\mathbf{A}^{**}\right) - 2\left[\varpi_{kj}\right]
\end{aligned}
$$

with

$\varpi_{kj} = \left(\mathbf{y}-\mathbf{A}\xi\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{Q}_{kj}^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi\right) = \left(\mathbf{y}-\mathbf{A}\xi\right)^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\mathbf{Q}_{kj}\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi\right)$.
$\mathbf{A}^{**} = \left[\mathbf{BQ}_1^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi\right),\ \ldots\ ,\mathbf{BQ}_u^T\left(\mathbf{BQ}_{\text{ll}}\mathbf{B}^T\right)^{-1}\left(\mathbf{y}-\mathbf{A}\xi\right)\right]$. Note that $\mathbf{A}^{**} = \mathbf{A}^*$ unless $\mathbf{Q}_{kj}^T = \mathbf{Q}_{kj}$ for any k and j.

Curriculum vitae

Persönlichen Daten

Name: Xing Fang
Geburtsdatum: 26.11.1981
Geburtsort: Hubei, VR China
Staatsangehörigkeit: Chinese

Bildung

10/2004-07/2007	Geodäsie und Geoinformatik (Dipl.-Ing) Leibniz Universität Hannover
09/2000-06/2004	Geodäsie und Geoinformatik (Bachelor) Wuhan Universität
09/1994-07/2000	Die Fremdsprache Schule in Wuhan (Abitur)

Beruf

09/2010-07/2011	Wissenschaftlicher Mitarbeiter am Institut für Erdmessung, Leibniz Universität Hannover
05/2009-08/2010	Wissenschaftlicher Mitarbeiter am Geodätischen Institut, Leibniz Universität Hannover
01/2008-04/2009	Studentische Hilfskraft am Geodätischen Institut bzw. Institut für Erdmessung, Leibniz Universität Hannover

Acknowledgements

This thesis could not be realized without the sincere support from many people. However, it is difficult to thank all the individuals who really help me to finish the work.

In the first place, I would like to thank my supervisor, Prof H. Kutterer, for his complete scientific support. His valuable suggestion from the initial to the final enabled me to develop my thesis step by step. He knew accurately which research topic is proper for me and gave me the enough freedom to concentrate on the study.

I am also very much thankful to Prof S. Schön and Prof W.D. Schuh not only for the acceptance of the exam committee members but also for constructive comments.

Prof J. Müller and Prof S. Schön are sincerely acknowledged for providing the extra financial support during the doctoral study.

In this place I would like to thank all of my colleagues at GIH for their support and contribution to a good research environment. The author would like to specially acknowledge Dr. -Ing I. Neumann, who discussed with me intensively when I am just involved in studying the theme. The discussion on the applications with J. Wang is also acknowledged.

For the English improvement of the early versions I would like to thank my IfE colleagues M. Smyrnaios and M. Naeimi, Prof X. Huang in Wuhan University and G. Wu in the geodetic institute of the university Stuttgart.

Last but not least, I am heartily thankful to my family for their care and support.

List of abbreviations

Best Linear Unbiased Estimation (BLUE)

element wise weighted total least-squares (EW-TLS)

errors-in-variables (EIV)

Gauss Helmert model (GHM)

Gauss Markov model (GMM)

generalized total least-squares (GTLS)

least squares (LS)

linear constrained weighted Total Least Squares (LCWTLS)

observed minus computed (O-C)

right-hand side (rhs)

singular value decomposition (SVD)

structured total least-squares (STLS)

total least-squares (TLS)

weighted Total Least Squares (WTLS)

variance covariance matrix (vcm)

Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover

(Eine vollständige Liste der Wiss. Arb. ist beim Geodätischen Institut, Nienburger Str. 1, 30167 Hannover erhältlich.)

Nr. 266	HAMPE, Mark:	Integration einer multiskaligen Datenbank in eine Webservice-Architektur (Diss. 2007)
Nr. 267	ANDERS, Frauke:	Mustererkennung in Straßennetzwerken - Verfahren zur Interpretation von Vektordaten. (Diss. 2007)
Nr. 268	HESSE, Christian:	Hochauflösende kinematische Objekterfassung mit terrestrischen Laserscannern (Diss. 2007)
Nr. 269	NEUNER, Hans-Berndt:	Zur Modellierung und Analyse instationärer Deformationsprozesse (Diss. 2007)
Nr. 270	WENDT, Axel:	Objektraumbasierte simultane multisensorale Orientierung (Diss. 2007)
Nr. 271	DILßNER, Florian:	Zum Einfluss des Antennenumfeldes auf die hochpräzise GNSS-Positionsbestimmung (Diss. 2007)
Nr. 272	BUTENUTH, Matthias:	Network Snakes (Diss. 2008)
Nr. 273	RAVANBAKHSH, Mehdi:	Road Junction Extraction from High Resolution Aerial Images Assisted by Topographic Database Information (Diss. 2008)
Nr. 274	BRZANK, Alexander:	Bestimmung Digitaler Geländemodelle in Wattgebieten aus Laserscannerdaten (Diss. 2008)
Nr. 275	SCHMIDT, Ralph:	Automatische Bestimmung von Verknüpfungspunkten für HRSC-Bilder der Mars Express-Mission (Diss. 2008)
Nr. 276	HAUNERT, Jan-Henrik:	Aggregation in Map Generalization by Combinatorial Optimization (Diss. 2008)
Nr. 277	NEUMANN, Ingo:	Zur Modellierung eines erweiterten Unsicherheitshaushaltes in Parameterschätzung und Hypothesentests (Diss. 2009)
Nr. 278	WEITKAMP, Alexandra:	Brachflächenrevitalisierung im Rahmen der Flächenkreislaufwirtschaft (Diss. 2009)
Nr. 279	HEUWOLD, Janet:	Automatische auflösungsabhängige Anpassung von Bildanalyse-Objektmodellen am Beispiel Straßen (Diss. 2009)
Nr. 280	KOCH, Christian:	Extraction of Mercury's tidal signal and libration amplitude from synthetic laser altimeter data sets (Diss. 2009)
Nr. 281	GITLEIN, Olga:	Absolutgravimetrische Bestimmung der Fennoskandischen Landhebung mit dem FG5-220 (Diss. 2009)
Nr. 282	ELING, Dirk:	Terrestrisches Laserscanning für die Bauwerksüberwachung (Diss. 2009)
Nr. 283	DOLD, Christoph:	Ebenenbasierte Verfahren für die automatische Registrierung terrestrischer Laserscans (Diss. 2010)
Nr. 284	RIPPERDA, Nora:	Rekonstruktion von Fassadenstrukturen mittels formaler Grammatiken und Reversible Jump Markov Chain Monte Carlo Sampling (Diss. 2010)
Nr. 285	KATTERFELD Christiane:	Interoperables Geodaten-basiertes E-Learning (Diss. 2010)
Nr. 286	JARECKI, Focke:	Spurkreuzungspunktdifferenzen zur Validierung satellitengradiometrischer Messungen (Diss. 2010)
Nr. 287	SCHRAMM, Matthias:	Unüberwachte lineare spektrale Entmischung mit Anwendungen für Baumdichteabschätzungen in semiariden Regionen Afrikas (Diss. 2010)
Nr. 288	ALSAYEL, Mohammed Y.:	Towards Standardizing International Boundary Determination and Quality Consequences on surveying and Mapping (Diss. 2010)
Nr. 289	TAVAKKOLI S., Seyed Mohammad:	Multi-Temporal Classification of Crops Using ENVISAT ASAR Data (Diss. 2011)
Nr. 290	VENNEGEERTS, Harald:	Objektraumgestützte kinematische Georeferenzierung für Mobile-Mapping-Systeme (Diss. 2011)
Nr. 291	BULATOV, Dimitri:	Textured 3D reconstruction of urban terrain from UAV-borne video sequences (Diss. 2011)
Nr. 292	GROTE, Anne:	Automatic road network extraction in suburban areas from aerial images (Diss. 2011)
Nr. 293	ALOBEID, Abdalla:	Assessment of Matching Algorithms for Urban DSM Generation from Very High Resolution Satellite Stereo Images (Diss. 2011)
Nr. 294	FANG, Xing:	Weighted Total Least Squares Solutions for Applications in Geodesy (Diss. 2011)

Die Arbeiten werden im Rahmen des wissenschaftlichen Schriftenaustausches verteilt und sind nicht im Buchhandel erhältlich. Der Erwerb ist zu einem Stückpreis von € 25,-- bei den herausgebenden Instituten möglich.