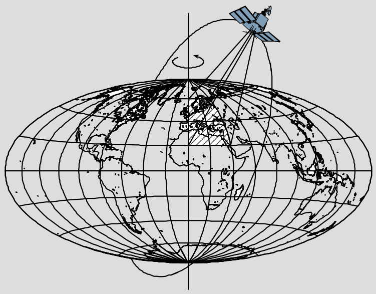
Topics in Total Least-Squares Adjustment within the Errors-In-Variables Model: Singular Cofactor Matrices and Prior Information

by

Kyle Snow



Report No. 502

Geodetic Science

The Ohio State University
Columbus, Ohio 43210

December 2012

Topics in Total Least-Squares Adjustment within the Errors-In-Variables Model: Singular Cofactor Matrices and Prior Information

by
Kyle Snow

Report No. 502
Geodetic Science
Ohio State University
Columbus, Ohio 43210

December 2012

This document was typeset by the author with $\LaTeX\,2_\varepsilon$.

Abstract

This dissertation is about *total least-squares* (TLS) adjustments within the *errors-in-variables* (EIV) model. In particular, it deals with symmetric positive-(semi)definite cofactor matrices that are otherwise quite arbitrary, including the case of cross-correlation between cofactor matrices for the observation vector and the coefficient matrix and also the case of *singular cofactor matrices*. The former case has been addressed already in a recent dissertation by Fang [2011], whereas the latter case has not been treated until very recently in a presentation by Schaffrin et al. [2012b], which was developed in conjunction with this dissertation. The second primary contribution of this work is the introduction of *prior information* on the parameters to the EIV model, thereby resulting in an errors-in-variables with random effects model (*EIV-REM*) [Snow and Schaffrin, 2012]. The (total) least-squares predictor within this model is herein called *weighted total least-squares collocation* (WTLSC), which was introduced just a few years ago by Schaffrin [2009] as TLSC for the case of independent and identically distributed (iid) data. Here the restriction of iid data is removed.

The EIV models treated in this work are presented in detail, and thorough derivations are given for various TLS estimators and predictors within these models. Algorithms for their use are also presented. In order to demonstrate the usefulness of the presented algorithms, basic geodetic problems in 2-D line-fitting and 2-D similarity transformations are solved numerically. The new extensions to the EIV model presented here will allow the model to be used by both researchers and practitioners to solve a wider range of problems than was hitherto feasible.

In addition, the Gauss-Helmert model (GHM) is reviewed, including details showing how to update the model properly during iteration in order to avoid certain pitfalls pointed out by Pope [1972]. After this, some connections between the GHM and the EIV model are explored.

Though the dissertation is written with a certain bent towards geodetic science, it is hoped that the work will be of benefit to those researching and working in other branches of applied science as well. Likewise, an important motivation of this work is to highlight the classical EIV model, and its recent extensions, within the geodetic science community, as it seems to have received little attention in this community until a few years ago when Professor Burkhard Schaffrin began publishing papers on the topic in both geodetic and applied mathematics publications.

Preface

This report is substantially the same as a dissertation that was prepared for and submitted to the Graduate School of The Ohio State University for the PhD degree. Except for the omission of some pages from the front matter, a different acknowledgment page, and a change from double-space to single-space format, this report is identical to the dissertation, which contains 15 pages with Roman numerals and 116 pages with Arabic numerals.

Acknowledgments

I thank Professor Frank Neitzel, of the Berlin University of Technology, for providing the data for the 2-D transformation problem solved in Chapter 6 of this report and for checking my solution to that problem.

I am indebted to my adviser, Professor Burkhard Schaffrin, not only for his advice and guidance but also for his critical review of my dissertation, which resulted in many improvements to the original manuscript. Any remaining errors or deficiencies are solely the responsibility of the author.

Contents

Chapter 1: Introduction

1.1 Contributions of this dissertation

This dissertation is about *total least-squares* (TLS) adjustments within the *errors-in-variables* (EIV) model. In particular, it deals with symmetric positive-(semi)definite cofactor matrices that are otherwise quite arbitrary, including the case of cross-correlation between cofactor matrices for the observation vector and the coefficient matrix and also the case of *singular cofactor matrices*. The former case has been addressed already in a recent dissertation by Fang [2011], whereas the latter case has not been treated until very recently in a presentation by Schaffrin et al. [2012b], which was developed in conjunction with this dissertation. The second primary contribution of this work is the introduction of *prior information* on the parameters to the EIV model, thereby resulting in an errors-in-variables with random effects model (*EIV-REM*) [Snow and Schaffrin, 2012]. The (total) least-squares predictor within this model is herein called *weighted total least-squares collocation* (WTLSC), which was introduced just a few years ago by Schaffrin [2009] as TLSC for the case of independent and identically distributed (iid) data. Here the restriction of iid data is removed.

The EIV models treated in this work are presented in detail, and thorough derivations are given for various TLS estimators and predictors within these models. Algorithms for their use are also presented. In order to demonstrate the usefulness of the presented algorithms, basic geodetic problems in 2-D line-fitting and 2-D similarity transformations are solved numerically. The new extensions to the EIV model presented here will allow the model to be used by both researchers and practitioners to solve a wider range of problems than was hitherto feasible.

In addition, the Gauss-Helmert model (GHM) is reviewed, including details showing how to update the model properly during iteration in order to avoid certain pitfalls pointed out by Pope [1972]. After this, some connections between the GHM and the EIV model are explored.

Though the dissertation is written with a certain bent towards geodetic science, it is hoped that the work will be of benefit to those researching and working in other branches of applied science as well. Likewise, an important motivation of this work is to highlight the classical EIV model, and its recent extensions, within the geodetic science community, as it seems to have received little attention in this community until a few years ago when Professor Burkhard Schaffrin began publishing papers on the topic in both geodetic and applied mathematics publications.

1.2 A brief introduction to the EIV model and TLS adjustment

Before introducing the formal EIV model and the principle of TLS, a brief selection of historical developments in linear algebra, least-squares theory, and statistical estimation is given, especially as related to the field of geodetic science.

1.2.1 Some historic connections of geodesy to linear algebra, least-squares, and statistical estimation theory

The importance of linear algebra, least-squares minimization, and statistical estimation in the field of geodetic science (or geodesy) cannot be overstated. The connection between geodetic science to least-squares adjustment dates back to Gauss, who discovered the least-squares principle in 1794 and used it for many years before first publishing the theory in Gauss [1809] (translated into English by Davis [1857]), followed by his famous 1823 work (Gauss [1823], translated into English by Stewart [1995]), where the important statistical connection to the minimum-variance principle was made. It is perhaps less well-known that, during the period between discovery and the later publication, least-squares adjustment found a home in applied geodesy, as Gauss devoted a significant amount of time to the geodetic surveying and mapping of Hannover and other Germanic regions. See Dunnington et al. [2004, Ch. 10] for a fairly detailed account of Gauss' extensive work in, and contributions to, the field of geodesy.

The German geodesist F.R. Helmert also made significant contributions to the theory of least-squares [Helmert, 1907]; the least-squares solution within the *Gauss-Helmert model* (GHM) is perhaps still one of the most versatile techniques available for estimating unknown (fixed) parameters within a nonlinear functional model.

The Swedish geodesist Bjerhammar is credited with rediscovering Moore's generalized inverse and linking it to solutions of linear systems [Ben-Israel and Greville, 2003, p. 4]. A particular generalized inverse, originally introduced as "stochastic ring inverse," was eventually called Moore-Penrose inverse and is now perhaps the most important generalized inverse for treating rank-deficient models in the context of parameter estimation.

The Austrian geodesist H. Moritz introduced *least-squares prediction* [1970, 1978, for example] in order to integrate a wide range of observables measured for determination of the gravity field of the Earth. These ideas were further promoted and advanced by the Danish geodesist T. Krarup [1969, 2006], who apparently coined the phrase *least-squares collocation* for this predictor.

The number of less famous connections between geodetic science and advances in linear algebra, least-squares minimization, and statistical estimation is too long to list. But it is interesting to note that in more modern times a popular college textbook on linear algebra [Strang, 1988, p. 145] mentions that a US geodetic institution (the National Geodetic Survey) planned to solve the largest system of equations ever attempted at that time (approximately 6,000,000 equations in 400,000 unknowns). This project motivated the work of Golub and Plemmons [1980], who proposed an orthogonal decomposition strategy to handle efficiently large, sparse systems of equations. The title of the popular textbook *Linear Algebra, Geodesy, and GPS* [Strang and Borre, 1997] also underscores the linkage between these fields.

The tradition of geodetic-science influence in estimation and prediction theory has continued in more recent years with new extensions to EIV modeling published by geodesists. For example, the EIV model with constraints [Schaffrin, 2006], the EIV model with stochastic prior information [Schaffrin, 2009], the EIV model transformed to a system of nonlinear condition equations [Schaffrin and Wieser, 2011], the EIV model in the presence of outliers [Schaffrin and Uzun, 2011], the EIV model with full cofactor matrices and with cross-correlation between them [Fang, 2011], structured EIV models [Schaffrin et al., 2012a], the EIV model with a singular covariance matrix [Schaffrin et al., 2012b], the EIV model with full cofactor matrices but *without* cross-correlation [Mahboub, 2012], just to mention some of the important contributions.

1.2.2 The notion of the total least-squares principle

A historical review

Essentially, the idea of total least-squares consists of minimizing *the errors in all measurement variables* that enter into the model, rather than to minimize only the errors in the dependent variables. The notion is often explained in the context of linear regression or in data fitting to linear functions, such as a line in 2-D space. Adcock [1877] is generally the first to be credited with a TLS-problem statement in the English literature, which he succinctly made in a publication that barely spans a full page. His brief paper mentions applications to point, line, and surface fitting.

Pearson [1901] developed a solution that treated all errors in q variables of line and plane fitting by minimizing the "*mean square residual.*" He showed that the solution "*depends only on a knowledge of the means, standard-deviations, and correlations of the q variables*" and solved the problem by finding the least (non-zero) root of the characteristic equation for a certain matrix, based only on these quantities. Interestingly, the form of the matrix for which he found the least root of its characteristic equation (see his p. 562) resembles the matrices shown in Schaffrin et al.

[2006, eq. (1.15)] and Schaffrin [2007, eq. (38)], for which the minimum eigenvalues were required. (See Appendix B for a comparison between Pearson's and Schaffrin's formulations.) Pearson also proved that "*The best-fitting straight line for a system of points in a space of any order goes through the centroid of the system*," which is certainly true for iid data.

The least-squares solution within the Gauss-Helmert model by Helmert [1907] is quite versatile for minimizing errors in all variables, though its use seems to be missing from the English literature until much more recent times, as evidenced by several authors who investigated the treatment of errors in all measurement variables from the 1930's through the 1980's, such as Aitken [1935], Wald [1940], Plackett [1949], Linnik [1961], York [1966], and Demmel [1985].

Another concept associated with TLS is *orthogonal regression*, which gets its name from the fact that, in the case of iid data, the observed data are projected orthogonally onto the fitted line or surface. The idea dates back to at least Pearson [1901], who did not use the term orthogonal regression but obviously expressed the idea in the statement "*…a good fit will clearly be obtained if we make the sum of the squares of the* perpendiculars *from the system of points upon the line or plane a minimum*." An illustration of this concept is shown in Figure 1.1, which resembles Pearson's graph. It is important to note, however, that as soon as a weight matrix is incorporated into the minimization problem, the geometric orthogonality gives way to projections having directions that depend on a ratio of standard deviations, as pointed out by Schaffrin and Wieser [2008, see their Figure 1] and also earlier by Gerhold [1969, see his footnote number 2].

Models for minimizing all the errors in 2D line-fitting

Given the oft-used example of 2-D line-fitting in explaining the concept of total least squares, it seems apropos to show here how this could be accomplished within various different models, before proceeding to the theoretical development of TLS adjustment within the EIV model.

A very common and straight forward way to model a line in 2-D space as a function of n coordinate pairs (x_i, y_i), $i = 1, \ldots, n$, is

$$y_i = \xi_1 x_i + \xi_2 + e_{y_i}, \quad i = 1, \ldots, n, \tag{1.1a}$$

$$\boldsymbol{e}_y := \left[e_{y_1}, \ldots, e_{y_n}\right]^T \sim (\boldsymbol{0}, \sigma_0^2 P^{-1}), \tag{1.1b}$$

where ξ_1 and ξ_2 are the unknown slope and intercept parameters, respectively, for the 2-D line; e_{y_i} is the random error in the dependent variable y_i, with the $n \times n$ given positive-definite weight matrix P for the errors e_{y_i}, and σ_0^2 as the unknown variance component. This represents a classical *Gauss-Markov model* (GMM). The least-squares estimator within this model is often called ordinary least-squares estimator (OLSE), if the weight matrix is replaced by $P = I_n$, to distinguish it from the

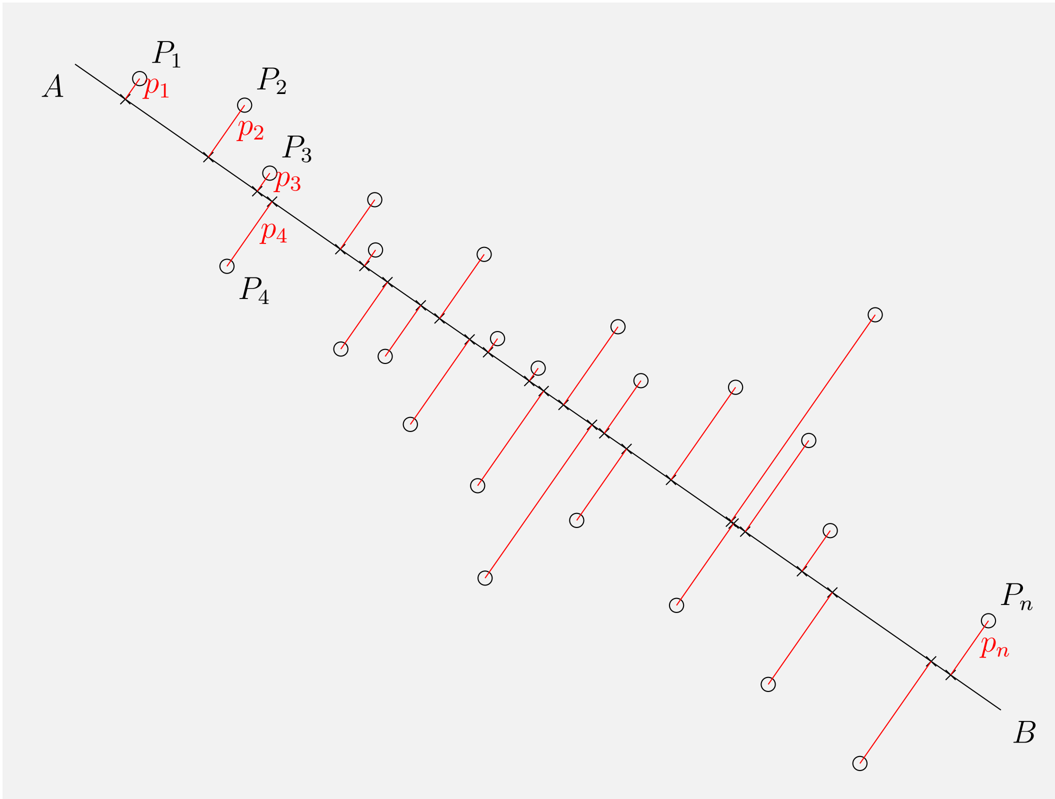


Figure 1.1: Fitted line AB based on data points $P_1, \ldots, P_n$ and associated orthogonal distances $p_1, \ldots, p_n$, c.f. Pearson [1901].

weighted (or generalized) least-squares estimator (GLSE) in the case that weights are considered. (See Rao et al. [2008, Chapters 3 and 4].)

The problem oftentimes is as follows: not only are the y-coordinates contaminated by measurement error, but so are the x-coordinates, and the choice of which coordinate to model as the independent, errorless one might be completely arbitrary. However, the estimation results may be quite different depending on which variable is chosen as the independent one. As Karl Pearson [1901] aptly stated, "*the most probable stature of a man with a given length of leg l being s, the most probable length of leg for a man of stature s will not be l.*"

In any case, if both coordinate variables are contaminated by measurement error, a more accurate model could be found for the problem. The model used should somehow incorporate the errors in the x-coordinates $\boldsymbol{e}_x := \left[e_{x_1}, \ldots, e_{x_n}\right]^T$, and the least-squares estimator for the unknown parameters should minimize the norm of the *total (extended) error-vector* $\boldsymbol{e} := \left[\boldsymbol{e}_x^T, \quad \boldsymbol{e}_y^T\right]^T$, i.e., minimize the quadratic form

$\boldsymbol{e}^T P \boldsymbol{e}$. Of course, in this case the dimension of the weight matrix P is doubled to account for the n additional errors in the independent variables.

There is more than one model that could be used to handle the above described "total least-squares problem." In fact, the GMM itself would work if the problem were stated in the form of the so-called parametric equation for the line. In such a model, n "nuisance parameters" would need to be introduced, one for each coordinate pair, and also n additional equations would be written. The extra parameters and equations cancel each other in terms of model redundancy. One option for such a (now nonlinear) GMM could look like

$$x_i = \xi_{i+2} + e_{x_i}, \tag{1.2a}$$

$$y_i = \xi_1 \cdot \xi_{i+2} + \xi_2 + e_{y_i}, \tag{1.2b}$$

$$\boldsymbol{e} := \begin{bmatrix} e_{x_1}, \ldots, e_{x_n}, e_{y_1}, \ldots, e_{y_n} \end{bmatrix}^T \sim (\mathbf{0}, \sigma_0^2 P^{-1} =: \sigma_0^2 \begin{bmatrix} Q_x & Q_{xy} \\ Q_{yx} & Q_y \end{bmatrix}), \tag{1.2c}$$

where $i = 1, \ldots, n$; ξ_1 is the slope of the line; ξ_2 is the y-intercept, and ξ_{i+2} is the i-th nuisance parameter. This approach may be unappealing if the data set is large, since the size of the unknown parameter vector $\boldsymbol{\xi}$ grows with the data set. But the point to make is that the least-squares estimator that minimizes $\boldsymbol{e}^T P \boldsymbol{e}$ will minimize the (weighted) errors of *all* measurement variables in the model. It is noted that this approach is analogous to fitting circles in parametric form, as in Gander et al. [1994].

Like the preceding approach, a different technique by Reinking [2001], who suggested to apply "Helmert's knack" (i.e., "Helmerts Kunstgriff," cf. Helmert (1907, p. 286)), also introduces an additional n parameters to the model. However, Schaffrin [2007] showed how these extra parameters could be removed before construction of the normal equations, thereby leading to a Gauss-Helmert model (GHM). In fact, the GHM, as described in detail in Chapter 4, provides another feasible model for minimizing all measurement errors in 2-D line-fitting. The least-squares solution within the GHM will produce the same parameter estimates and residuals as those generated by the least-squares solution within the nonlinear GMM for the parametric form of the line described above.

Finally, the total least-squares solution within the EIV model also can be used to minimize all the measurement errors and to generate the same parameter estimates and residuals as the respective estimators within the GMM and GHM described above. Moreover, provided that the weight matrix has a certain form, the TLS problem within the EIV model can be cast as a *minimum eigenvalue problem* for a certain augmentation of the system of normal equations, see Van Huffel and Vandewalle [1991, p. 37] or Felus and Schaffrin [2005, §2.2], for example. Or, the techniques described later on in this dissertation can be used for the case of more general weight matrices within the EIV model.

Though it may be possible to minimize equivalently the total (extended) error-vector via least-squares estimators within different models, e.g, GMM, GHM, or EIV model,

> *the terms total least-squares (TLS) and TLS solution as used in this dissertation will mean the least-squares solution within the EIV model without linearization.*

1.2.3 The development of TLS adjustment within the EIV model

Golub and Reinsch [1970, §2.4] presented a model with errors in the coefficient matrix A as well as in the observation vector $\boldsymbol{y}$ (they used $\boldsymbol{b}$); however, the only weighting permitted was that of a scalar quantity used to weight the inner product of the observation vector in their minimization problem. A few years later, Golub [1973] presented what eventually became known as the classical errors-in-variables model, where two positive-definite diagonal matrices were introduced in the least-squares minimization statement. (See his equations (6.1) and (6.2).) There he solved the parameter estimation problem by a singular value decomposition (SVD) of the matrix for which the Frobenius norm was minimized. In Golub and Van Loan [1979] the model was extended for the treatment of multiple vectors on the right side (i.e., the observation vector $\boldsymbol{b}$ becomes an observation matrix B, meaning also that the vector of parameters becomes a matrix of parameters). It was in this publication that the phrase "the total least squares (TLS) problem" was first used. Shortly thereafter, these authors published their well known paper [Golub and Van Loan, 1980] on total least squares, which included an SVD analysis of the TLS problem, a sensitivity analysis, and a comparison of TLS to the ordinary least-squares problem.

The TLS problem has also been treated briefly in the well known textbooks by Golub and Van Loan [1996, §12.3] and Björck [1996, §4.6], where the SVD technique is used to estimate the model parameters. Beginning in the early 1990's, the work of Van Huffel and others [Van Huffel and Vandewalle, 1991, Van Huffel, 1997, 2004] drew much attention to errors-in-variables modeling and the total least-squares problem, where by then the phrase errors-in-variables model had taken hold as a name for the model first introduced by Golub and Van Loan. The mathematical model by Golub and Van Loan is introduced in some detail in §2.3 of this work and will from here on be called *the classical EIV model*.

1.2.4 A progression of weighting schemes within the EIV model

As mentioned in the introduction, recent developments in EIV modeling allow for very arbitrary weighting of data, assuming that the weight matrices are symmetric positive-definite [Fang, 2011]. In contrast, the earliest EIV models were very limited in their handling of data weighting. (See §2.3 below for further details.) In fact, weights were often only admitted in the least-squares minimization statement, rather than introduced as a stochastic component of the model. The recent extensions, including the handling of singular cofactor matrices in this work, represent a maturation of the EIV model that should make its use more appealing for a wider range of problems. To gain an appreciation of the progression to the current state, one may be interested in reviewing the works enumerated below, which are but a sample of the efforts made to accommodate data weighting within the EIV model (or in some instances in the TLS objective function rather than the model) since around the time that the classical EIV model was introduced.

1. Heteroscedastic weighting for the observation vector only: Golub and Reinsch [1970, §2.4] presented a TLS minimization that permitted a nonsingular, diagonal weight matrix for the errors of the observation vector. However, the problem did not include weights for the errors of the data matrix.

2. Row- and column-dependent weighting: the TLS minimization presented by Golub [1973, eq. (6.2)], Golub and Van Loan [1979, eqs. (7), (8)], and Golub and Van Loan [1980, eqs. (1.1), (1.4)] introduced two positive-definite diagonal weight matrices. One pre-multiplies and the other post-multiplies the augmented error matrix appearing in the matrix norm to be minimized. This weighting scheme gives rise to row and column dependence among the weights, which turns out, in general, not to be very useful for problems in geodetic science. See §2.3 below for further explanation.

3. Demmel [1985] generalized the weighting introduced by Golub and Van Loan (previous item) for the case where only some elements of the data matrix and the data vector are contaminated by error.

4. Here the work of Deming [1964] is mentioned, not because it extends the classical EIV model (it actually precedes it), but rather because it was cited by Golub and Van Loan [1980] as being more general than their TLS problem. In fact, Deming's approach is the same as the least-squares solution within the GHM with heteroscedastic data (see his Part D). So it was indeed more general than Golub and Van Loan as far as data weighting goes. However, the application of Deming's method could lead to the pitfalls discussed by Pope [1972], as Deming states that it is sufficient to evaluate the derivatives needed for the (linear) normal equations at the observed data points and approximate parameters rather than at the adjusted data points and estimated parameters (ibid, p. 139). By now it should be widely known that the risk of taking such short cuts could lead to erroneous estimates for the parameters or preclude convergence of the solution, and thus that practice must be avoided. (See §4.1 for further discussion.)

Also mentioned by Golub and Van Loan in this context is the work of Gerhold [1969], who used a Lagrangian approach to treat independent weights for all variables and employed the Newton-Raphson method to solve the resulting nonlinear system of equations.

5. Van Huffel and Vandewalle [1989] wrote about Generalized Total Least-Squares (GTLS), where a "weight" matrix was permitted for the errors in the data matrix A, given that A could contain some error-free columns. However, as in the case of Golub and Van Loan [1980] mentioned in the first item, their weight matrix does not refer to a variance-covariance matrix for the data and thus also turns out, in general, not to be very useful in geodetic applications. Schaffrin and Wieser [2008] have called this GTLS approach "equilibrated TLS" to distinguish it from weighted TLS.

6. Markovsky et al. [2006] introduced element-wise weighting to the TLS problem, where the random errors of the data matrix A, augmented with the random errors of the observation vector $\boldsymbol{y}$ (a matrix B in their paper), were considered to be row-wise independent. Thus, for n observation equations and m unknown parameters, their approach allows for n non-singular covariance matrices, each of size $(m+1) \times (m+1)$. They further generalized the problem to handle the case where some data were considered noise free, thus allowing for some of the n covariance matrices to be singular (ibid, §3).

7. Schaffrin and Wieser [2008] made an important step towards "general variance-covariance matrices" within the EIV model by allowing a completely general cofactor matrix for errors of the observation vector $\boldsymbol{y}$ and a slightly more restricted cofactor matrix for the errors in the data matrix A, which had a certain Kronecker-product structure imposed upon it. An attractive feature of their estimator was that it handled constant columns of the data matrix (e.g., problems having an intercept parameter) without any special considerations beyond the Kronecker-product structure for the data cofactor matrix.

8. The dissertation by Fang [2011] presented an estimator that could handle arbitrary, symmetric positive-definite weight matrices, including correlation between the errors of the data vector and the data matrix. Fang proposed three different, but algebraically equivalent, forms of the estimator. In independent work, Mahboub [2012] derived an estimator identical to one of the forms presented by Fang, except that Mahboub's work did not account for correlation between the observation vector and the data matrix. The developments of these authors are important, as they represent a maturation of EIV modeling that now permits very general covariance matrices, as long as they are not singular. As shown in Appendix D, both Fang's and Mahboub's works are generalizations of Schaffrin and Wieser [2008], with Fang's estimator being the most general.

9. The work of this dissertation now allows for singular cofactor matrices to be treated within the EIV model. This development is analogous to the work of Neitzel and Schaffrin [2012], where a least-squares estimator within the GHM having a singular dispersion matrix was presented, as well as a criterion for this estimator to be unique.

1.3 Comments on the notation used in this dissertation

An important task in writing mathematical works is to employ a clear and consistent notation. To do otherwise places too much burden on the reader to understand the meaning of variables and notation that may change from place to place. Of course, some context-dependent reuse of variable names might be unavoidable in lengthy or detailed works but should be kept to a minimum. In particular, when the meanings of variable names change, the new meanings should be clearly stated and their scope of use should be clearly delineated, a practice adhered to in this work.

Many textbooks on linear algebra use boldface type for matrices and vectors, using uppercase characters for the former and lowercase for the latter. Most authors follow this case-convention, but many do not use boldface at all. In Ben-Israel and Greville [2003] boldface is used for vectors but not for matrices; this is the style adopted in this work. Also, in this dissertation, matrices are always uppercase, whereas vectors are always lowercase (with an exception in Chapter 4 where $\boldsymbol{Y}$ and $\boldsymbol{\Xi}$ are used for vectors) and are always typeset with boldface font. Furthermore, boldface font is only used for vectors, and vectors are always column vectors. If a row vector is needed, it is written as the transpose of a column vector. With all but one exception, lowercase characters are used for scalars. The exception is the use of Ω for the SSR (sum of squared residuals), which is common among some authors.

Another notation convention employed here is one long used by Professor Burkhard Schaffrin. The convention is helpful because it reveals something about the nature of the unknown variables. Greek characters are used for unknown, nonrandom variables, and Latin characters are used for unknown, random variables. An example of the former is the parameter vector $\boldsymbol{\xi}$ in the Gauss-Helmert model, while an example of the latter is the random error vector $\boldsymbol{e}$ in the same model. Of course, the estimates of unknown, nonrandom variables are themselves random, and for these a hat is placed on top. We speak of predictions of unknown, random variables; for these a tilde is used. So, when one reads $\hat{\boldsymbol{\xi}}$, one immediately realizes that the symbol refers to the estimator (or estimate in the context of the actual realization of the estimator) of the unknown, nonrandom vector $\boldsymbol{\xi}$. Likewise, the symbol $\tilde{\boldsymbol{e}}$ would refer to the predictor (or possibly the prediction) of the unknown, random vector $\boldsymbol{e}$.

Abbreviations used herein are mostly standard in the geodetic science and/or statistics literature. They are always spelled out in full in the first place they occur. However, for convenience of reference, all abbreviations used are defined in Table 1.1.

Table 1.1: List of abbreviations

Abbr.	Meaning
BLUUE	Best Linear Uniformly Unbiased Estimator
EIV	Errors-In-Variables
EIV-REM	Errors-In-Variables with Random Effects Model
GLS	Generalized Least Squares
GLSE	Generalized Least-Squares Estimator
GHM	Gauss-Helmert Model
GMM	Gauss-Markov Model
iid	Independent and Identically Distributed
LESS	LEast-Squares Solution
LESS-GHM	LEast-Squares Solution within the GHM
LSC	Least-Squares Collocation
nnd	nonnegative definite
OLS	Ordinary Least Squares
OLSE	Ordinary Least-Squares Estimator
REM	Random Effects Model
SSR	Sum of Squared Residuals
Std dev	Standard deviation
SVD	Singular Value Decomposition
TLS	Total Least Squares
TLSC	Total Least-Squares Collocation
TSSR	Total Sum of Squared Residuals
WLS	Weighted Least Squares
WTLS	Weighted Total Least Squares
WTLSC	Weighted Total Least-Squares Collocation
WTLSS-EIV	Weighted Total Least-Squares Solution within the EIV model

1.4 Linear algebra references

This dissertation makes heavy use of linear algebra. Some operators and special matrices, such as the vec operator, the Kronecker product, and the commutation matrix, may be unfamiliar to the reader. For this reason, many of their properties used in this work are listed in Appendix E. For further reference to these topics in linear algebra any of these works is recommended: Lütkepohl [1996], Horn and Johnson [1994], Harville [1997], or Magnus and Neudecker [2007].

Chapter 2: The EIV Model with Correlation

The errors-in-variables (EIV) model with *correlated errors* among the observation vector $\boldsymbol{y}$ and among the *data matrix* A, as well as cross-correlation between the random errors in $\boldsymbol{y}$ and A, can be written as

$$\boldsymbol{y} = (\underset{n\times m}{A} - E_A)\boldsymbol{\xi} + \boldsymbol{e}_y, \quad \operatorname{rk} A = m, \tag{2.1a}$$

$$\underset{n(m+1)\times 1}{\begin{bmatrix} \boldsymbol{e}_y \\ \boldsymbol{e}_A \end{bmatrix}} := \begin{bmatrix} \boldsymbol{e}_y \\ \operatorname{vec} E_A \end{bmatrix} \sim \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \sigma_0^2 Q \right), \tag{2.1b}$$

where

$$Q = \underset{n(m+1)\times n(m+1)}{\begin{bmatrix} Q_y & Q_{yA} \\ Q_{Ay} & Q_A \end{bmatrix}}, \quad \operatorname{rk} Q \leq n(m+1), \tag{2.1c}$$

is a symmetric nnd cofactor matrix. The model variables are defined as follows:

$\boldsymbol{y}$	denotes the $n \times 1$ observation vector,
$\boldsymbol{\xi}$	the $m \times 1$ (unknown) parameter vector,
A	the $n \times m$ data matrix with $n > m = \operatorname{rk} A$,
$\boldsymbol{e}_y$	the $n \times 1$ (unknown) random error vector associated with $\boldsymbol{y}$,
E_A	the $n \times m$ (unknown) random error matrix associated with A,
$\boldsymbol{e}_A$	the $nm \times 1$ vectorized form of E_A,
σ_0^2	the (unknown) variance component, and
Q	the $n \times n$ symmetric nnd *cofactor matrix*.

If the cofactor matrix Q is non-singular, there exists a unique *weight matrix* P defined as

$$P := Q^{-1} = \begin{bmatrix} \underset{n\times n}{P_{11}} & \underset{n\times nm}{P_{12}} \\ \underset{nm\times n}{P_{21}} & \underset{nm\times nm}{P_{22}} \end{bmatrix} \text{ if, and only if, } \operatorname{rk} Q = n(m+1). \tag{2.2}$$

The product of the variance component σ_0^2 and the cofactor matrix Q is called the *covariance matrix*, defined by $\Sigma := \sigma_0^2 Q$, which is also called variance-covariance matrix (or vcm for short) in the literature. In this work, it is the cofactor matrix Q (or its inverse, the weight matrix P) that is mainly discussed, with very few references to the covariance matrix Σ specifically. As in this work, the cofactor matrix Q appears as a (potentially) fully populated matrix in Fang [2011], where it was defined as non-singular. However, in this contribution, the non-singular restriction has been removed. The case of *singular cofactor matrices* is dealt with in detail in Chapter 3.

Note also that the *data matrix* A has full column rank in model (2.1). It is also called design matrix, coefficient matrix, or information matrix in other contexts, but here the use of the term data matrix underscores the fact that some or all of the elements of A are comprised of random entries representing measurement data. Moreover, in keeping with the customary usage of the EIV model, it is assumed that all random entries of A are *linear* in the measurement variables.

In the following, total least-squares (TLS) estimators are derived for the unknown parameters of the EIV model, and algorithms for their use in numerical computations are presented. This is followed by numerical applications to 2-D line-fitting and 2-D similarity transformations.

2.1 WTLS adjustment within the EIV model with correlation

Let us assume for the moment that the cofactor matrix Q of model (2.1) is regular (i.e., non-singular) so that $\operatorname{rk} Q = n(m+1)$, and therefore $Q^{-1} = P$ exists. Then the *weighted total least-squares (WTLS) problem* can be stated as

$$\underset{\boldsymbol{\xi}}{\text{minimize}}: \begin{bmatrix} \boldsymbol{e}_y^T, & \boldsymbol{e}_A^T \end{bmatrix} \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{e}_y \\ \boldsymbol{e}_A \end{bmatrix} \quad \text{such that} \quad \boldsymbol{y} = (A - E_A)\boldsymbol{\xi} + \boldsymbol{e}_y. \tag{2.3}$$

Here a Lagrangian approach is taken by introducing an unknown $n \times 1$ vector of Lagrange multipliers $\boldsymbol{\lambda}$ to solve for the parameter vector $\boldsymbol{\xi}$. The Lagrange target function is written as

$$\begin{aligned} \phi(\boldsymbol{e}_y, \boldsymbol{e}_A, \boldsymbol{\xi}, \boldsymbol{\lambda}) &= \boldsymbol{e}_y^T P_{11} \boldsymbol{e}_y + 2\boldsymbol{e}_y^T P_{12} \boldsymbol{e}_A + \boldsymbol{e}_A^T P_{22} \boldsymbol{e}_A + \\ &+ 2\boldsymbol{\lambda}^T \Big(\boldsymbol{y} - A\boldsymbol{\xi} - \boldsymbol{e}_y + \big(\boldsymbol{\xi}^T \otimes I_n \big) \boldsymbol{e}_A \Big) = \text{stationary}. \end{aligned} \tag{2.4}$$

The symbol $\otimes$ represents the Kronecker product (or Kronecker-Zehfuss product), several rules for which are listed in Appendix E.

The Euler-Lagrange (or first-order) necessary conditions are satisfied by

$$\frac{1}{2}\frac{\partial \phi}{\partial \boldsymbol{e}_y} = P_{11}\tilde{\boldsymbol{e}}_y + P_{12}\tilde{\boldsymbol{e}}_A - \hat{\boldsymbol{\lambda}} \doteq \boldsymbol{0}, \tag{2.5a}$$

$$\frac{1}{2}\frac{\partial \phi}{\partial \boldsymbol{e}_A} = P_{22}\tilde{\boldsymbol{e}}_A + P_{21}\tilde{\boldsymbol{e}}_y + \left(\hat{\boldsymbol{\xi}} \otimes I_n\right)\hat{\boldsymbol{\lambda}} \doteq \boldsymbol{0}, \tag{2.5b}$$

$$\frac{1}{2}\frac{\partial \phi}{\partial \boldsymbol{\xi}} = -A^T\hat{\boldsymbol{\lambda}} + \tilde{E}_A^T\hat{\boldsymbol{\lambda}} \doteq \boldsymbol{0}, \tag{2.5c}$$

$$\frac{1}{2}\frac{\partial \phi}{\partial \boldsymbol{\lambda}} = \boldsymbol{y} - A\hat{\boldsymbol{\xi}} - \tilde{\boldsymbol{e}}_y + \left(\hat{\boldsymbol{\xi}}^T \otimes I_n\right)\tilde{\boldsymbol{e}}_A \doteq \boldsymbol{0}. \tag{2.5d}$$

Here hats and tildes are placed over the unknown nonrandom and random vectors, respectively, to indicate the particular quantities that satisfy the first-order, homogeneous system of condition equations. The Hessian matrix can be derived from the second partial-derivatives of ϕ with respect to $\boldsymbol{e}_y^T$ and $\boldsymbol{e}_A^T$, which yields

$$\frac{1}{2}\frac{\partial^2 \phi}{\partial \begin{bmatrix} \boldsymbol{e}_y \\ \boldsymbol{y}_A \end{bmatrix} \partial \begin{bmatrix} \boldsymbol{e}_y^T, & \boldsymbol{e}_A^T \end{bmatrix}} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}. \tag{2.6}$$

Thus, the *sufficient condition* for minimization — namely that the Hessian matrix (2.6) is positive-(semi)definite — is satisfied.

Substituting (2.5b) into (2.5a) and then, vice versa, substituting (2.5a) into (2.5b), leads to predicted error vectors (or *residual* vectors)

$$\tilde{\boldsymbol{e}}_y = \left(P_{11} - P_{12}P_{22}^{-1}P_{21}\right)^{-1}\left[P_{12}P_{22}^{-1}\left(\hat{\boldsymbol{\xi}} \otimes I_n\right) + I_n\right]\hat{\boldsymbol{\lambda}}, \tag{2.7a}$$

$$\tilde{\boldsymbol{e}}_A = -\left(P_{22} - P_{21}P_{11}^{-1}P_{12}\right)^{-1}\left[P_{21}P_{11}^{-1} + \left(\hat{\boldsymbol{\xi}} \otimes I_n\right)\right]\hat{\boldsymbol{\lambda}} \tag{2.7b}$$

as a function of the block components of the weight matrix P. Continuing under the assumption that Q is regular (and thus also P_{11} and P_{22}), the following identities, based on the *Banachiewicz inversion formula* [Zhang, 2005, p. 11] for block matrices, relate the blocks of Q to those of P:

$$\left(P_{11} - P_{12}P_{22}^{-1}P_{21}\right)^{-1} = Q_y, \tag{2.8a}$$

$$\left(P_{11} - P_{12}P_{22}^{-1}P_{21}\right)^{-1}P_{12}P_{22}^{-1} = -Q_{yA} = -Q_{Ay}^T, \tag{2.8b}$$

$$\left(P_{22} - P_{21}P_{11}^{-1}P_{12}\right)^{-1} = Q_A, \tag{2.8c}$$

$$\left(P_{22} - P_{21}P_{11}^{-1}P_{12}\right)^{-1}P_{21}P_{11}^{-1} = -Q_{Ay} = -Q_{yA}^T. \tag{2.8d}$$

Using the equations (2.8), the predicted error vectors can be rewritten in terms of the block components of the cofactor matrix Q, rather than the weight matrix P, as

$$\tilde{\boldsymbol{e}}_y = \left[-Q_{yA}\left(\hat{\boldsymbol{\xi}} \otimes I_n\right) + Q_y\right]\hat{\boldsymbol{\lambda}}, \tag{2.9a}$$

$$\tilde{\boldsymbol{e}}_A = \left[Q_{Ay} - Q_A\left(\hat{\boldsymbol{\xi}} \otimes I_n\right)\right]\hat{\boldsymbol{\lambda}}. \tag{2.9b}$$

Substituting (2.9a) and (2.9b) into (2.5d) leads to

$$
\begin{aligned}
&\boldsymbol{y} - A\hat{\boldsymbol{\xi}} = \\
&= -Q_{yA}\Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)\hat{\boldsymbol{\lambda}} + Q_y\hat{\boldsymbol{\lambda}} - \Big(\hat{\boldsymbol{\xi}}^T \otimes I_n\Big)Q_{Ay}\hat{\boldsymbol{\lambda}} + \Big(\hat{\boldsymbol{\xi}}^T \otimes I_n\Big)Q_A\Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)\hat{\boldsymbol{\lambda}} = \\
&= \Big[Q_y - Q_{yA}\Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big) - \Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)^T Q_{Ay} + \Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)^T Q_A\Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)\Big]\hat{\boldsymbol{\lambda}} = \\
&= Q_1\hat{\boldsymbol{\lambda}},
\end{aligned}
\tag{2.10a}
$$

with

$$
\underset{n\times n}{Q_1} := \Big[Q_y - Q_{yA}\Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big) - \Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)^T Q_{Ay} + \Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)^T Q_A\Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)\Big] = Q_1(\hat{\boldsymbol{\xi}}).
\tag{2.10b}
$$

Since in this section it is assumed that Q (and therefore Q_y and Q_A) is regular, the matrix Q_1 is also regular, which permits the solution

$$
\hat{\boldsymbol{\lambda}} = Q_1^{-1}(\boldsymbol{y} - A\hat{\boldsymbol{\xi}})
\tag{2.11}
$$

for the vector of Lagrange multipliers $\boldsymbol{\lambda}$ in terms of the estimated parameter vector $\hat{\boldsymbol{\xi}}$. Following the logic of Schaffrin et al. [2012a], (2.5c) is rewritten as

$$
\begin{aligned}
A^T\hat{\boldsymbol{\lambda}} = \tilde{E}_A^T\hat{\boldsymbol{\lambda}} = \Big(\hat{\boldsymbol{\lambda}}^T \otimes I_m\Big)\operatorname{vec}\Big(\tilde{E}_A^T\Big) &= \\
= \Big(\hat{\boldsymbol{\lambda}}^T \otimes I_m\Big)K_{nm}\tilde{\boldsymbol{e}}_A &= \Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)^T\tilde{\boldsymbol{e}}_A
\end{aligned}
\tag{2.12}
$$

for a unique commutation (vec-permutation) matrix K_{nm} of dimension $nm \times nm$. (See Appendix E for properties of commutation matrices.) Using (2.9b) and (2.11), equation (2.12) can be further developed as

$$
-A^T\hat{\boldsymbol{\lambda}} = \Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)^T\Big[-Q_{Ay} + Q_A\Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)\Big]\hat{\boldsymbol{\lambda}},
\tag{2.13a}
$$

respectively,

$$
\begin{aligned}
&-A^TQ_1^{-1}(\boldsymbol{y} - A\hat{\boldsymbol{\xi}}) = \\
&\qquad = \Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)^T\Big[-Q_{Ay} + Q_A\Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)\Big]Q_1^{-1}(\boldsymbol{y} - A\hat{\boldsymbol{\xi}}) = \\
&\qquad = \Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)^T\Big[-Q_{Ay}Q_1^{-1} + Q_A\Big(\hat{\boldsymbol{\xi}} \otimes Q_1^{-1}\Big)\Big](\boldsymbol{y} - A\hat{\boldsymbol{\xi}}) = \\
&\qquad = R_1(\boldsymbol{y} - A\hat{\boldsymbol{\xi}}),
\end{aligned}
\tag{2.13b}
$$

with

$$
\underset{m\times n}{R_1} := \Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)^T\Big[-Q_{Ay}Q_1^{-1} + Q_A\Big(\hat{\boldsymbol{\xi}} \otimes Q_1^{-1}\Big)\Big] = R_1(\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\lambda}}).
\tag{2.13c}
$$

Equation (2.13) yields what Schaffrin et al. [2012a] referred to as "the generalized normal equations"

$$
\boxed{\Big(A^TQ_1^{-1} + R_1\Big)A \cdot \hat{\boldsymbol{\xi}} = \Big(A^TQ_1^{-1} + R_1\Big)\boldsymbol{y}},
\tag{2.14}
$$

which is identical in form to their normal equations; however their formulation does not incorporate the cross-correlation matrix Q_{yA}. On the other hand, when differences in symbols are taken into account, (2.14) is found to be identical to equation (4.25) in Fang [2011], which has the form $\hat{\boldsymbol{\xi}} = \Big[(A - \tilde{E}_A)^T Q_1^{-1} A\Big]^{-1} (A - \tilde{E}_A)^T Q_1^{-1}\boldsymbol{y}$. This becomes even more obvious when noting that $R_1 = -\tilde{E}_A^T Q_1^{-1}$, which is proved below.

Proof that $R_1 = -\tilde{E}_A^T Q_1^{-1}$:
An expression for $\operatorname{vec}\tilde{E}_A =: \tilde{\boldsymbol{e}}_A$ is given in (2.9b), thus it is helpful to employ the vec operator:

$$
\begin{aligned}
\operatorname{vec}(-\tilde{E}_A^T Q_1^{-1}) &= -\Big(Q_1^{-1} \otimes I_m\Big) \operatorname{vec}\tilde{E}_A^T = \\
&= -\Big(Q_1^{-1} \otimes I_m\Big) K_{nm} \tilde{\boldsymbol{e}}_A = \\
&= -\Big(Q_1^{-1} \otimes I_m\Big) K_{nm} \Big[Q_{Ay} - Q_A\Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)\Big]\hat{\boldsymbol{\lambda}} = \\
&= -\Big(\hat{\boldsymbol{\lambda}}^T \otimes (Q_1^{-1} \otimes I_m) K_{nm}\Big) \operatorname{vec}\Big[Q_{Ay} - Q_A\Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)\Big] = \\
&= -\Big(Q_1^{-1} \otimes I_m \otimes \hat{\boldsymbol{\lambda}}^T\Big) \operatorname{vec}\Big[Q_{Ay} - Q_A\Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)\Big].
\end{aligned}
$$

Now apply the vec operator to R_1:

$$
\begin{aligned}
\operatorname{vec} R_1 &= \operatorname{vec}\Big\{\Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)^T \Big[-Q_{Ay} + Q_A(\hat{\boldsymbol{\xi}} \otimes I_n)\Big] Q_1^{-1}\Big\} = \\
&= -\Big(Q_1^{-1} \otimes I_m \otimes \hat{\boldsymbol{\lambda}}^T\Big) \operatorname{vec}\Big[Q_{Ay} - Q_A\Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)\Big] = \\
&= -\operatorname{vec}(\tilde{E}_A^T Q_1^{-1}).
\end{aligned}
$$

Since $\dim R_1 = \dim(\tilde{E}_A^T Q_1^{-1})$, therefore

$$R_1 = -\tilde{E}_A^T Q_1^{-1}. \qquad \square$$

Along these lines, Appendix D makes some comparisons between the works of Schaffrin and Wieser [2008] and Fang [2011].

Following the approach of Schaffrin et al. [2012a], an algorithm to solve (2.14) is presented below. The algorithm is identical in structure to Schaffrin (ibid), but, in contrast to theirs, will handle the case of correlation between Q_y and Q_A. The algorithm is labeled Algorithm 1 for reference purposes. Algebraically, it is identical to "Algorithm 2" of Fang [2011]. Note that the superscript i in Algorithm 1 denotes the iteration number.

Algorithm 1 For the WTLS solution within the EIV model

Step 1: Compute an initial solution

$$\hat{\boldsymbol{\xi}}^{(0)} := N^{-1}\boldsymbol{c} \quad \text{for} \quad \Big[N, \quad \boldsymbol{c}\Big] := A^T Q_y^{-1} \Big[A, \quad \boldsymbol{y}\Big]. \tag{2.15a}$$

Step 2:
repeat For $i \in \mathbb{N}$, compute

$$Q_1^{(i)} := \Big[Q_y - Q_{yA}\Big(\hat{\boldsymbol{\xi}}^{(i-1)} \otimes I_n\Big) - \Big(\hat{\boldsymbol{\xi}}^{(i-1)} \otimes I_n\Big)^T Q_{Ay} + \\ + \Big(\hat{\boldsymbol{\xi}}^{(i-1)} \otimes I_n\Big)^T Q_A \Big(\hat{\boldsymbol{\xi}}^{(i-1)} \otimes I_n\Big)\Big] \tag{2.15b}$$

$$\hat{\boldsymbol{\lambda}}^{(i)} := \Big(Q_1^{(i)}\Big)^{-1}\Big(\boldsymbol{y} - A\hat{\boldsymbol{\xi}}^{(i-1)}\Big) \tag{2.15c}$$

$$R_1^{(i)} := \Big(I_m \otimes \hat{\boldsymbol{\lambda}}^{(i)}\Big)^T \Big[-Q_{yA}\Big(Q_1^{(i)}\Big)^{-1} + Q_A\Big(\hat{\boldsymbol{\xi}}^{(i-1)} \otimes \Big(Q_1^{(i)}\Big)^{-1}\Big)\Big] \tag{2.15d}$$

$$\hat{\boldsymbol{\xi}}^{(i)} = \Big[\Big(A^T\Big(Q_1^{(i)}\Big)^{-1} + R_1^{(i)}\Big)A\Big]^{-1}\Big(A^T\Big(Q_1^{(i)}\Big)^{-1} + R_1^{(i)}\Big)\boldsymbol{y} \tag{2.15e}$$

until

$$\Big\|\hat{\boldsymbol{\xi}}^{(i)} - \hat{\boldsymbol{\xi}}^{(i-1)}\Big\| < \delta \tag{2.15f}$$

for a chosen threshold δ.

2.1.1 Alternative solution using $(A - \tilde{E}_A)$

Now we wish to derive an alternative algorithm that allows us to work with a symmetric positive-definite normal-equations matrix and that also directly accounts for the errors in the design matrix by replacing A in Algorithm 1 with $A - \tilde{E}_A$. The derivation is as follows: using equations (2.10a), (2.9b), and (2.5c), the system of equations (2.5) can be expressed equivalently as

$$Q_1\hat{\boldsymbol{\lambda}} = \boldsymbol{y} - A\hat{\boldsymbol{\xi}}, \tag{2.16a}$$

$$\tilde{\boldsymbol{e}}_A = \Big[Q_{Ay} - Q_A\Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)\Big]\hat{\boldsymbol{\lambda}}, \tag{2.16b}$$

$$\Big(A - \tilde{E}_A\Big)^T\hat{\boldsymbol{\lambda}} = \boldsymbol{0}, \tag{2.16c}$$

or as

$$\begin{bmatrix} Q_1 & A - \tilde{E}_A \\ \Big(A - \tilde{E}_A\Big)^T & 0 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\lambda}} \\ \hat{\boldsymbol{\xi}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} - \tilde{E}_A\hat{\boldsymbol{\xi}} \\ \boldsymbol{0} \end{bmatrix}, \tag{2.17a}$$

$$\text{and } \operatorname{vec}\tilde{E}_A =: \tilde{\boldsymbol{e}}_A = \Big[Q_{Ay} - Q_A\Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)\Big]\hat{\boldsymbol{\lambda}}. \tag{2.17b}$$

The variable $\hat{\boldsymbol{\lambda}}$ can be eliminated from the first row of (2.17a) via premultiplication by $(A - \tilde{E}_A)^T Q_1^{-1}$, which then yields

$$\left[\left(A - \tilde{E}_A\right)^T Q_1^{-1}\left(A - \tilde{E}_A\right)\right]\hat{\boldsymbol{\xi}} = \left(A - \tilde{E}_A\right)^T Q_1^{-1}\left(\boldsymbol{y} - \tilde{E}_A\hat{\boldsymbol{\xi}}\right). \tag{2.18}$$

The matrix on the left side of (2.18) is symmetric and positive-definite, assuming $\operatorname{rk}(A - \tilde{E}_A) = \operatorname{rk} A = m$, and thus can be inverted by Cholesky factorization, though the equation must be solved by iteration since the parameter vector $\hat{\boldsymbol{\xi}}$ appears on both sides of the equation.

The following Algorithm 2 can be used to solve for $\hat{\boldsymbol{\xi}}$, while at the same time generating solutions $\hat{\boldsymbol{\lambda}}$ and $\tilde{\boldsymbol{e}}_A$ for the vector of Lagrange multipliers and the residual vector, respectively. Following Schaffrin et al. [2012a], the symbol Invec is introduced for the inverse of the vec operator as follows:

$$\operatorname{Invec}(\operatorname{vec} A) = A = [a_{ij}],\text{ where } a_{ij} \text{ is the } [(j-1)m+i]\text{-th element of } \operatorname{vec} A. \tag{2.19}$$

See Harville [1997], p. 340, for the relationship between the elements of a matrix and the elements of the vec of the same matrix. Note that the superscript i in Algorithm 2 denotes the iteration number.

Algorithm 2 For the WTLS solution within the EIV model

Step 1: Compute an initial solution

$$\hat{\boldsymbol{\xi}}^{(0)} := N^{-1}\boldsymbol{c} \quad \text{for } \begin{bmatrix} N, & \boldsymbol{c} \end{bmatrix} := A^T Q_y^{-1}\begin{bmatrix} A, & \boldsymbol{y} \end{bmatrix},\text{ and assign } \tilde{E}_A^{(0)} := 0. \tag{2.20a}$$

Step 2:

repeat For $i \in \mathbb{N}$, compute

$$\begin{aligned} Q_1^{(i)} := \Big[&Q_y - Q_{yA}\left(\hat{\boldsymbol{\xi}}^{(i-1)} \otimes I_n\right) - \left(\hat{\boldsymbol{\xi}}^{(i-1)} \otimes I_n\right)^T Q_{Ay} + \\ &+ \left(\hat{\boldsymbol{\xi}}^{(i-1)} \otimes I_n\right)^T Q_A\left(\hat{\boldsymbol{\xi}}^{(i-1)} \otimes I_n\right)\Big] \end{aligned} \tag{2.20b}$$

$$\begin{aligned} \hat{\boldsymbol{\xi}}^{(i)} = &\left[\left(A - \tilde{E}_A^{(i-1)}\right)^T \left(Q_1^{(i)}\right)^{-1}\left(A - \tilde{E}_A^{(i-1)}\right)\right]^{-1} \times \\ &\left[\left(A - \tilde{E}_A^{(i-1)}\right)^T\left(Q_1^{(i)}\right)^{-1}\left(\boldsymbol{y} - \tilde{E}_A^{(i-1)}\hat{\boldsymbol{\xi}}^{(i-1)}\right)\right] \end{aligned} \tag{2.20c}$$

$$\hat{\boldsymbol{\lambda}}^{(i)} = \left(Q_1^{(i)}\right)^{-1}\left[\left(\boldsymbol{y} - \tilde{E}_A^{(i-1)}\hat{\boldsymbol{\xi}}^{(i-1)}\right) - \left(A - \tilde{E}_A^{(i-1)}\right)\hat{\boldsymbol{\xi}}^{(i)}\right] \tag{2.20d}$$

$$\tilde{\boldsymbol{e}}_A^{(i)} = \left[Q_{Ay} - Q_A\left(\hat{\boldsymbol{\xi}}^{(i)} \otimes I_n\right)\right]\hat{\boldsymbol{\lambda}}^{(i)} \text{ and } \tilde{E}_A^{(i)} = \operatorname{Invec}\tilde{\boldsymbol{e}}_A^{(i)} \tag{2.20e}$$

until

$$\left\|\hat{\boldsymbol{\xi}}^{(i)} - \hat{\boldsymbol{\xi}}^{(i-1)}\right\| < \delta \tag{2.20f}$$

for a chosen threshold δ.

It is noted that (2.18) is the same as equation (4.26) in Fang [2011], which is associated with "Algorithm 3" therein.

Here it is important to note that either, or both, of the cofactor matrices Q_y or Q_A could actually be singular in the above algorithms. This is true as long as the matrix Q_1 that they are incorporated into does not become singular, in which case Algorithms 1 and 2 could not be used. The case where singularities in cofactor matrices Q_y and/or Q_A lead to a singular Q_1 matrix is investigated in Chapter 3.

2.2 Sum of squared residuals (SSR) for the WTLS solution

In the context of total least-squares (TLS), the sum of squared residuals (SSR) has been called *total sum of squared residuals (TSSR)* by Schaffrin et al. [2012a]. This is certainly a more descriptive label in this context, and it should not be confused with the term *total sum of squares* used in the context of variance analysis, as in Davis [2002, pp. 80, 195], for example.

The TSSR is a scalar-valued vector function, being the square of the (weighted) norm of the *total residual vector* computed by

$$\Omega = \left\| \begin{bmatrix} \tilde{\boldsymbol{e}}_y^T, & \tilde{\boldsymbol{e}}_A^T \end{bmatrix}^T \right\|_P^2 = \begin{bmatrix} \tilde{\boldsymbol{e}}_y^T, & \tilde{\boldsymbol{e}}_A^T \end{bmatrix} P \begin{bmatrix} \tilde{\boldsymbol{e}}_y \\ \tilde{\boldsymbol{e}}_A \end{bmatrix} = \tag{2.21a}$$

$$= \hat{\boldsymbol{\lambda}}^T Q_1 \hat{\boldsymbol{\lambda}} = \tag{2.21b}$$

$$= \hat{\boldsymbol{\lambda}}^T (\boldsymbol{y} - A\hat{\boldsymbol{\xi}}). \tag{2.21c}$$

Then a suitable approximation for the *estimated variance component* $\hat{\sigma}_0^2$ is given by dividing the TSSR by the model degrees of freedom (or redundancy) as in

$$\hat{\sigma}_0^2 = \Omega / r, \tag{2.22}$$

where the redundancy r is defined as $r := n - \text{rk}\, A$, or $r = n - m$ if A has full column rank m [cf. Schaffrin et al., 2012a].

2.3 The TLS problem within the classical EIV model

Though the name "errors-in-variables model" does not actually appear in the works of Golub cited above, the model presented in Golub and Van Loan [1980], from which their TLS solution was derived, will herein be referred to as *the classical EIV model*. It is given here in its original form in order to compare and contrast it with the *extended EIV model* shown in (2.1); however, some of the symbols are changed from what Golub and Van Loan used in order to be consistent with the notation used in this work.

The system of equations

$$\underbrace{(A - E_A)}_{n\times m}\boldsymbol{\xi} = \boldsymbol{y} - \boldsymbol{e}_y \tag{2.23}$$

is herein called the *classical EIV model*. With the introduction of positive-definite "weighting matrices"

$$D = \mathrm{Diag}[d_1, \ldots, d_n], \qquad d_i > 0, i = 1, \ldots, n, \text{ and} \tag{2.24a}$$

$$T = \mathrm{Diag}[t_1, \ldots, t_{(m+1)}], \quad t_i > 0, i = 1, \ldots, m+1, \tag{2.24b}$$

the *classical TLS problem* is to

$$\underset{[E_A\,|\,\boldsymbol{e}_y]}{\text{minimize}}\left(\left\| D\left[E_A \mid \boldsymbol{e}_y\right] T\right\|_F^2 = \left\|(T \otimes D)\begin{bmatrix}\boldsymbol{e}_A\\ \boldsymbol{e}_y\end{bmatrix}\right\|^2\right). \tag{2.25}$$

Note that the weighting matrices D and T have been excluded intentionally from the model (2.23), as they are not associated necessarily with the dispersion of the random errors of the data, though under certain conditions perhaps they could be. (See comments on page 22.)

The (squared) Frobenius norm of an $n \times m$ matrix A is defined as $\|A\|_F^2 := \mathrm{tr}(A^T A) = \sum_{i=1}^n \sum_{j=1}^m \left|a_{ij}\right|^2$. Using this definition, and the definition $\boldsymbol{e}_A := \mathrm{vec}\, E_A$, the matrix norm is transformed to a vector norm by

$$\begin{aligned}\left\| D\left[E_A \mid \boldsymbol{e}_y\right] T\right\|_F^2 &= \mathrm{tr}\Big(T^T \left[E_A \mid \boldsymbol{e}_y\right]^T D^T D \left[E_A \mid \boldsymbol{e}_y\right] T\Big) = \\ = \mathrm{tr}\Big(\left[E_A \mid \boldsymbol{e}_y\right]^T D^2 \left[E_A \mid \boldsymbol{e}_y\right] T^2\Big) &= \left[\boldsymbol{e}_A^T, \quad \boldsymbol{e}_y^T\right](T^2 \otimes D^2)\begin{bmatrix}\boldsymbol{e}_A\\ \boldsymbol{e}_y\end{bmatrix}.\end{aligned} \tag{2.26}$$

Now let $T = \left[\begin{smallmatrix} T_1 & 0 \\ 0 & t_{m+1}\end{smallmatrix}\right]$, with $T_1 := \mathrm{Diag}[t_1, \ldots, t_m]$. Then the equivalent TLS problem is to

$$\underset{[E_A\,|\,\boldsymbol{e}_y]}{\text{minimize}}\left\| D\left[E_A \mid \boldsymbol{e}_y\right] T\right\|_F^2 = \underset{[E_A\,|\,\boldsymbol{e}_y]}{\text{minimize}}\Big\{\boldsymbol{e}_A^T(T_1^2 \otimes D^2)\boldsymbol{e}_A + \boldsymbol{e}_y^T(t_{m+1}^2 D^2)\boldsymbol{e}_y\Big\}. \tag{2.27}$$

Golub and Van Loan [1980] gave a solution based on the SVD of $D\,[A\,|\,\boldsymbol{y}]\,T$, which is presented below. Once again, it is noted that some symbols used here differ from theirs in order to maintain consistency with this dissertation. Define the $n \times (m+1)$ matrix

$$C := D\left[A \mid \boldsymbol{y}\right] T \tag{2.28}$$

and let

$$\underset{n\times(m+1)}{U^T C V} = \begin{bmatrix}\mathrm{Diag}(\sigma_1, \ldots, \sigma_{m+1})\\ \mathbf{0}\end{bmatrix}, \tag{2.29a}$$

$$U = \left[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\right], V = \left[\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{m+1}\right], \boldsymbol{u}_i \in \mathbb{R}^n, \boldsymbol{v}_i \in \mathbb{R}^{m+1}, \tag{2.29b}$$

$$\sigma_1 \geq \cdots \geq \sigma_m > \sigma_{m+1} > 0 \tag{2.29c}$$

be the SVD of C, with $U^T U = I_n$ and $V^T V = I_{m+1}$. If $\hat{\boldsymbol{\xi}} \in \mathbb{R}^m$ such that

$$C^T C T^{-1} \begin{bmatrix} \hat{\boldsymbol{\xi}} \\ -1 \end{bmatrix} = \sigma_{m+1}^2 T^{-1} \begin{bmatrix} \hat{\boldsymbol{\xi}} \\ -1 \end{bmatrix} \tag{2.30a}$$

$$\Rightarrow \boxed{\hat{\boldsymbol{\xi}} = T_1 \Big(T_1 A^T D^2 A T_1 - \sigma_{m+1}^2 I_m \Big)^{-1} T_1 A^T D^2 \boldsymbol{y}}, \tag{2.30b}$$

then $\hat{\boldsymbol{\xi}}$ solves the TLS problem. Obviously, σ_{m+1}^2 is the smallest eigenvalue of $C^T C$. The solution is unique as long as $\sigma_m > \sigma_{m+1}$ holds. In the case that $\sigma_m = \sigma_{m+1}$, a solution may still exist, but it might not be unique.

In contrast to TLS minimization within the extended EIV model (2.1), the minimization (2.27) does not contain a bilinear term in $\boldsymbol{e}_A$ and $\boldsymbol{e}_y$. Moreover, the matrices $(T_1^2 \otimes D^2)$ and $t_{m+1}^2 D^2$ could not, in general, be interpreted as weight matrices P_{22} and P_{11}, respectively (associated with $\boldsymbol{e}_A$ and $\boldsymbol{e}_y$, respectively), from the extended EIV model.

Another way to visualize the distribution of the weighting matrices is to multiply out a few terms in (2.25), as in

$$D \Big[E_A \mid \boldsymbol{e}_y \Big] T = \left[\begin{array}{ccc|c} d_1 e_{A_{11}} t_1 & \cdots & d_1 e_{A_{1m}} t_m & d_1 e_{y_1} t_{m+1} \\ \cdots & \cdots & \cdots & \cdots \\ d_n e_{A_{n1}} t_1 & \cdots & d_n e_{A_{nm}} t_m & d_n e_{y_n} t_{m+1} \end{array} \right], \tag{2.31}$$

which clearly shows dependence among the columns due to the matrix D and dependence among the rows due to the matrix T. If D is the identity matrix (or a scalar multiple thereof), the data are row-wise iid. Likewise, if T is the identity matrix, the data are column-wise iid. If both D and T are identities (or scalar multiples thereof), the data are all iid.

In light of the above discussion, it is noted that the solution (2.30b) is only viable within the extended EIV model (2.1) for a somewhat restricted category of weight matrices. In fact, the TLS problem within the extended EIV model cannot be solved by the SVD technique unless $P_y = t_{m+1}^2 \cdot D^2$ and $P_A = (T_1^2 \otimes D^2)$ (also implying that $Q_{yA} = 0$), meaning that the weights for each column of the variables in matrix A would have to be scalar multiples of the weights for the observation vector $\boldsymbol{y}$. The restriction of the range of weight matrices admissible in the classical EIV model was surely a primary motivation for the further development of the extended EIV model (2.1), after Schaffrin and Wieser's [2008] extension.

Chapter 3: The EIV Model with Singular Cofactor Matrices

Algorithms 1 and 2 above cannot be used if the matrix Q_1 of (2.10b) turns out to be singular, which would suggest that either Q_y or Q_A is singular or that both are singular (or perhaps rather the case where a nonzero Q_{yA} gives rise to a singularity in Q_1). Therefore, it is desirable to derive an estimator, and corresponding algorithm, that works for a singular matrix Q_1. First, the case of no cross-correlation, i.e., $Q_{yA} = 0$, is considered in §3.1. This simplifies the development somewhat and is a useful reference for problems without cross-correlation between the matrices Q_y and Q_A. Then in §3.2 the case of non-zero Q_{yA} is developed. Note that in this context the term *cross-correlation* refers to correlation between random errors in the observation vector $\boldsymbol{y}$ and random errors in the data matrix A.

3.1 Singular cofactor matrices without cross-correlation between Q_y and Q_A

Let us begin by writing a system of equations comprised of (2.9b), (2.10a), (2.10b), and (2.12), omitting the matrix Q_{yA}, i.e., assuming it is zero:

$$\tilde{\boldsymbol{e}}_A = -Q_A\Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)\hat{\boldsymbol{\xi}}, \tag{3.1a}$$

$$\boldsymbol{y} - A\hat{\boldsymbol{\xi}} = Q_1\hat{\boldsymbol{\lambda}}, \tag{3.1b}$$

$$A^T\hat{\boldsymbol{\lambda}} = \Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)^T\tilde{\boldsymbol{e}}_A, \tag{3.1c}$$

$$\text{where} \quad Q_1 := \Big[Q_y + \Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)^T Q_A\Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)\Big] = Q_1(\hat{\boldsymbol{\xi}}). \tag{3.1d}$$

In matrix form, the system of equations reads

$$\begin{bmatrix} Q_1 & 0 & A \\ 0 & -Q_A & Q_A\Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big) \\ A^T & \Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)^T Q_A & 0 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\lambda}} \\ \hat{\boldsymbol{\gamma}} \\ \hat{\boldsymbol{\xi}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \tag{3.2}$$

with $\tilde{\boldsymbol{e}}_A = \operatorname{vec} \tilde{E}_A = -Q_A\hat{\boldsymbol{\gamma}}$, for $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\xi}} \otimes \hat{\boldsymbol{\lambda}}) = (I_m \otimes \hat{\boldsymbol{\lambda}})\hat{\boldsymbol{\xi}}$.

Now introduce an $m \times m$ symmetric positive-(semi)definite matrix S; then perform the following row operations on the system of equations (3.2):

1. Multiply row three from the left by AS and add to row one.
2. Multiply row three from the left by $Q_A(I_m \otimes \hat{\boldsymbol{\lambda}})S = Q_A(S \otimes \hat{\boldsymbol{\lambda}})$ and add to row two.

These two row operations result in the following modified system of (necessary-condition) equations:

$$\begin{bmatrix} Q_1 + ASA^T & \left(AS \otimes \hat{\boldsymbol{\lambda}}^T\right)Q_A & A \\ Q_A\left(SA^T \otimes \hat{\boldsymbol{\lambda}}\right) & Q_A\left(S \otimes \hat{\boldsymbol{\lambda}}\hat{\boldsymbol{\lambda}}^T\right)Q_A - Q_A & Q_A\left(I_m \otimes \hat{\boldsymbol{\lambda}}\right) \\ A^T & \left(I_m \otimes \hat{\boldsymbol{\lambda}}^T\right)Q_A & 0 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\lambda}} \\ \hat{\gamma} \\ \hat{\boldsymbol{\xi}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \tag{3.3}$$

Now define

$$Q'_A := Q_A\left(S \otimes \hat{\boldsymbol{\lambda}}\hat{\boldsymbol{\lambda}}^T\right)Q_A - Q_A = Q'_A(\hat{\boldsymbol{\lambda}}), \tag{3.4a}$$

$$Q_2 := Q_1 + ASA^T = Q_2(\hat{\boldsymbol{\xi}}), \tag{3.4b}$$

so that the symmetric system of equations (3.3) is simplified somewhat as

$$\begin{bmatrix} Q_2 & \left(AS \otimes \hat{\boldsymbol{\lambda}}^T\right)Q_A & A \\ Q_A\left(SA^T \otimes \hat{\boldsymbol{\lambda}}\right) & Q'_A & Q_A\left(I_m \otimes \hat{\boldsymbol{\lambda}}\right) \\ A^T & \left(I_m \otimes \hat{\boldsymbol{\lambda}}^T\right)Q_A & 0 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\lambda}} \\ \hat{\gamma} \\ \hat{\boldsymbol{\xi}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \tag{3.5}$$

Note that the dimension of the range space of the augmented matrix $[Q_1, AS]$ should satisfy

$$\dim \mathcal{R}\Big[\underset{n\times n}{Q_1}, \quad \underset{n\times m}{AS} \Big] = \operatorname{rk}\Big[Q_1, \quad AS\Big] = n \tag{3.6a}$$

$$\Rightarrow \operatorname{rk} Q_2 = \operatorname{rk}(Q_1 + ASA^T) = \operatorname{rk}\Big[Q_1, \quad AS\Big] \begin{bmatrix} Q_1^- & 0 \\ 0 & S^- \end{bmatrix} \begin{bmatrix} Q_1 \\ SA^T \end{bmatrix} = n, \tag{3.6b}$$

where Q_1^- and S^- denote (nonsingular) generalized inverses of Q_1 and S, respectively. Equations (3.6) ensure that $Q_2 = Q_1 + ASA^T$ is invertible. Moreover, equation (3.6b) satisfies the *Neitzel/Schaffrin criterion* [2012] for uniqueness of the least-squares solution $\hat{\boldsymbol{\xi}}$ in the presence of a singular dispersion matrix. (See Appendix C for further details.)

From row one of (3.5), the estimator for the Lagrange multipliers is found to be

$$\hat{\boldsymbol{\lambda}} = Q_2^{-1} \cdot \Big[(\boldsymbol{y} - A\hat{\boldsymbol{\xi}}) - \left(AS \otimes \hat{\boldsymbol{\lambda}}^T\right)Q_A \cdot \hat{\gamma}\Big]. \tag{3.7}$$

Substituting (3.7) into the vector on the left side of (3.5), yields for the second, resp. third rows of (3.5)

$$\begin{aligned} &\Big[Q_A\left(SA^T \otimes \hat{\boldsymbol{\lambda}}\right)Q_2^{-1}\left(AS \otimes \hat{\boldsymbol{\lambda}}^T\right)Q_A\Big] \cdot \hat{\gamma} - Q'_A \cdot \hat{\gamma} = \\ &= Q_A\left(I_m \otimes \hat{\boldsymbol{\lambda}}\right)\hat{\boldsymbol{\xi}} + Q_A\left(SA^T \otimes \hat{\boldsymbol{\lambda}}\right)Q_2^{-1}\left(\boldsymbol{y} - A\hat{\boldsymbol{\xi}}\right) \end{aligned} \tag{3.8a}$$

and

$$A^T Q_2^{-1}\Big(\boldsymbol{y} - A\hat{\boldsymbol{\xi}}\Big) = -\Big(I_m \otimes \hat{\boldsymbol{\lambda}}^T\Big) Q_A \cdot \hat{\boldsymbol{\gamma}} + A^T Q_2^{-1}\Big(AS \otimes \hat{\boldsymbol{\lambda}}^T\Big) Q_A \cdot \hat{\boldsymbol{\gamma}} \tag{3.8b}$$

or, combining the two equations in matrix form,

$$\left[\begin{array}{c|c} A^T Q_2^{-1} A & -\Big[\Big(I_m \otimes \hat{\boldsymbol{\lambda}}^T\Big) - \Big(A^T Q_2^{-1} AS \otimes \hat{\boldsymbol{\lambda}}^T\Big)\Big] Q_A \\ \hline -Q_A\Big[\Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big) - \Big(SA^T Q_2^{-1} A \otimes \hat{\boldsymbol{\lambda}}\Big)\Big] & Q_A\Big(SA^T Q_2^{-1} AS \otimes \hat{\boldsymbol{\lambda}}\hat{\boldsymbol{\lambda}}^T\Big) Q_A - Q_A' \end{array}\right] \times$$

$$\times \begin{bmatrix} \hat{\boldsymbol{\xi}} \\ \hat{\boldsymbol{\gamma}} \end{bmatrix} = \begin{bmatrix} A^T Q_2^{-1} \boldsymbol{y} \\ Q_A\Big(SA^T Q_2^{-1} \otimes \hat{\boldsymbol{\lambda}}\Big)\boldsymbol{y} \end{bmatrix}. \tag{3.9}$$

From (3.1a), the relation $\hat{\boldsymbol{\gamma}} = (I_m \otimes \hat{\boldsymbol{\lambda}}) \cdot \hat{\boldsymbol{\xi}}$ holds, which upon substituting into the first row of (3.9) leads to the solution

$$\begin{aligned} A^T Q_2^{-1}\Big(\boldsymbol{y} - A\hat{\boldsymbol{\xi}}\Big) &= \\ = -\Big[\Big(I_m \otimes \hat{\boldsymbol{\lambda}}^T\Big) - \Big(A^T Q_2^{-1} AS \otimes \hat{\boldsymbol{\lambda}}^T\Big)\Big] Q_A \cdot \hat{\boldsymbol{\gamma}} &= \\ = \Big(A^T Q_2^{-1} AS - I_m\Big) \cdot \Big[\Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)^T Q_A \Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)\Big] \cdot \hat{\boldsymbol{\xi}}&. \end{aligned} \tag{3.10}$$

Now define R_2 as

$$R_2 := \Big(A^T Q_2^{-1} AS - I_m\Big) \cdot \Big[\Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)^T Q_A \Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)\Big] = R_2(\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\lambda}}). \tag{3.11}$$

Then the solution $\hat{\boldsymbol{\xi}}$ is written as

$$\boxed{\hat{\boldsymbol{\xi}} = \Big(A^T Q_2^{-1} A + R_2\Big)^{-1} A^T Q_2^{-1} \boldsymbol{y}}. \tag{3.12}$$

Using (3.7) and the relation $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\xi}} \otimes I_n) \cdot \hat{\boldsymbol{\lambda}}$, the estimator for the unknown vector of Lagrange multipliers $\boldsymbol{\lambda}$ is derived as

$$Q_2 \hat{\boldsymbol{\lambda}} = (\boldsymbol{y} - A\hat{\boldsymbol{\xi}}) - \Big(AS \otimes \hat{\boldsymbol{\lambda}}^T\Big) Q_A \Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big) \cdot \hat{\boldsymbol{\lambda}} \tag{3.13a}$$

$$\Rightarrow \boxed{\hat{\boldsymbol{\lambda}} = \Big(Q_2 + AS \cdot \Big[\Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)^T Q_A \Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)\Big]\Big)^{-1} (\boldsymbol{y} - A\hat{\boldsymbol{\xi}})}. \tag{3.13b}$$

3.1.1 Alternative solution using $(A - \tilde{E}_A)$

Note that the formula (3.12) is nonlinear in the unknown parameters, and thus must be computed by iteration, though the matrix A remains constant through the iteration steps. In the following, an alternative to (3.12) is developed that allows the matrix difference $A - \tilde{E}_A$ to be used instead of just A, as long as $\operatorname{rk}(A - \tilde{E}_A) = \operatorname{rk} A$.

The system (3.2) can be rewritten as

$$Q_1\hat{\boldsymbol{\lambda}} = \boldsymbol{y} - A\hat{\boldsymbol{\xi}}, \tag{3.14a}$$

$$\tilde{\boldsymbol{e}}_A = -Q_A\Big(\hat{\boldsymbol{\xi}} \otimes \hat{\boldsymbol{\lambda}}\Big), \tag{3.14b}$$

$$\Big(A - \tilde{E}_A\Big)^T\hat{\boldsymbol{\lambda}} = \mathbf{0}. \tag{3.14c}$$

Equation (3.14c) can be premultiplied by $(A - \tilde{E}_A)S$ to arrive at

$$\Big(A - \tilde{E}_A\Big)S\Big(A - \tilde{E}_A\Big)^T\hat{\boldsymbol{\lambda}} = 0, \tag{3.15}$$

which combined with (3.14a) gives

$$Q_3\hat{\boldsymbol{\lambda}} = \Big[Q_1 + \Big(A - \tilde{E}_A\Big)S\Big(A - \tilde{E}_A\Big)^T\Big]\hat{\boldsymbol{\lambda}} = \boldsymbol{y} - A\hat{\boldsymbol{\xi}}, \tag{3.16}$$

where Q_3 is defined by

$$Q_3 := \Big[Q_1 + \Big(A - \tilde{E}_A\Big)S\Big(A - \tilde{E}_A\Big)^T\Big] = Q_3(\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\lambda}}). \tag{3.17}$$

Its inverse exists whenever $\operatorname{rk}[Q_1, (A - \tilde{E}_A)S] = n$, in analogy to (3.6a).

This leads to the equivalent system of equations

$$\begin{bmatrix} Q_3 & A - \tilde{E}_A \\ \Big(A - \tilde{E}_A\Big)^T & 0 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\lambda}} \\ \hat{\boldsymbol{\xi}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} - \tilde{E}_A\hat{\boldsymbol{\xi}} \\ \mathbf{0} \end{bmatrix}, \tag{3.18a}$$

$$\text{with } \operatorname{vec}\tilde{E}_A = \tilde{\boldsymbol{e}}_A = -Q_A\Big(\hat{\boldsymbol{\xi}} \otimes \hat{\boldsymbol{\lambda}}\Big). \tag{3.18b}$$

Then the vector $\hat{\boldsymbol{\lambda}}$ can be eliminated from the first row of (3.18a) via premultiplication by $(A - \tilde{E}_A)^T Q_3^{-1}$, which leads to the solution

$$\boxed{\hat{\boldsymbol{\xi}} = \Big[\Big(A - \tilde{E}_A\Big)^T Q_3^{-1}\Big(A - \tilde{E}_A\Big)\Big]^{-1}\Big(A - \tilde{E}_A\Big)^T Q_3^{-1}\Big(\boldsymbol{y} - \tilde{E}_A\hat{\boldsymbol{\xi}}\Big)}, \tag{3.19}$$

provided Q_3 remains nonsingular.

The "normal-equation matrix" to invert on the right side of (3.19) is symmetric and positive-definite, if $\operatorname{rk}(A - \tilde{E}_A) = \operatorname{rk} A = m$, and thus can be inverted by Cholesky factorization, though the equation must be solved by iteration since the parameter vector $\hat{\boldsymbol{\xi}}$ appears on both sides of the equation, and $\hat{\boldsymbol{\xi}}$ depends on the prediction $\tilde{E}_A$ as well. Before an algorithm for numerical computation of (3.19) is presented, the extension to the case of both a singular matrix Q_1 and a non-zero matrix Q_{yA} is developed.

3.2 Singular cofactor matrices with cross-correlation

To account for cross-correlation between Q_y and Q_A, the system of equations (3.1) is modified, in accordance with (2.16), as

$$\tilde{\boldsymbol{e}}_A = \Big[Q_{Ay} - Q_A\Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)\Big]\hat{\boldsymbol{\lambda}}, \tag{3.20a}$$

$$\boldsymbol{y} - A\hat{\boldsymbol{\xi}} = Q_1\hat{\boldsymbol{\lambda}}, \tag{3.20b}$$

$$A^T\hat{\boldsymbol{\lambda}} = \Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)^T\tilde{\boldsymbol{e}}_A, \tag{3.20c}$$

where

$$Q_1 := \Big[Q_y - Q_{yA}\Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big) - \Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)^T Q_{Ay} + \Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)^T Q_A\Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)\Big] = Q_1(\hat{\boldsymbol{\xi}}), \tag{3.20d}$$

which, analogous to (3.2), can be expressed in matrix form as

$$\begin{bmatrix} Q_1 & 0 & A \\ 0 & -Q_A & Q_A\Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big) \\ A^T & \Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)^T Q_A & 0 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\lambda}} \\ \hat{\boldsymbol{\gamma}} \\ \hat{\boldsymbol{\xi}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{0} \\ \Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)^T Q_{Ay}\hat{\boldsymbol{\lambda}} \end{bmatrix}, \tag{3.21}$$

with $\tilde{\boldsymbol{e}}_A = \operatorname{vec}\tilde{E}_A = -Q_A\hat{\boldsymbol{\gamma}}$, for $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\xi}} \otimes \hat{\boldsymbol{\lambda}}) = (I_m \otimes \hat{\boldsymbol{\lambda}})\hat{\boldsymbol{\xi}}$. By eliminating $\hat{\boldsymbol{\lambda}}$ from the right side, while maintaining symmetry on the left side, the preceding system of equations is modified as

$$\begin{aligned} &\begin{bmatrix} Q_1 & 0 & A - Q_{yA}\Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big) \\ 0 & -Q_A & Q_A\Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big) \\ A^T - \Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)^T Q_{Ay} & \Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)^T Q_A & 0 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\lambda}} \\ \hat{\boldsymbol{\gamma}} \\ \hat{\boldsymbol{\xi}} \end{bmatrix} = \\ &\qquad = \begin{bmatrix} \boldsymbol{y} - Q_{yA}\Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big) \cdot \hat{\boldsymbol{\xi}} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}. \end{aligned} \tag{3.22}$$

For sake of compactness, define the the auxiliary matrix

$$Q_2' := Q_1 + [A - Q_{yA}(I_m \otimes \hat{\boldsymbol{\lambda}})]S[A^T - (I_m \otimes \hat{\boldsymbol{\lambda}})^T Q_{Ay}], \tag{3.23}$$

and then perform the following row operations:

1. Multiply row three from the left by $[A - Q_{yA}(I_m \otimes \hat{\boldsymbol{\lambda}})]S$ and add to row one.
2. Multiply row three from the left by $Q_A(I_m \otimes \hat{\boldsymbol{\lambda}})S = Q_A(S \otimes \hat{\boldsymbol{\lambda}})$ and add to row two.

This leads to the system of equations

$$\left[\begin{array}{c|c|c} Q_2' & \begin{array}{c}(AS\otimes\hat{\boldsymbol{\lambda}}^T)Q_A-\\ -Q_{yA}(S\otimes\hat{\boldsymbol{\lambda}}\hat{\boldsymbol{\lambda}}^T)Q_A\end{array} & A-Q_{yA}(I_m\otimes\hat{\boldsymbol{\lambda}}) \\ \hline \begin{array}{c}Q_A(SA^T\otimes\hat{\boldsymbol{\lambda}})-\\ -Q_A(S\otimes\hat{\boldsymbol{\lambda}}\hat{\boldsymbol{\lambda}}^T)Q_{Ay}\end{array} & Q_A' & Q_A(I_m\otimes\hat{\boldsymbol{\lambda}}) \\ \hline A^T-(I_m\otimes\hat{\boldsymbol{\lambda}})^TQ_{Ay} & (I_m\otimes\hat{\boldsymbol{\lambda}})^TQ_A & 0 \end{array}\right]\begin{bmatrix}\hat{\boldsymbol{\lambda}}\\ \hat{\gamma}\\ \hat{\boldsymbol{\xi}}\end{bmatrix}= \tag{3.24}$$
$$=\begin{bmatrix}\boldsymbol{y}-Q_{yA}\Big(I_m\otimes\hat{\boldsymbol{\lambda}}\Big)\cdot\hat{\boldsymbol{\xi}}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix},$$

where the matrix Q_A' is defined in (3.4a).

The first row of (3.24) yields

$$\hat{\boldsymbol{\lambda}}=(Q_2')^{-1}\cdot\Big[(\boldsymbol{y}-A\hat{\boldsymbol{\xi}})-\Big[(AS\otimes\hat{\boldsymbol{\lambda}}^T)-Q_{yA}(S\otimes\hat{\boldsymbol{\lambda}}\hat{\boldsymbol{\lambda}}^T)\Big]Q_A\cdot\hat{\gamma}\Big]. \tag{3.25}$$

Then substituting this expression for $\hat{\boldsymbol{\lambda}}$ into the second, resp. third rows of (3.24) yields

$$\begin{aligned}&Q_A\Big[(SA^T\otimes\hat{\boldsymbol{\lambda}})-(S\otimes\hat{\boldsymbol{\lambda}}\hat{\boldsymbol{\lambda}}^T)Q_{Ay}\Big](Q_2')^{-1}\times\\ &\times\Big[(\boldsymbol{y}-A\hat{\boldsymbol{\xi}})-\Big[(AS\otimes\hat{\boldsymbol{\lambda}}^T)-Q_{yA}(S\otimes\hat{\boldsymbol{\lambda}}\hat{\boldsymbol{\lambda}}^T)\Big]Q_A\cdot\hat{\gamma}\Big]+Q_A'\cdot\hat{\gamma}+Q_A(I_m\otimes\hat{\boldsymbol{\lambda}})\cdot\hat{\boldsymbol{\xi}}=\mathbf{0}\end{aligned} \tag{3.26a}$$

and

$$\begin{aligned}&\Big[A^T-(I_m\otimes\hat{\boldsymbol{\lambda}})^TQ_{Ay}\Big](Q_2')^{-1}(\boldsymbol{y}-A\hat{\boldsymbol{\xi}})=\\ &=\Big[A^T-(I_m\otimes\hat{\boldsymbol{\lambda}})^TQ_{Ay}\Big](Q_2')^{-1}\Big[(AS\otimes\hat{\boldsymbol{\lambda}}^T)-Q_{yA}(S\otimes\hat{\boldsymbol{\lambda}}\hat{\boldsymbol{\lambda}}^T)\Big]Q_A\cdot\hat{\gamma}-(I_m\otimes\hat{\boldsymbol{\lambda}})^TQ_A\cdot\hat{\gamma}.\end{aligned} \tag{3.26b}$$

Equations (3.26) can be expressed in matrix form as

$$\left[\begin{array}{c|c} \Big[A^T-(I_m\otimes\hat{\boldsymbol{\lambda}})^TQ_{Ay}\Big](Q_2')^{-1}A & \begin{array}{c}\Big[A^T-(I_m\otimes\hat{\boldsymbol{\lambda}})^TQ_{Ay}\Big](Q_2')^{-1}\cdot\\ \cdot\Big[A-Q_{yA}(I_m\otimes\hat{\boldsymbol{\lambda}})\Big](S\otimes\hat{\boldsymbol{\lambda}})^TQ_A-(I_m\otimes\hat{\boldsymbol{\lambda}})Q_A\end{array} \\ \hline \text{Symmetric} & \begin{array}{c}Q_A(S\otimes\hat{\boldsymbol{\lambda}})\Big[A^T-(I_m\otimes\hat{\boldsymbol{\lambda}})^TQ_{Ay}\Big](Q_2')^{-1}\cdot\\ \cdot\Big[A-Q_{yA}(I_m\otimes\hat{\boldsymbol{\lambda}})\Big](S\otimes\hat{\boldsymbol{\lambda}})^TQ_A-Q_A'\end{array}\end{array}\right]\times$$
$$\times\begin{bmatrix}\hat{\boldsymbol{\xi}}\\ \hat{\gamma}\end{bmatrix}=\begin{bmatrix}\Big[A^T-(I_m\otimes\hat{\boldsymbol{\lambda}})^TQ_{Ay}\Big](Q_2')^{-1}\boldsymbol{y}\\ Q_A(S\otimes\hat{\boldsymbol{\lambda}})\Big[A^T-(I_m\otimes\hat{\boldsymbol{\lambda}})^TQ_{Ay}\Big](Q_2')^{-1}\cdot(\boldsymbol{y}-Q_{yA}(I_m\otimes\hat{\boldsymbol{\lambda}})\hat{\boldsymbol{\xi}})\end{bmatrix}. \tag{3.27}$$

Using the relation $\hat{\boldsymbol{\gamma}} = (I_m \otimes \hat{\boldsymbol{\lambda}})\hat{\boldsymbol{\xi}}$ and manipulating the first row of (3.27), leads to an expression for the parameter estimator $\hat{\boldsymbol{\xi}}$ as follows:

$$\begin{aligned}
&\Big[A^T - (I_m \otimes \hat{\boldsymbol{\lambda}})^T Q_{Ay}\Big](Q_2')^{-1}(\boldsymbol{y} - A\hat{\boldsymbol{\xi}}) = \\
&= \Big\{\Big[A^T - (I_m \otimes \hat{\boldsymbol{\lambda}})^T Q_{Ay}\Big](Q_2')^{-1}\Big[A - Q_{yA}(I_m \otimes \hat{\boldsymbol{\lambda}})\Big](S \otimes \hat{\boldsymbol{\lambda}})^T - (I_m \otimes \hat{\boldsymbol{\lambda}})^T\Big\}Q_A \cdot \hat{\boldsymbol{\gamma}} = \\
&= \Big\{\Big[A^T - (I_m \otimes \hat{\boldsymbol{\lambda}})^T Q_{Ay}\Big](Q_2')^{-1}\Big[A - Q_{yA}(I_m \otimes \hat{\boldsymbol{\lambda}})\Big]S - I_m\Big\}\times \\
&\times\Big[(I_m \otimes \hat{\boldsymbol{\lambda}})^T Q_A (I_m \otimes \hat{\boldsymbol{\lambda}})\Big] \cdot \hat{\boldsymbol{\xi}} = R_2'\hat{\boldsymbol{\xi}},
\end{aligned}$$

with the auxiliary matrix R_2' defined as

$$R_2' := \Big\{\Big[A^T - (I_m \otimes \hat{\boldsymbol{\lambda}})^T Q_{Ay}\Big](Q_2')^{-1}\Big[A - Q_{yA}(I_m \otimes \hat{\boldsymbol{\lambda}})\Big]S - I_m\Big\} \cdot \Big[(I_m \otimes \hat{\boldsymbol{\lambda}})^T Q_A (I_m \otimes \hat{\boldsymbol{\lambda}})\Big], \tag{3.28}$$

finally resulting in

$$\boxed{\hat{\boldsymbol{\xi}} = \Big\{\Big[A^T - (I_m \otimes \hat{\boldsymbol{\lambda}})^T Q_{Ay}\Big](Q_2')^{-1}A + R_2'\Big\}^{-1} \cdot \Big[A^T - (I_m \otimes \hat{\boldsymbol{\lambda}})^T Q_{Ay}\Big](Q_2')^{-1}\boldsymbol{y}}. \tag{3.29}$$

Substituting $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\xi}} \otimes I_n)\hat{\boldsymbol{\lambda}}$ into the vector on the left side of (3.24) yields, for the first row of (3.24),

$$Q_2'\hat{\boldsymbol{\lambda}} = (\boldsymbol{y} - A\hat{\boldsymbol{\xi}}) - \Big[A - Q_{yA}\Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)\Big]S\Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)^T Q_A\Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big) \cdot \hat{\boldsymbol{\lambda}}, \tag{3.30}$$

leading to

$$\boxed{\hat{\boldsymbol{\lambda}} = \Big\{Q_2' + \Big[A - Q_{yA}\Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)\Big] \cdot S \cdot \Big[\Big(I_m \otimes \hat{\boldsymbol{\lambda}}\Big)^T Q_A\Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)\Big]\Big\}^{-1}(\boldsymbol{y} - A\hat{\boldsymbol{\xi}})} \tag{3.31}$$

as an expression for the estimator for the vector of Lagrange multipliers.

As expected, when $Q_{Ay} = 0$, the estimators derived in this section become equivalent to those of §3.1, where it was assumed that there was no cross-correlation between the random errors in $\boldsymbol{y}$ and A. Table 3.1 summarizes the respective equation numbers associated with the cases where $Q_{Ay} = 0$ and $Q_{Ay} \neq 0$.

It is important to note the role of the matrix A in both cases shown in Table 3.1. In the first case (when $Q_{Ay} = 0$), the matrix A (or its transpose) appears without having a matrix directly subtracted from it. In contrast, when $Q_{Ay} \neq 0$ the term $A - Q_{yA}(I_m \otimes \hat{\boldsymbol{\lambda}})$ appears in the equations listed in the last column of Table 3.1, rather than A alone. The question then naturally arises as to whether such a reduction of the matrix A might increase the condition numbers of the matrices that must be inverted in equations (3.29) and (3.31) for $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\lambda}}$, respectively. If the condition numbers were to increase significantly, these formulas may become numerically unstable. Such a risk is perhaps much lower in the case where the matrix difference $A - \tilde{E}_A$ appears, since the expectation of $\tilde{E}_A$ is zero.

Table 3.1: Summary of equation numbers for the cases of $Q_{Ay} = 0$ and $Q_{Ay} \neq 0$

Variables	When $Q_{Ay} = 0$	When $Q_{Ay} \neq 0$
Q_2 or Q_2'	(3.4b)	(3.23)
R_2 or R_2'	(3.11)	(3.28)
$\hat{\boldsymbol{\xi}}$	(3.12)	(3.29)
$\hat{\boldsymbol{\lambda}}$	(3.13b)	(3.31)

3.2.1 Alternative solution using $(A - \tilde{E}_A)$

In light of the preceding discussion, we desire to find a solution analogous to (3.19), which incorporates the corrected data matrix $A - \tilde{E}_A$. The system (3.18) is easily extended to the case of a nonzero matrix Q_{Ay} by simply using (3.20a) for $\tilde{\boldsymbol{e}}_A$, rather than the form shown in (3.18b), which is done in the following:

$$\begin{bmatrix} Q_3 & A - \tilde{E}_A \\ \left(A - \tilde{E}_A\right)^T & 0 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\lambda}} \\ \hat{\boldsymbol{\xi}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} - \tilde{E}_A \hat{\boldsymbol{\xi}} \\ \boldsymbol{0} \end{bmatrix}, \tag{3.32a}$$

$$\text{with } \operatorname{vec} \tilde{E}_A = \tilde{\boldsymbol{e}}_A = \left[Q_{Ay} - Q_A\left(\hat{\boldsymbol{\xi}} \otimes I_n\right)\right]\hat{\boldsymbol{\lambda}}. \tag{3.32b}$$

Then the solution for the parameter estimator $\hat{\boldsymbol{\xi}}$ takes the same form as (3.19), which is repeated here for completeness:

$$\boxed{\hat{\boldsymbol{\xi}} = \left[\left(A - \tilde{E}_A\right)^T Q_3^{-1} \left(A - \tilde{E}_A\right)\right]^{-1} \left(A - \tilde{E}_A\right)^T Q_3^{-1} \left(\boldsymbol{y} - \tilde{E}_A \hat{\boldsymbol{\xi}}\right)}. \tag{3.33}$$

Finally, Algorithm 3 is presented for computation of the estimated parameter vector $\hat{\boldsymbol{\xi}}$ shown in (3.33), which handles both cross-correlation (i.e., $Q_{yA} = Q_{Ay}^T \neq 0$) and singular cofactor matrices (more specifically, singular Q_1).

Algorithm 3 For the WTLS solution within the EIV model having rank-deficient cofactor matrices (with singular matrix Q_1)

Step 1: Compute an initial solution

$$\hat{\boldsymbol{\xi}}^{(0)} := N^{+}\boldsymbol{c} \quad \text{for} \quad \Big[N, \quad \boldsymbol{c}\Big] := A^T Q_y^{+} \Big[A, \quad \boldsymbol{y}\Big], \text{ and assign } \tilde{E}_A^{(0)} = 0, \tag{3.34a}$$

where N^{+} represents the pseudo-inverse (or Moore-Penrose inverse) of N.

Step 2:
repeat For $i \in \mathbb{N}$, compute

$$\begin{aligned} Q_1^{(i)} &:= \Big[Q_y - Q_{yA}\Big(\hat{\boldsymbol{\xi}}^{(i-1)} \otimes I_n\Big) - \Big(\hat{\boldsymbol{\xi}}^{(i-1)} \otimes I_n\Big)^T Q_{Ay} + \\ &\qquad + \Big(\hat{\boldsymbol{\xi}}^{(i-1)} \otimes I_n\Big)^T Q_A\Big(\hat{\boldsymbol{\xi}}^{(i-1)} \otimes I_n\Big)\Big] \end{aligned} \tag{3.34b}$$

$$Q_3^{(i)} := Q_1^{(i)} + \Big(A - \tilde{E}_A^{(i-1)}\Big) S \Big(A - \tilde{E}_A^{(i-1)}\Big)^T \tag{3.34c}$$

$$\begin{aligned} \hat{\boldsymbol{\xi}}^{(i)} &= \Big[\Big(A - \tilde{E}_A^{(i-1)}\Big)^T \Big(Q_3^{(i)}\Big)^{-1} \Big(A - \tilde{E}_A^{(i-1)}\Big)\Big]^{-1} \times \\ &\qquad \times \Big[\Big(A - \tilde{E}_A^{(i-1)}\Big)^T \Big(Q_3^{(i)}\Big)^{-1} \Big(\boldsymbol{y} - \tilde{E}_A^{(i-1)} \hat{\boldsymbol{\xi}}^{(i-1)}\Big)\Big] \end{aligned} \tag{3.34d}$$

$$\hat{\boldsymbol{\lambda}}^{(i)} = \Big(Q_3^{(i)}\Big)^{-1} \Big[\Big(\boldsymbol{y} - \tilde{E}_A^{(i-1)} \hat{\boldsymbol{\xi}}^{(i-1)}\Big) - \Big(A - \tilde{E}_A^{(i-1)}\Big)\hat{\boldsymbol{\xi}}^{(i)}\Big] \tag{3.34e}$$

$$\tilde{\boldsymbol{e}}_A^{(i)} = \Big[Q_{Ay} - Q_A\Big(\hat{\boldsymbol{\xi}}^{(i)} \otimes I_n\Big)\Big]\hat{\boldsymbol{\lambda}}^{(i)} \quad \text{and} \quad \tilde{E}_A^{(i)} = \operatorname{Invec} \tilde{\boldsymbol{e}}_A^{(i)} \tag{3.34f}$$

until

$$\Big\|\hat{\boldsymbol{\xi}}^{(i)} - \hat{\boldsymbol{\xi}}^{(i-1)}\Big\| < \delta \tag{3.34g}$$

for a chosen threshold δ.

Table 3.2 summarizes the WTLS algorithms presented thus far:

Table 3.2: Summary of WTLS algorithms

Eq. No.	Algorithm	Comment
(2.14)	Algorithm 1	Same as eq. 4.25 ("Algorithm 2") of Fang [2011]. [1]
(2.18)	Algorithm 2	Same as eq. 4.26 ("Algorithm 3") of Fang [2011].
(3.33)	Algorithm 3	Handles both cross-correlation ($Q_{yA} \neq 0$) and singular cofactor matrices (singular Q_1); new in this contribution.

[1]See also Mahboub [2012].

Chapter 4: Analytical Comparisons Between the Least-Squares Solutions within the EIV and Gauss-Helmert Models

In this chapter the Gauss-Helmert model (GHM) is reviewed, and some relations between solutions within the GHM and the errors-in-variables (EIV) model are presented. First, the EIV model (2.1) is rewritten in (4.1) with a slight variation in notation, following Schaffrin and Snow [2010], so as to clearly distinguish between symbols used for the EIV and Gauss-Helmert models.

$$\boldsymbol{y} = \underbrace{(X - E_X)}_{k\times m} \cdot \boldsymbol{\beta}_\mu + \boldsymbol{e}_y, \quad \begin{bmatrix} \boldsymbol{e}_y \\ \operatorname{vec} E_X \end{bmatrix} \sim \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \sigma_0^2 \begin{bmatrix} \underset{k\times k}{Q_y} & Q_{yX} \\ Q_{Xy} & \underset{km\times km}{Q_X} \end{bmatrix} = \sigma_0^2 Q \right) \tag{4.1}$$

The meanings of the terms used in (4.1) should be clear from the descriptions immediately following (2.1). By introducing new symbols

$$\boldsymbol{Y} := \operatorname{vec}\left[\boldsymbol{y}, X\right], \quad \boldsymbol{e} := \operatorname{vec}\left[\boldsymbol{e}_y, E_X\right], \quad \boldsymbol{\Xi} := \boldsymbol{\beta}_\mu, \tag{4.2}$$

the EIV model (4.1) may be written as

$$\boldsymbol{b}(\underbrace{\boldsymbol{Y} - \boldsymbol{e}}_{n\times 1}, \underbrace{\boldsymbol{\Xi}}_{m\times 1}) := \left[I_k, \quad -(\boldsymbol{\Xi} \otimes I_k)^T\right] \begin{bmatrix} \boldsymbol{y} - \boldsymbol{e}_y \\ \operatorname{vec}(X - E_X) \end{bmatrix} = \mathbf{0}, \quad \boldsymbol{e} \sim (\mathbf{0}, \sigma_0^2 \underbrace{P^{-1}}_{n\times n}), \tag{4.3}$$

where $\boldsymbol{b} : \mathbb{R}^{m+n} \Rightarrow \mathbb{R}^{m+r}$ denotes a given multivariate *nonlinear* function. Furthermore, $n = k(m + 1)$ is the total number of observations (assuming no fixed columns of X), and P is the $n \times n$ weight matrix such that $P = Q^{-1}$, assuming Q has full rank. Finally, r stands for the redundancy of the model, with $r \leq n - m$. In this form, (4.3) represents the *nonlinear Gauss-Helmert model*, following Helmert [1907]. Obviously, then, the EIV model can be classified as a nonlinear GHM.

By introducing the "true" $n \times 1$ vector of observables

$$\boldsymbol{\mu} := \boldsymbol{Y} - \boldsymbol{e} = E\{\boldsymbol{Y}\}, \tag{4.4}$$

the least-squares objective for model (4.3) is defined by

$$\boldsymbol{e}^T P \boldsymbol{e} = (\boldsymbol{Y} - \boldsymbol{\mu})^T P (\boldsymbol{Y} - \boldsymbol{\mu}) = \min \text{ subject to } \boldsymbol{b}(\boldsymbol{\mu}, \boldsymbol{\Xi}) = \boldsymbol{0}. \tag{4.5}$$

The solution resulting from (4.5), though it minimizes the weighted squared sum of all random errors, will herein be called the least-squares solution (LESS) within the GHM to distinguish it from the WTLS solutions within the EIV model derived in the previous chapters.

4.1 Iterative linearization of the Gauss-Helmert model

The solution resulting from (4.5) can be determined by iterative linearization of the model (4.3) with subsequent standard least-squares approximation. The linearization scheme of Schaffrin and Snow [2010] is shown in Algorithm 4, where subscripts are used for the expansion point of the iterative linearization, and superscripts are used with the coefficient matrices and residual vector.

Algorithm 4 For the iterative linearization of the GHM, and the associated LESS

repeat for $j \in \mathbb{N}_0$

Step 1: Use the truncated Taylor series about $(\boldsymbol{\mu}_j, \Xi_j)$, namely

$$\left[\frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\mu}^T}\Big|_{\boldsymbol{\mu}_j, \Xi_j}, \quad \frac{\partial \boldsymbol{b}}{\partial \Xi^T}\Big|_{\boldsymbol{\mu}_j, \Xi_j} \right] \cdot \begin{bmatrix} \boldsymbol{\mu} - \boldsymbol{\mu}_j \\ \Xi - \Xi_j \end{bmatrix} + \boldsymbol{b}\Big(\boldsymbol{\mu}_j, \Xi_j\Big) = \mathbf{0}, \tag{4.6a}$$

introduce an initial approximation Ξ_0 for Ξ_j, and replace $\boldsymbol{\mu}$ with $\boldsymbol{Y} - \boldsymbol{e}$, in accordance with (4.4), to introduce

$$\underset{m \times 1}{\boldsymbol{\xi}_{j+1}} := \Xi - \Xi_j, \quad \underset{(m+r)\times m}{A^{(j)}} := -\frac{\partial \boldsymbol{b}}{\partial \Xi^T}\Big|_{\boldsymbol{\mu}_j, \Xi_j}, \quad \underset{(m+r)\times n}{B^{(j)}} := \frac{\partial \boldsymbol{b}}{\partial \boldsymbol{\mu}^T}\Big|_{\boldsymbol{\mu}_j, \Xi_j}, \tag{4.6b}$$

$$\underset{(m+r)\times 1}{\boldsymbol{w}_j} := \boldsymbol{b}(\boldsymbol{\mu}_j, \Xi_j) + B^{(j)} \cdot (\boldsymbol{Y} - \boldsymbol{\mu}_j) \approx \boldsymbol{b}(\boldsymbol{Y}, \Xi_j), \ \boldsymbol{\mu}_0 := \boldsymbol{Y} - \underset{\sim}{\mathbf{0}}, \tag{4.6c}$$

and to form the linearized GHM:

$$\boldsymbol{w}_j = A^{(j)} \boldsymbol{\xi}_{j+1} + B^{(j)} \boldsymbol{e}, \quad \boldsymbol{e} \sim (\mathbf{0}, \sigma_0^2 P^{-1}). \tag{4.6d}$$

▷ Here it is noted that matrix $A^{(j)}$ has full column rank, and matrix $B^{(j)}$ has full row rank. It is also noted that $\underset{\sim}{\mathbf{0}}$ denotes a "random zero vector" (or vector of "pseudo-observations") of suitable size, in accordance with the notion in Harville [1986].

Step 2: Produce the $(j+1)$-th least-squares solution for (4.6d), following Koch [1999], e.g., namely:

$$\hat{\boldsymbol{\xi}}_{j+1} = \left[\left(A^{(j)}\right)^T \left[B^{(j)} P^{-1} \left(B^{(j)}\right)^T \right]^{-1} A^{(j)} \right]^{-1} \left(A^{(j)}\right)^T \left[B^{(j)} P^{-1} \left(B^{(j)}\right)^T \right]^{-1} \boldsymbol{w}_j, \tag{4.6e}$$

$$\tilde{\boldsymbol{e}}^{(j+1)} = P^{-1} \left(B^{(j)}\right)^T \left[B^{(j)} P^{-1} \left(B^{(j)}\right)^T \right]^{-1} \left(\boldsymbol{w}_j - A^{(j)} \hat{\boldsymbol{\xi}}_{j+1} \right). \tag{4.6f}$$

Step 3: Obtain new approximate values (non-random) through:

$$\Xi_{j+1} := \hat{\Xi}^{(j+1)} - \underset{\sim}{\mathbf{0}} = \Xi_j + \hat{\boldsymbol{\xi}}_{j+1} - \underset{\sim}{\mathbf{0}}, \tag{4.6g}$$

$$\boldsymbol{\mu}_{j+1} := \hat{\boldsymbol{\mu}}^{(j+1)} - \underset{\sim}{\mathbf{0}} = \boldsymbol{Y} - \tilde{\boldsymbol{e}}^{(j+1)} - \underset{\sim}{\mathbf{0}}. \tag{4.6h}$$

▷ Note that the use of the random zero vector $\underset{\sim}{\mathbf{0}}$ means that the j-th (approximate) estimates are stripped of their randomness while retaining their numerical values.

until

$$\left\| \hat{\boldsymbol{\xi}}_{j+1} \right\| < \delta \text{ and } \left\| \tilde{\boldsymbol{e}}^{(j+1)} - \tilde{\boldsymbol{e}}^{(j)} \right\| < \epsilon \tag{4.6i}$$

for chosen thresholds δ and ϵ.

After the condition (4.6i) of Algorithm 4 is fulfilled for $\hat{\boldsymbol{\xi}}_{j+1}$ and $\tilde{\boldsymbol{e}}^{(j+1)}$, the respective *Mean Squared Error (MSE)* and *dispersion matrices* are obtained in first-order approximation via:

$$D\{\hat{\boldsymbol{\Xi}} := \boldsymbol{\Xi}_j + \hat{\boldsymbol{\xi}}_{j+1}\} = \sigma_0^2 \Big[\big(A^{(j)}\big)^T \Big[B^{(j)} P^{-1} \big(B^{(j)}\big)^T \Big]^{-1} A^{(j)} \Big]^{-1} \approx \text{MSE}\{\hat{\boldsymbol{\Xi}}\} \tag{4.7a}$$

and

$$\text{MSE}\{\tilde{\boldsymbol{e}}\} = D\{\tilde{\boldsymbol{e}} - \boldsymbol{e}\} = D\{\boldsymbol{e}\} - D\{\tilde{\boldsymbol{e}}\}, \tag{4.7b}$$

with

$$\begin{aligned} D\{\tilde{\boldsymbol{e}} := \tilde{\boldsymbol{e}}_{j+1}\} &= P^{-1} \big(B^{(j)}\big)^T \Big[B^{(j)} P^{-1} \big(B^{(j)}\big)^T \Big]^{-1} \times \\ &\times \Big[B^{(j)} D\{\boldsymbol{e}\} \big(B^{(j)}\big)^T - A^{(j)} D\{\hat{\boldsymbol{\Xi}}\} \big(A^{(j)}\big)^T \Big] \Big[B^{(j)} P^{-1} \big(B^{(j)}\big)^T \Big]^{-1} B^{(j)} P^{-1}, \end{aligned} \tag{4.7c}$$

while the *variance component* is estimated by

$$\hat{\sigma}_0^2 = r^{-1} \cdot \boldsymbol{w}_j^T \Big[B^{(j)} P^{-1} \big(B^{(j)}\big)^T \Big]^{-1} \big(\boldsymbol{w}_j - A^{(j)} \hat{\boldsymbol{\xi}}_{j+1} \big). \tag{4.8}$$

It is worth noting that Algorithm 4 conforms to the recommendations by Pope [1972], who pointed out some "pitfalls" risked by taking short cuts in the iterative adjustment of nonlinear problems. The main points made by Pope and treated in this algorithm are the following:

1. The coefficients for $A^{(j)}$ and $B^{(j)}$ are evaluated at the most recent adjusted values of *all* parameters *and* observables.

2. The vector $\boldsymbol{w}_j$ does not necessarily have the form $\boldsymbol{b}(\boldsymbol{Y}, \boldsymbol{\Xi}_j)$, except at the first iteration, although it might be a numerically sufficient approximation.

3. In general, the second term of $\boldsymbol{w}_j := \boldsymbol{b}(\boldsymbol{\mu}_j, \boldsymbol{\Xi}_j) + B^{(j)} \cdot (\boldsymbol{Y} - \boldsymbol{\mu}_j)$ for $j > 0$ does not vanish numerically.

4. The vectors $\boldsymbol{\Xi}_j$ and $\boldsymbol{\mu}_j$ are updated differently. $\boldsymbol{\Xi}_j$ is updated by adding the estimated corrections $\hat{\boldsymbol{\xi}}_{j+1}$ to the solution from the previous iteration (and subtracting the random zero vector, at least theoretically), whereas $\boldsymbol{\mu}_j$ is updated by subtracting the newly predicted error vector $\tilde{\boldsymbol{e}}^{j+1}$ from the observation vector $\boldsymbol{Y}$ (and then subtracting the random zero vector), not by subtracting from the adjusted observation vector of the previous iteration.

4.2 Analytical comparisons between the least-squares solutions within the EIV and Gauss-Helmert Models

The main objective in this section is to compare Algorithm 2, developed within the EIV model, to Algorithm 4, which was developed within the GHM. Fang [2011] has made some comparisons already. Prior to his work, Schaffrin and Snow [2010] had shown already that the (quasilinear) EIV model could be classified as a nonlinear GHM, whereas Neitzel and Petrovic [2008] made comparisons at the adjustment (i.e., numerical) level in showing that the LESS within the GHM generated the same optimally fitted 2-D line as other TLS approaches. Here this topic is explored in a little further detail with the aim of better understanding similarities and differences between the models that Algorithms 2 and 4 were developed within.

In this context, comparisons between the models mean that each model represents the same underlying physical phenomenon that the data are generated by. (Note that models are said to generate the data, not vice versa [see Rao et al., 2008, page 3].) The investigation is restricted to problems where the measurement variables in the EIV model appear only in linear form, which is consistent with the description of the data matrix on page 14. It is also assumed that the data matrix X has full column rank; thus, the redundancy r of the EIV model satisfies $r = k - m$, implying that the number of rows of the matrix B in the GHM is $m + r = k$.

Here the data matrix X in the EIV model (4.1) is considered as a matrix of observations, though one or more columns could be constant in some cases. In any case, it obviously contains coefficients of the unknown parameter vector $\boldsymbol{\Xi}$. Likewise, the matrix A in the GHM also contains coefficients of the unknown parameter vector $\boldsymbol{\Xi}$, as indicated in Step 1 of Algorithm 4. These coefficients of A also represent measurement variables, since we presently focus exclusively on problems where measurement variables appear only in linear form (e.g., line fitting and plane fitting), as opposed to other problems of higher degree in the measurement variables (e.g., conic sections). In such cases, the matrix X in the EIV model (4.1) is numerically identical to the initial matrix $A^{(0)}$ in the GHM (4.6b), since $A^{(0)} = (X - E_X)|_{\boldsymbol{\mu}_0} = X - \underset{\sim}{0}$, with $\underset{\sim}{0}$ as "random zero matrix." For $j \in \mathbb{N}$, the matrices $A^{(j)}$ become $A^{(j)} = (X - E_X)|_{\boldsymbol{\mu}_j} = (X - \tilde{E}_X^{(j)}) - \underset{\sim}{0}$, in accordance with (4.6b).

Using the notation for the EIV-model presented in (4.1), equations (2.10b) and (3.20d) for the matrices $Q_1^{(j)}$ are replaced by

$$Q_1^{(j)} := Q_y - Q_{yX}\Big(\boldsymbol{\Xi}_j \otimes I_k\Big) - \Big(\boldsymbol{\Xi}_j \otimes I_k\Big)^T Q_{Xy} + \Big(\boldsymbol{\Xi}_j \otimes I_k\Big)^T Q_X\Big(\boldsymbol{\Xi}_j \otimes I_k\Big). \tag{4.9}$$

Likewise, the normal equations (2.18) may be rewritten as

$$\Big[\Big(X - \tilde{E}_X\Big)^T Q_1^{-1}\Big(X - \tilde{E}_X\Big)\Big]\hat{\boldsymbol{\Xi}} = \Big(X - \tilde{E}_X\Big)^T Q_1^{-1}\Big(\boldsymbol{y} - \tilde{E}_X\hat{\boldsymbol{\Xi}}\Big), \tag{4.10a}$$

or, equivalently, as

$$\Big[\Big(A^{(j)} + \underset{\sim}{0}\Big)^T (Q_1^j)^{-1}\Big(A^{(j)} + \underset{\sim}{0}\Big)\Big]\hat{\boldsymbol{\Xi}}^{(j)} = \Big(A^{(j)} + \underset{\sim}{0}\Big)^T (Q_1^j)^{-1}\Big(\boldsymbol{y} - \tilde{E}_X^{(j-1)} \cdot \hat{\boldsymbol{\Xi}}^{(j-1)}\Big) \tag{4.10b}$$

for $j \in \mathbb{N}$.

Now the matrix product BQB^T from Algorithm 4 is compared to the matrix Q_1 of Algorithm 2. By definition, the matrix $B^{(j)}$ of (4.6b) is comprised of the first partial-derivatives of the nonlinear function $\boldsymbol{b}$ with respect to the true observables $\boldsymbol{\mu} = \boldsymbol{Y} - \boldsymbol{e}$, which in the current context will then be comprised of an identity matrix associated with the vector of observables $\boldsymbol{y} - \boldsymbol{e}_y$ and a second matrix containing the approximate parameters $\hat{\boldsymbol{\Xi}} - \underset{\sim}{0}$ (with opposite sign) associated with the corrected data matrix $\boldsymbol{X} - E_X$, as shown in the EIV model (4.1). That is, $B^{(j)}$ can be expressed as

$$B^{(j)} = \Big[I_k, \quad -\big(\boldsymbol{\Xi}_j \otimes I_k\big)^T \Big] \text{ with } j \in \mathbb{N}_0, \tag{4.11}$$

leading to the product

$$B^{(j)}Q(B^{(j)})^T = \Big[I_k, \quad -\big(\boldsymbol{\Xi}_j \otimes I_k\big)^T \Big] \begin{bmatrix} \underset{k\times k}{Q_y} & Q_{yX} \\ Q_{Xy} & \underset{km\times km}{Q_X} \end{bmatrix} \begin{bmatrix} I_k \\ -\big(\boldsymbol{\Xi}_j \otimes I_k\big) \end{bmatrix}, \tag{4.12}$$

which when multiplied out gives

$$\begin{aligned} B^{(j)}Q(B^{(j)})^T = Q_y - Q_{yX}\Big((\hat{\boldsymbol{\Xi}}^{(j)} - \underset{\sim}{0}) \otimes I_k\Big) - \Big((\hat{\boldsymbol{\Xi}}^{(j)} - \underset{\sim}{0}) \otimes I_k\Big)^T Q_{Xy} \\ + \Big((\hat{\boldsymbol{\Xi}}^{(j)} - \underset{\sim}{0}) \otimes I_k\Big)^T Q_X \Big((\hat{\boldsymbol{\Xi}}^{(j)} - \underset{\sim}{0}) \otimes I_k\Big), \end{aligned} \tag{4.13}$$

thereby showing numerical coincidence with $Q_i^{(j)}$ in (4.9). It is again noted that $\boldsymbol{\Xi}_j$ denotes the approximate, non-random value (i.e., the expansion point) of the parameter vector $\boldsymbol{\Xi}$ at the j-th iteration, whereas $\hat{\boldsymbol{\Xi}}^{(j)}$ denotes the estimated (and therefore random) parameter vector at the j-th iteration, both in accordance with (4.6g).

Therefore, for the types of underlying functional models that have been considered in this section, namely those that can be treated by the EIV model described in Chapter 2, it has been shown that the matrices $Q_1^{(j)}$ in the algorithms for the WTLS within the EIV model correspond to the matrices $B^{(j)}Q(B^{(j)})^T$ in the algorithms for the LESS within the GHM, as long as they are evaluated at the same values for the estimated, resp. approximate parameters. These numerical equivalencies are shown for specific problems in §6.1.3 and §6.2.1 for 2-D line-fitting and 2-D similarity transformations, respectively.

Now let us compare the matrices $A^{(j)}$ in Algorithm 4, where the the superscript j denotes their evaluation at the jth iteration of the model, to the matrix difference $X - \tilde{E}_X^{(j)}$. It was already stressed that $A^{(j)}$ must be evaluated at the current expansion point for the measurement variables, as implied by (4.6h), namely the measured values minus the residuals minus the random zero vector. But for the type of underlying functions considered here, this is nothing more than $X - \tilde{E}_X^{(j)} - \underset{\sim}{0}$ in the EIV model. Thus, the numerical connection between $A^{(j)} + \underset{\sim}{0}$ in the LESS algorithm within the GHM and $X - \tilde{E}_X^{(j)}$ in the WTLS algorithm within the EIV model has been made.

It is easy to verify this numerical equivalence, which will be done in the experiments of Chapter 6.

Up to now, all the terms in the LESS within the GHM have been compared to the corresponding terms in the WTLS within the EIV model, except for the so-called misclosure vectors, which appear as $\boldsymbol{w}_j$ in the LESS-GHM (4.6e) and as $\boldsymbol{y} - \tilde{E}_X \hat{\boldsymbol{\Xi}}$ in the WTLSS-EIV (4.10a). The relationship between these two terms will be established in the following:

Firstly, from (4.6c), in conjunction with (4.3), we obtain

$$\begin{aligned} \boldsymbol{w}_j &= \begin{bmatrix} I_k, & -(\boldsymbol{\Xi}_j \otimes I_k)^T \end{bmatrix} \begin{bmatrix} \boldsymbol{y} - \tilde{\boldsymbol{e}}_y^{(j)} - \underset{\sim}{\boldsymbol{0}} \\ \operatorname{vec}(X - \tilde{E}_X^{(j)} - \underset{\sim}{0}) \end{bmatrix} + B^{(j)} \begin{bmatrix} \tilde{\boldsymbol{e}}^{(j)} + \underset{\sim}{\boldsymbol{0}} \\ \operatorname{vec}(\tilde{E}_X^{(j)} + \underset{\sim}{0}) \end{bmatrix} = \\ &= \begin{bmatrix} I_k, & -(\boldsymbol{\Xi}_j \otimes I_k)^T \end{bmatrix} \begin{bmatrix} \boldsymbol{y} \\ \operatorname{vec} X \end{bmatrix} = \boldsymbol{y} - \left(A^{(j)} + \tilde{E}_X^{(j)} + \tilde{0} \right) \cdot \boldsymbol{\Xi}_j, \end{aligned} \tag{4.14a}$$

or, equivalently,

$$\boldsymbol{w}_j + A^{(j)} \cdot \boldsymbol{\Xi}_j = \boldsymbol{y} - \left(\tilde{E}_X^{(j)} + \underset{\sim}{0} \right) \left(\hat{\boldsymbol{\Xi}}^{(j)} - \underset{\sim}{\boldsymbol{0}} \right). \tag{4.14b}$$

Secondly, it follows from (4.6e), (4.6g), and (4.14b) that

$$\begin{aligned} \hat{\boldsymbol{\Xi}}^{(j+1)} = &\left[(A^{(j)})^T \left[B^{(j)} P^{-1} (B^{(j)})^T \right]^{-1} A^{(j)} \right]^{-1} (A^{(j)})^T \left[B^{(j)} P^{-1} (B^{(j)})^T \right]^{-1} \times \\ &\times \left[\boldsymbol{y} - (\tilde{E}_X^{(j)} + \underset{\sim}{0}) \cdot (\hat{\boldsymbol{\Xi}}^{(j)} + \underset{\sim}{\boldsymbol{0}}) \right], \end{aligned} \tag{4.15}$$

showing that this solution is similar, but *not entirely* identical, to (4.10b) under consideration of (4.13).

Chapter 5: Weighted Total Least-Squares Collocation

The previous chapters have dealt only with an unknown (and unobservable) parameter vector $\boldsymbol{\xi}$ of type *fixed effects*. In this chapter, stochastic prior information is admitted for the parameter vector such that it becomes a vector of *random effects*, now denoted by $\boldsymbol{x}$. The model that incorporates stochastic prior information for the parameter vector $\boldsymbol{x}$ is called the *random effects model* (REM). First a review of the traditional REM is presented, and then the EIV model is modified to include stochastic prior information.

The revised model is called the *errors-in-variables with random effects* model (EIV-REM) as introduced by Schaffrin [2009], but now with consideration of arbitrary cofactor matrices, including cross-correlation between certain cofactor matrices. The least-squares predictor within the REM can be said to belong to the class of predictors called *least-squares collocation*, a predictor originally developed by Moritz in the early 1960's for the purpose of combining various geodetic data types in order to predict functions of the Earth's gravity field; see, e.g., Heiskanen and Moritz [1967] or Moritz [1970]. The work by Krarup [1969] also advanced the concept of least-squares collocation, as emphasized by Borre in the Preface to Krarup [2006].

5.1 Brief review of the random effects model (REM) and least-squares collocation

Following Schaffrin [2001], for instance, the *random effects model* (REM) is defined as

$$\begin{aligned} \boldsymbol{y} &= A\boldsymbol{x} + \boldsymbol{e}, \\ \boldsymbol{\beta}_0 &= \boldsymbol{x} + \boldsymbol{e}_0, \end{aligned} \qquad \begin{bmatrix} \boldsymbol{e} \\ \boldsymbol{e}_0 \end{bmatrix} \sim (\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \sigma_0^2 \begin{bmatrix} P^{-1} & 0 \\ 0 & Q_0 \end{bmatrix}), \tag{5.1}$$

with

$\boldsymbol{y}$	the $n \times 1$ observation vector,
$\boldsymbol{x}$	the $m \times 1$ (unknown) random effects vector,
A	the $n \times m$ (known) non-random coefficient matrix of rank $q \leq \min\{m, n\}$,
$\boldsymbol{e}$	the $n \times 1$ (unknown) random error vector associated with $\boldsymbol{y}$,
$\boldsymbol{\beta}_0$	the $m \times 1$ vector of (given) expected values $\boldsymbol{\beta}_0 = E\{\boldsymbol{x}\}$,
$\boldsymbol{e}_0$	the $m \times 1$ vector of (unknown) random errors of the *prior information*,
σ_0^2	the (unknown) variance component,
P	the $n \times n$ symmetric positive-definite weight matrix associated with $\boldsymbol{y}$, such that $P^{-1} = Q$, with Q being the cofactor matrix, and
Q_0	the $m \times m$ symmetric positive-(semi)definite (or nnd) cofactor matrix associated with $\boldsymbol{x}$.

No correlations between $\boldsymbol{e}$ and $\boldsymbol{e}_0$ are introduced, as they are assumed to come from completely different sources in most cases. Furthermore, it is assumed that they share a common variance component σ_0^2. Note that, in contrast to the previously presented models, the coefficient matrix A could be rank deficient, whereas the cofactor matrix Q is assumed to be non-singular, though the cofactor matrix Q_0 could be singular.

According to Schaffrin [2001], an inhomogeneous linear predictor $\tilde{\boldsymbol{x}}$, called the *inhomBLIP* of $\boldsymbol{x}$, can be derived based on the principle of minimum mean-squared error. The predictor is known to be *weakly unbiased* based on the equality $E\{\tilde{\boldsymbol{x}}\} = \boldsymbol{\beta}_0 = E\{\boldsymbol{x}\}$. An equivalent linear predictor can be derived from the principle of weighted least-squares by forming the Lagrange target function

$$\phi(\boldsymbol{x}) = (\boldsymbol{y} - A\boldsymbol{x})^T P (\boldsymbol{y} - A\boldsymbol{x}) + (\boldsymbol{\beta}_0 - \boldsymbol{x})^T Q_0^{-1} (\boldsymbol{\beta}_0 - \boldsymbol{x}) = \min_{\boldsymbol{x}}. \tag{5.2}$$

Note that the matrix Q_0 is shown to be nonsingular in the target function, a requirement that will be removed in the final step of the derivation. Defining the terms $[N, \boldsymbol{c}] := A^T P [A, \boldsymbol{y}]$, the Euler-Lagrange (first-order) necessary condition is

$$\frac{1}{2}\frac{\partial \phi}{\partial \boldsymbol{x}} = -\boldsymbol{c} + N\tilde{\boldsymbol{x}} - Q_0^{-1}\boldsymbol{\beta}_0 + Q_0^{-1}\tilde{\boldsymbol{x}} \doteq \boldsymbol{0}, \tag{5.3}$$

with the sufficient condition for minimization provided by

$$\frac{1}{2}\frac{\partial^2 \phi}{\partial \boldsymbol{x} \partial \boldsymbol{x}^T} = N + Q_0^{-1}, \tag{5.4}$$

which is positive-definite as long as the augmented matrix $[A^T \,|\, Q_0]$ has full row rank m. The normal equations are then written as

$$\left(N + Q_0^{-1}\right)\tilde{\boldsymbol{x}} = \boldsymbol{c} + Q_0^{-1}\boldsymbol{\beta}_0, \tag{5.5a}$$

$$\text{or} \quad (I_m + Q_0 N)\,\tilde{\boldsymbol{x}} = \boldsymbol{\beta}_0 + Q_0\boldsymbol{c}, \tag{5.5b}$$

the second of which is valid for an either singular or nonsingular matrix Q_0. The normal equations lead to

$$\tilde{\boldsymbol{x}} = \left(N + Q_0^{-1}\right)^{-1}(\boldsymbol{c} + Q_0^{-1}\boldsymbol{\beta}_0) = \tag{5.6a}$$

$$= \boldsymbol{\beta}_0 + Q_0\left(I_m + NQ_0\right)^{-1}A^T P(\boldsymbol{y} - A\boldsymbol{\beta}_0) \tag{5.6b}$$

for the least-squares predictor within the REM, which is also called *least-squares collocation*. Note that (5.6b) appears as an update to the given prior-information vector $\boldsymbol{\beta}_0$.

Since the expectation of $\boldsymbol{c}$ is $E\{\boldsymbol{c}\} = A^T PE\{\boldsymbol{y}\} = N\boldsymbol{\beta}_0$, then indeed we find the predictor $\tilde{\boldsymbol{x}}$ to be weakly unbiased due to

$$E\{\tilde{\boldsymbol{x}}\} = \boldsymbol{\beta}_0 = E\{\boldsymbol{x}\}, \tag{5.7}$$

which leads to the *mean-square prediction error*

$$\text{MSE}\{\tilde{\boldsymbol{x}}\} = D\{\tilde{\boldsymbol{x}} - \boldsymbol{x}\} = \sigma_0^2\left(N + Q_0^{-1}\right)^{-1} = \sigma_0^2 Q_0\left(I_m + NQ_0\right)^{-1}, \tag{5.8}$$

where $D\{\cdot\}$ denotes the dispersion (or variance) of its argument.

Analytically, the predictor $\tilde{\boldsymbol{x}}$ also can be expressed as a combination of the first and second statistical moments of the true (but unknown) variables $\boldsymbol{x}$ and $\boldsymbol{y}$:

$$\tilde{\boldsymbol{x}} = E\{\boldsymbol{x}\} + C\{\boldsymbol{x}, \boldsymbol{y}\} \cdot [D\{\boldsymbol{y}\}]^{-1}(\boldsymbol{y} - E\{\boldsymbol{y}\}), \tag{5.9a}$$

$$\text{with} \quad C\{\boldsymbol{x}, \boldsymbol{y}\} = \sigma_0^2 Q_0 A^T \text{ and } D\{\boldsymbol{y}\} = \sigma_0^2\left(P^{-1} + AQ_0A^T\right), \tag{5.9b}$$

where $C\{\cdot, \cdot\}$ denotes the covariance of its arguments.

The predicted residual vectors $\tilde{\boldsymbol{e}}$ and $\tilde{\boldsymbol{e}}_0$ are then given by

$$\tilde{\boldsymbol{e}}_0 = \boldsymbol{\beta}_0 - \tilde{\boldsymbol{x}} = -Q_0\hat{\boldsymbol{\nu}}^0 \quad \text{for} \quad \hat{\boldsymbol{\nu}}^0 := \left(I_m + NQ_0\right)^{-1}(\boldsymbol{c} - N\boldsymbol{\beta}_0), \tag{5.10a}$$

$$\tilde{\boldsymbol{e}} = \boldsymbol{y} - A\tilde{\boldsymbol{x}} = \left[I_n - AQ_0\left(I_m + NQ_0\right)^{-1}A^T P\right](\boldsymbol{y} - A\boldsymbol{\beta}_0) \Rightarrow A^T P\tilde{\boldsymbol{e}} = \hat{\boldsymbol{\nu}}^0, \tag{5.10b}$$

which permits the estimated variance component

$$\hat{\sigma}_0^2 = n^{-1} \cdot \left(\tilde{\boldsymbol{e}}^T P\tilde{\boldsymbol{e}} + (\hat{\boldsymbol{\nu}}^0)^T Q_0\hat{\boldsymbol{\nu}}^0\right) = n^{-1} \cdot \left(\boldsymbol{y}^T P\boldsymbol{y} - \boldsymbol{c}^T\tilde{\boldsymbol{x}} - \boldsymbol{\beta}_0^T\hat{\boldsymbol{\nu}}^0\right) \tag{5.11}$$

to be formed without inversion of the matrix Q_0.

5.2 The EIV with random effects model (EIV-REM)

Certain problems with measurement variables appearing in both the observation vector $\boldsymbol{y}$ and in the coefficient matrix A are best treated by the *errors-in-variables with random effects model* (EIV-REM) when stochastic prior information for the parameters is available. Such a model is formed by combining the EIV model of (2.1) with the REM of (5.1) as follows:

$$\begin{aligned} \boldsymbol{y} &= (A - E_A)\boldsymbol{x} + \boldsymbol{e}_y, \\ \boldsymbol{\beta}_0 &= \boldsymbol{x} + \boldsymbol{e}_0, \end{aligned} \qquad \begin{bmatrix} \boldsymbol{e}_y \\ \boldsymbol{e}_A \\ \boldsymbol{e}_0 \end{bmatrix} \sim \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \sigma_0^2 \left[\begin{array}{cc|c} Q_y & Q_{yA} & 0 \\ Q_{Ay} & Q_A & 0 \\ \hline 0 & 0 & Q_0 \end{array}\right] \right). \tag{5.12}$$

All the terms of (5.12) have been defined already in the lists following (2.1) and (5.1), so those definitions are not repeated here. However, it is noted that here both the coefficient (data) matrix A and the prior-information cofactor matrix Q_0 could be rank deficient, as long as $[A^T \,|\, Q_0]$ has full row rank m.

5.2.1 Total least-squares collocation (TLSC)

Consistent with previous chapters, the predictor $\tilde{\boldsymbol{x}}$ of the unknown random effects vector $\boldsymbol{x}$ is derived from the principle of least squares:

$$\boldsymbol{e}_y^T P_{11} \boldsymbol{e}_y + 2\boldsymbol{e}_y^T P_{12} \boldsymbol{e}_A + \boldsymbol{e}_A^T P_{22} \boldsymbol{e}_A + \boldsymbol{e}_0^T Q_0^{-1} \boldsymbol{e}_0 = \min, \tag{5.13}$$

subject to the model (5.12), leading to the Lagrange target function

$$\begin{aligned} \phi(\boldsymbol{e}_y, \boldsymbol{e}_A, \boldsymbol{e}_0, \boldsymbol{\lambda}) = \boldsymbol{e}_y^T P_{11} \boldsymbol{e}_y + 2\boldsymbol{e}_y^T P_{12} \boldsymbol{e}_A + \boldsymbol{e}_A^T P_{22} \boldsymbol{e}_A + \boldsymbol{e}_0^T Q_0^{-1} \boldsymbol{e}_0 + \\ + 2\boldsymbol{\lambda}^T \Big(\boldsymbol{y} - A\boldsymbol{\beta}_0 - \boldsymbol{e}_y + \Big(\boldsymbol{\beta}_0^T \otimes I_n \Big) \boldsymbol{e}_A + A\boldsymbol{e}_0 - E_A \boldsymbol{e}_0 \Big) = \text{stationary}, \end{aligned} \tag{5.14}$$

where $\boldsymbol{\lambda}$ is an $n \times 1$ vector of Lagrange multipliers and Q_0 is momentarily assumed to be nonsingular. The weight matrix $P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$ has already been defined in (2.2).

The first-order partial derivatives of ϕ, set to zero and evaluated at $\tilde{\boldsymbol{e}}_y, \tilde{\boldsymbol{e}}_A, \tilde{\boldsymbol{e}}_0, \hat{\boldsymbol{\lambda}}$, provide the Euler-Lagrange necessary conditions

$$\frac{1}{2}\frac{\partial \phi}{\partial \boldsymbol{e}_y} = P_{11}\tilde{\boldsymbol{e}}_y + P_{12}\tilde{\boldsymbol{e}}_A \; - \hat{\boldsymbol{\lambda}} \doteq \mathbf{0}, \tag{5.15a}$$

$$\frac{1}{2}\frac{\partial \phi}{\partial \boldsymbol{e}_A} = P_{21}\tilde{\boldsymbol{e}}_y + P_{22}\tilde{\boldsymbol{e}}_A + \Big(\boldsymbol{\beta}_0 \otimes I_n \Big)\hat{\boldsymbol{\lambda}} - \Big(\tilde{\boldsymbol{e}}_0 \otimes I_n \Big)\hat{\boldsymbol{\lambda}} \doteq \mathbf{0}, \tag{5.15b}$$

$$\frac{1}{2}\frac{\partial \phi}{\partial \boldsymbol{e}_0} = Q_0^{-1}\tilde{\boldsymbol{e}}_0 + A^T\hat{\boldsymbol{\lambda}} - \tilde{E}_A^T\hat{\boldsymbol{\lambda}} \doteq \mathbf{0}, \tag{5.15c}$$

$$\frac{1}{2}\frac{\partial \phi}{\partial \boldsymbol{\lambda}} = \boldsymbol{y} - A\boldsymbol{\beta}_0 - \tilde{\boldsymbol{e}}_y + \tilde{E}_A\boldsymbol{\beta}_0 + (A - \tilde{E}_A)\tilde{\boldsymbol{e}}_0 \doteq \mathbf{0}, \tag{5.15d}$$

which are manipulated algebraically below in order to derive the predictor $\tilde{\boldsymbol{x}}$.

Substituting (5.15b) into (5.15a) and then, vise versa, substituting (5.15a) into (5.15b), leads to the residual (predicted error) vectors

$$\tilde{\boldsymbol{e}}_y = \Big(P_{11} - P_{12}P_{22}^{-1}P_{21} \Big)^{-1} \Big[P_{12}P_{22}^{-1}\Big([\boldsymbol{\beta}_0 - \tilde{\boldsymbol{e}}_0] \otimes I_n \Big) + I_n \Big] \hat{\boldsymbol{\lambda}}, \tag{5.16}$$

$$\tilde{\boldsymbol{e}}_A = -\Big(P_{22} - P_{21}P_{11}^{-1}P_{12} \Big)^{-1} \Big[P_{21}P_{11}^{-1} + \Big([\boldsymbol{\beta}_0 - \tilde{\boldsymbol{e}}_0] \otimes I_n \Big) \Big] \hat{\boldsymbol{\lambda}}, \tag{5.17}$$

which, upon applying the relations (2.8), permits the predicted error vectors $\tilde{\boldsymbol{e}}_y$ and $\tilde{\boldsymbol{e}}_A$ to be written in terms of the cofactor matrices and the estimated Lagrange multipliers:

$$\tilde{\boldsymbol{e}}_y = \Big[-Q_{yA}\Big([\boldsymbol{\beta}_0 - \tilde{\boldsymbol{e}}_0] \otimes I_n\Big) + Q_y\Big]\hat{\boldsymbol{\lambda}}, \tag{5.18}$$

$$\tilde{\boldsymbol{e}}_A = \Big[Q_{Ay} - Q_A\Big([\boldsymbol{\beta}_0 - \tilde{\boldsymbol{e}}_0] \otimes I_n\Big)\Big]\hat{\boldsymbol{\lambda}}. \tag{5.19}$$

For compactness, substitute $\tilde{\boldsymbol{x}}$ for $\boldsymbol{\beta}_0 - \tilde{\boldsymbol{e}}_0$, and then combine (5.15d), (5.18), and (5.19) to arrive at

$$\begin{aligned}\boldsymbol{y} - A\tilde{\boldsymbol{x}} &= \tilde{\boldsymbol{e}}_y - (\tilde{\boldsymbol{x}} \otimes I_n)^T \tilde{\boldsymbol{e}}_A = \\ &= \Big[Q_y - Q_{yA}(\tilde{\boldsymbol{x}} \otimes I_n) - (\tilde{\boldsymbol{x}} \otimes I_n)^T Q_{Ay} + (\tilde{\boldsymbol{x}} \otimes I_n)^T Q_A (\tilde{\boldsymbol{x}} \otimes I_n)\Big]\hat{\boldsymbol{\lambda}} = \qquad (5.20\text{a}) \\ &= Q_1 \hat{\boldsymbol{\lambda}},\end{aligned}$$

with

$$Q_1 := \Big[Q_y - Q_{yA}(\tilde{\boldsymbol{x}} \otimes I_n) - (\tilde{\boldsymbol{x}} \otimes I_n)^T Q_{Ay} + (\tilde{\boldsymbol{x}} \otimes I_n)^T Q_A (\tilde{\boldsymbol{x}} \otimes I_n)\Big]. \tag{5.20b}$$

If Q_1 is nonsingular, though neither Q_y nor Q_A would necessarily have to be nonsingular, then

$$\hat{\boldsymbol{\lambda}} = Q_1^{-1}(\boldsymbol{y} - A\tilde{\boldsymbol{x}}), \text{ if } Q_1^{-1} \text{ exists.} \tag{5.21}$$

Continuing under the assumption that both Q_1^{-1} and Q_0^{-1} exist, and using (5.15c), we can write

$$A^T\hat{\boldsymbol{\lambda}} = \tilde{E}_A^T\hat{\boldsymbol{\lambda}} - Q_0^{-1}\tilde{\boldsymbol{e}}_0 = \tag{5.22a}$$

$$= \tilde{E}_A^T Q_1^{-1}(\boldsymbol{y} - A\boldsymbol{\beta}_0) + \Big(\tilde{E}_A^T Q_1^{-1} A - Q_0^{-1}\Big)\tilde{\boldsymbol{e}}_0 = \tag{5.22b}$$

$$= A^T Q_1^{-1}(\boldsymbol{y} - A\boldsymbol{\beta}_0) + A^T Q_1^{-1} A \tilde{\boldsymbol{e}}_0. \tag{5.22c}$$

Then subtracting (5.22b) from (5.22c) leads to the *normal equations*

$$\Big[(A - \tilde{E}_A)^T Q_1^{-1} A + Q_0^{-1}\Big]\tilde{\boldsymbol{e}}_0 = -(A - \tilde{E}_A)^T Q_1^{-1}(\boldsymbol{y} - A\boldsymbol{\beta}_0), \tag{5.23a}$$

$$\Big[(A - \tilde{E}_A)^T Q_1^{-1}(A - \tilde{E}_A) + Q_0^{-1}\Big]\tilde{\boldsymbol{e}}_0 = -(A - \tilde{E}_A)^T Q_1^{-1}(\boldsymbol{y} - (A - \tilde{E}_A)\boldsymbol{\beta}_0 - \tilde{E}_A\tilde{\boldsymbol{x}}). \tag{5.23b}$$

Solving for the predicted error vector yields

$$\boxed{\tilde{\boldsymbol{e}}_0 = -\Big[(A - \tilde{E}_A)^T Q_1^{-1}(A - \tilde{E}_A) + Q_0^{-1}\Big]^{-1}(A - \tilde{E}_A)^T Q_1^{-1}(\boldsymbol{y} - (A - \tilde{E}_A)\boldsymbol{\beta}_0 - \tilde{E}_A\tilde{\boldsymbol{x}})}, \tag{5.24}$$

or for the predicted vector of random effects

$$\boxed{\begin{aligned}\tilde{\boldsymbol{x}} &= \boldsymbol{\beta}_0 + \Big[(A - \tilde{E}_A)^T Q_1^{-1}(A - \tilde{E}_A) + Q_0^{-1}\Big]^{-1} \cdot \\ &\quad \cdot (A - \tilde{E}_A)^T Q_1^{-1}(\boldsymbol{y} - A\boldsymbol{\beta}_0 + \tilde{E}_A\tilde{\boldsymbol{e}}_0) = \qquad (5.25\text{a}) \\ &= \boldsymbol{\beta}_0 + Q_0\Big[I_m + (A - \tilde{E}_A)^T Q_1^{-1}(A - \tilde{E}_A)Q_0\Big]^{-1} \cdot \\ &\quad \cdot (A - \tilde{E}_A)^T Q_1^{-1}(\boldsymbol{y} - A\boldsymbol{\beta}_0 + \tilde{E}_A\tilde{\boldsymbol{e}}_0), \qquad (5.25\text{b})\end{aligned}}$$

which provide update formulas with respect to the prior information $\boldsymbol{\beta}_0$. The prediction $\tilde{\boldsymbol{x}}$ of the random effects vector $\boldsymbol{x}$ by (5.25a) will herein be called *weighted total least-squares collocation* (WTLSC).

After prediction of the random effects vector $\boldsymbol{x}$ and the error vectors $\boldsymbol{e}_y$, $\boldsymbol{e}_A$, and $\boldsymbol{e}_0$, as well as estimation of the vector of Lagrange multipliers $\boldsymbol{\lambda}$, the total sum of squared residuals (TSSR) can be computed by

$$\Omega = \begin{bmatrix}\tilde{\boldsymbol{e}}_y^T, & \tilde{\boldsymbol{e}}_A^T\end{bmatrix} \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22}\end{bmatrix} \begin{bmatrix}\tilde{\boldsymbol{e}}_y \\ \tilde{\boldsymbol{e}}_A\end{bmatrix} + \tilde{\boldsymbol{e}}_0^T Q_0^{-1} \tilde{\boldsymbol{e}}_0 = \tag{5.26a}$$

$$= \hat{\boldsymbol{\lambda}}^T \Big(Q_1 + (A - \tilde{E}_A) Q_0 (A - \tilde{E}_A)^T \Big) \hat{\boldsymbol{\lambda}}, \tag{5.26b}$$

which provides formulas for an either regular or singular matrix Q_0.

§6.4 describes an experiment where the WTLSC predictor is applied to a 2-D line-fitting problem, using Algorithm 5.

Algorithm 5 For TLSC within the EIV with random effects model

Step 1: Assign $\tilde{\boldsymbol{x}}^{(0)} = \boldsymbol{\beta}_0$, $\tilde{E}_A^{(0)} = 0$, and $\tilde{\boldsymbol{e}}_0^{(0)} = \mathbf{0}$.

Step 2:

repeat For $i \in \mathbb{N}$, compute

$$Q_1^{(i)} := \Big[Q_y - Q_{yA}\Big(\tilde{\boldsymbol{x}}^{(i-1)} \otimes I_n\Big) - \Big(\tilde{\boldsymbol{x}}^{(i-1)} \otimes I_n\Big)^T Q_{Ay} + \\ + \Big(\tilde{\boldsymbol{x}}^{(i-1)} \otimes I_n\Big)^T Q_A\Big(\tilde{\boldsymbol{x}}^{(i-1)} \otimes I_n\Big)\Big] \tag{5.27a}$$

if Q_0 is regular **then**

$$\tilde{\boldsymbol{x}}^{(i)} = \boldsymbol{\beta}_0 + \Big[\Big(A - \tilde{E}_A^{(i-1)}\Big)^T \Big(Q_1^{(i)}\Big)^{-1}\Big(A - \tilde{E}_A^{(i-1)}\Big) + Q_0^{-1}\Big]^{-1} \times \\ \times \Big(A - \tilde{E}_A^{(i-1)}\Big)^T \Big(Q_1^{(i)}\Big)^{-1}\Big(\boldsymbol{y} - A\boldsymbol{\beta}_0 + \tilde{E}_A^{(i-1)} \cdot \tilde{\boldsymbol{e}}_0^{(i-1)}\Big) \tag{5.27b}$$

else

$$\tilde{\boldsymbol{x}}^{(i)} = \boldsymbol{\beta}_0 + Q_0\Big[I_m + (A - \tilde{E}_A^{(i-1)})^T \Big(Q_1^{(i)}\Big)^{-1}(A - \tilde{E}_A^{(i-1)})Q_0\Big]^{-1} \times \\ \times (A - \tilde{E}_A^{(i-1)})^T \Big(Q_1^{(i)}\Big)^{-1}\Big(\boldsymbol{y} - A\boldsymbol{\beta}_0 + \tilde{E}_A^{(i-1)} \cdot \tilde{\boldsymbol{e}}_0^{(i-1)}\Big) \tag{5.27c}$$

▷ This equation can also be used when Q_0 is regular.

end if

$$\hat{\boldsymbol{\lambda}}^{(i)} = \Big(Q_1^{(i)}\Big)^{-1}\Big[\Big(\boldsymbol{y} - \tilde{E}_A^{(i-1)}\tilde{\boldsymbol{x}}^{(i-1)}\Big) - \Big(A - \tilde{E}_A^{(i-1)}\Big)\tilde{\boldsymbol{x}}^{(i)}\Big] \tag{5.27d}$$

$$\tilde{\boldsymbol{e}}_A^{(i)} = \Big[Q_{Ay} - Q_A\Big(\tilde{\boldsymbol{x}}^{(i)} \otimes I_n\Big)\Big]\hat{\boldsymbol{\lambda}}^{(i)} \quad \text{and} \quad \tilde{E}_A^{(i)} = \operatorname{Invec} \tilde{\boldsymbol{e}}_A^{(i)} \tag{5.27e}$$

$$\tilde{\boldsymbol{e}}_0^{(i)} = \boldsymbol{\beta}_0 - \tilde{\boldsymbol{x}}^{(i)} \tag{5.27f}$$

until

$$\Big\|\tilde{\boldsymbol{x}}^{(i)} - \tilde{\boldsymbol{x}}^{(i-1)}\Big\| < \delta \tag{5.27g}$$

for a chosen threshold δ.

5.2.2 The special case of iid data within the EIV-REM

In some applications, the measurement data are independent and identically distributed (iid), while the distribution of the prior information may remain quite arbitrary. In fact, the assumption of iid data within the EIV model is fairly common in the literature, where very often the error distribution is assumed to be characterized by the dispersion matrix $\sigma_0^2 \cdot I_{n(m+1)}$; see Van Huffel and Vandewalle [1991, p. 230 and assumption (8.6)] or Markovsky et al. [2010, Eqs. (2) and (3)], for example. Therefore, it is worthwhile to show how the development from the preceding section can be simplified somewhat when the data are iid. Rather than repeat every step in detail, only some highlights are given.

The EIV-REM with iid data is provided by

$$\begin{aligned}\boldsymbol{y} &= (A - E_A)\boldsymbol{x} + \boldsymbol{e}_y,\\ \boldsymbol{\beta}_0 &= \boldsymbol{x} + \boldsymbol{e}_0,\end{aligned} \qquad \begin{bmatrix}\boldsymbol{e}_y\\ \boldsymbol{e}_A\\ \boldsymbol{e}_0\end{bmatrix} \sim \left(\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}, \sigma_0^2\left[\begin{array}{cc|c} I_n & 0 & 0\\ 0 & I_{mn} & 0\\ \hline 0 & 0 & Q_0\end{array}\right]\right), \tag{5.28}$$

where $\boldsymbol{e}_A$ is the vectorized form of E_A as before.

Here the dispersion $D\{\boldsymbol{e}_A\} = \sigma_0^2 I_{mn}$ could actually be reduced by use of a singular *selection matrix* S, such that $\sigma_0^2(S \otimes I_n) := \sigma_0^2\left[\begin{smallmatrix} I_{(m-c)n} & 0\\ 0 & 0\end{smallmatrix}\right]$, where c is the number of fixed columns in A. This would require no special treatment as long as the matrix Q_1 defined in (5.31) remains nonsingular.

Temporarily assuming that the cofactor matrix Q_0 is nonsingular, the Lagrange target function for the least-squares approach can be written as

$$\begin{aligned}\phi(\boldsymbol{e}_y, \boldsymbol{e}_A, \boldsymbol{e}_0, \boldsymbol{\lambda}) &= \boldsymbol{e}_y^T\boldsymbol{e}_y + \boldsymbol{e}_A^T\boldsymbol{e}_A + \boldsymbol{e}_0^T Q_0^{-1}\boldsymbol{e}_0 +\\ &+ 2\boldsymbol{\lambda}^T\Big(\boldsymbol{y} - A\boldsymbol{\beta}_0 - \boldsymbol{e}_y + \big(\boldsymbol{\beta}_0^T \otimes I_n\big)\boldsymbol{e}_A + A\boldsymbol{e}_0 - E_A\boldsymbol{e}_0\Big) = \text{stationary},\end{aligned} \tag{5.29}$$

leading to the following Euler-Lagrange necessary conditions:

$$\frac{1}{2}\frac{\partial\phi}{\partial\boldsymbol{e}_y} = \tilde{\boldsymbol{e}}_y - \hat{\boldsymbol{\lambda}} \doteq \boldsymbol{0}, \tag{5.30a}$$

$$\frac{1}{2}\frac{\partial\phi}{\partial\boldsymbol{e}_A} = \tilde{\boldsymbol{e}}_A + \big(\boldsymbol{\beta}_0 \otimes I_n\big)\hat{\boldsymbol{\lambda}} - \big(\tilde{\boldsymbol{e}}_0 \otimes I_n\big)\hat{\boldsymbol{\lambda}} \doteq \boldsymbol{0}, \tag{5.30b}$$

$$\frac{1}{2}\frac{\partial\phi}{\partial\boldsymbol{e}_0} = Q_0^{-1}\tilde{\boldsymbol{e}}_0 + A^T\hat{\boldsymbol{\lambda}} - \tilde{E}_A^T\hat{\boldsymbol{\lambda}} \doteq \boldsymbol{0}, \tag{5.30c}$$

$$\frac{1}{2}\frac{\partial\phi}{\partial\boldsymbol{\lambda}} = \boldsymbol{y} - A\boldsymbol{\beta}_0 - \tilde{\boldsymbol{e}}_y + \tilde{E}_A\boldsymbol{\beta}_0 + (A - \tilde{E}_A)\tilde{\boldsymbol{e}}_0 \doteq \boldsymbol{0}. \tag{5.30d}$$

The matrix Q_1 of (5.20b) reduces to

$$Q_1 = \Big[I_n + \big(\tilde{\boldsymbol{x}}^T \otimes I_n\big)I_{mn}\big(\tilde{\boldsymbol{x}} \otimes I_n\big)\Big] = (1 + \tilde{\boldsymbol{x}}^T\tilde{\boldsymbol{x}})I_n, \tag{5.31}$$

unless the coefficient matrix A has fixed columns, meaning that $\tilde{E}_A$ has the corresponding columns fixed to zero. In that case, the quadratic term $\tilde{\boldsymbol{x}}^T\tilde{\boldsymbol{x}}$ would be reduced to $\tilde{\boldsymbol{x}}^T S\tilde{\boldsymbol{x}}$ so as not to contain elements associated with the fixed columns. For example, if the last c columns of A are fixed, then Q_1 becomes

$$Q_1 = \Big[I_n + \Big(\tilde{\boldsymbol{x}}^T \otimes I_n\Big)\Big(\begin{bmatrix} I_{(m-c)n} & 0 \\ 0 & 0 \end{bmatrix} \otimes I_n\Big)\Big(\tilde{\boldsymbol{x}} \otimes I_n\Big)\Big] = \Big(1 + \sum_{i=1}^{m-c} \tilde{x}_i^2\Big) I_n. \tag{5.32}$$

For the balance of this section, it is assumed that A has no fixed columns, and thus equation (5.31) applies rather than equation (5.32). (Obviously, this restriction would be unnecessary if the *selection matrix* S is used, with I_m as a particular choice.)

Using (5.31) in (5.21), the estimator for the unknown vector of Lagrange multipliers $\boldsymbol{\lambda}$ becomes

$$\boxed{\hat{\boldsymbol{\lambda}} = \frac{1}{(1 + \tilde{\boldsymbol{x}}^T \tilde{\boldsymbol{x}})}(\boldsymbol{y} - A\tilde{\boldsymbol{x}})}, \tag{5.33}$$

resulting in the predicted errors from (5.30b), namely

$$\boxed{\tilde{\boldsymbol{e}}_A = -(\tilde{\boldsymbol{x}} \otimes I_n)\hat{\boldsymbol{\lambda}} \Rightarrow \tilde{E}_A = -\hat{\boldsymbol{\lambda}}\tilde{\boldsymbol{x}}^T}. \tag{5.34}$$

Revising equation (5.23a) with equations (5.31) and (5.34) yields

$$\Big[(A - \tilde{E}_A)^T A + (1 + \tilde{\boldsymbol{x}}^T \tilde{\boldsymbol{x}}) Q_0^{-1}\Big]\tilde{\boldsymbol{e}}_0 = -(A - \tilde{E}_A)^T(\boldsymbol{y} - A\boldsymbol{\beta}_0) \tag{5.35}$$

as the normal equations for the iid case, which, upon adopting $[N, \boldsymbol{c}] = A^T[A, \boldsymbol{y}]$, can be reduced to

$$-(1 + \tilde{\boldsymbol{x}}^T \tilde{\boldsymbol{x}}) Q_0^{-1}\tilde{\boldsymbol{e}}_0 = \boldsymbol{c} - N\tilde{\boldsymbol{x}} - \tilde{E}_A^T(\boldsymbol{y} - A\tilde{\boldsymbol{x}}) \Rightarrow \tag{5.36a}$$

$$\boldsymbol{c} - N\tilde{\boldsymbol{x}} + Q_0^{-1}\tilde{\boldsymbol{e}}_0 = -(\tilde{\boldsymbol{x}}^T \tilde{\boldsymbol{x}}) Q_0^{-1}\tilde{\boldsymbol{e}}_0 + \tilde{E}_A^T(\boldsymbol{y} - A\tilde{\boldsymbol{x}}). \tag{5.36b}$$

Now using the relations $\tilde{\boldsymbol{e}}_0 = \boldsymbol{\beta}_0 - \tilde{\boldsymbol{x}}$ and $\tilde{E}_A^T = -\tilde{\boldsymbol{x}}\hat{\boldsymbol{\lambda}}^T$, a summary of the formulas for the predicted random effects vector $\tilde{\boldsymbol{x}}$ for the case where Q_0^{-1} exists is given by

$$\begin{aligned} &(\boldsymbol{c} + Q_0^{-1}\boldsymbol{\beta}_0) - (N + Q_0^{-1})\tilde{\boldsymbol{x}} = \\ &\quad = -(\tilde{\boldsymbol{x}}^T \tilde{\boldsymbol{x}}) Q_0^{-1}\tilde{\boldsymbol{e}}_0 - \tilde{\boldsymbol{x}}(1 + \tilde{\boldsymbol{x}}^T \tilde{\boldsymbol{x}})^{-1}(\boldsymbol{y} - A\tilde{\boldsymbol{x}})^T(\boldsymbol{y} - A\tilde{\boldsymbol{x}}) = \qquad (5.37\text{a}) \\ &\quad = -\hat{\nu} \cdot \tilde{\boldsymbol{x}} + (\tilde{\boldsymbol{x}}^T \tilde{\boldsymbol{x}}) \cdot \hat{\boldsymbol{\nu}}^0, \\ &\text{where} \quad \hat{\nu} := (1 + \tilde{\boldsymbol{x}}^T \tilde{\boldsymbol{x}})^{-1}(\boldsymbol{y} - A\tilde{\boldsymbol{x}})^T(\boldsymbol{y} - A\tilde{\boldsymbol{x}}), \qquad (5.37\text{b}) \\ &\text{and} \quad \hat{\boldsymbol{\nu}}^0 := -Q_0^{-1}\tilde{\boldsymbol{e}}_0, \qquad (5.37\text{c}) \end{aligned}$$

which is in agreement with Schaffrin [2009]. In the case that Q_0 is singular, the following alternative system of equations can be used:

$$\begin{aligned} &(\boldsymbol{c} - N\boldsymbol{\beta}_0) - [(1 + \tilde{\boldsymbol{x}}^T \tilde{\boldsymbol{x}}) I_m + N Q_0]\hat{\boldsymbol{\nu}}^0 = -\hat{\nu} \cdot \tilde{\boldsymbol{x}}, \qquad (5.38\text{a}) \\ &\text{where} \quad \hat{\nu} := (1 + \tilde{\boldsymbol{x}}^T \tilde{\boldsymbol{x}})^{-1}(\boldsymbol{y} - A\tilde{\boldsymbol{x}})^T(\boldsymbol{y} - A\tilde{\boldsymbol{x}}), \qquad (5.38\text{b}) \\ &\text{and} \quad \tilde{\boldsymbol{x}} = \boldsymbol{\beta}_0 + Q_0\hat{\boldsymbol{\nu}}^0. \qquad (5.38\text{c}) \end{aligned}$$

The predictor $\tilde{\boldsymbol{x}}$ of the random effects vector $\boldsymbol{x}$ has been called *total least-squares collocation* (TLSC) by Schaffrin [2009] and can be extracted from the following algorithm:

Algorithm 6 For TLSC within the EIV with random effects model having iid data

---

Step 1: Assign $\tilde{\boldsymbol{x}}^{(0)} = \boldsymbol{\beta}_0$ and construct an $m \times m$ diagonal selection matrix S having a 0 at every diagonal element that corresponds to a fixed column in the matrix A and 1 at the other diagonal elements.

Step 2:
repeat For $i \in \mathbb{N}$, compute

$$\hat{\nu}^{(i)} = \Big(1 + (\tilde{\boldsymbol{x}}^{(i-1)})^T \cdot S \cdot \tilde{\boldsymbol{x}}^{(i-1)}\Big)^{-1} \Big(\boldsymbol{y} - A\tilde{\boldsymbol{x}}^{(i-1)}\Big)^T \Big(\boldsymbol{y} - A\tilde{\boldsymbol{x}}^{(i-1)}\Big) \tag{5.39a}$$

$$\Big(\hat{\boldsymbol{\nu}}^0\Big)^{(i)} = \Big[\Big(1 + (\tilde{\boldsymbol{x}}^{(i-1)})^T \cdot S \cdot \tilde{\boldsymbol{x}}^{(i-1)}\Big) I_m + NQ_0\Big]^{-1} \Big(\boldsymbol{c} - N\boldsymbol{\beta}_0 + \hat{\nu}^{(i)} \cdot S \cdot \tilde{\boldsymbol{x}}^{(i-1)}\Big) \tag{5.39b}$$

$$\tilde{\boldsymbol{x}}^{(i)} = \boldsymbol{\beta}_0 + Q_0 \cdot \Big(\hat{\boldsymbol{\nu}}^0\Big)^{(i)} \tag{5.39c}$$

▷ Here $[N,\, \boldsymbol{c}] := A^T[A,\, \boldsymbol{y}]$.

until

$$\Big\|\tilde{\boldsymbol{x}}^{(i)} - \tilde{\boldsymbol{x}}^{(i-1)}\Big\| < \delta \tag{5.39d}$$

for a chosen threshold δ.

Step 3: Compute the residual vectors and the total sum of squared residuals (TSSR) as follows:

$$\tilde{\boldsymbol{e}}_y = \Big(1 + \tilde{\boldsymbol{x}}^T \cdot S \cdot \tilde{\boldsymbol{x}}\Big)\Big(\boldsymbol{y} - A\tilde{\boldsymbol{x}}\Big) \tag{5.39e}$$

$$\tilde{E}_A = -\tilde{\boldsymbol{e}}_y \cdot S \cdot \tilde{\boldsymbol{x}},\ \tilde{\boldsymbol{e}}_A = \operatorname{vec} \tilde{E}_A \tag{5.39f}$$

$$\tilde{\boldsymbol{e}}_0 = \boldsymbol{\beta}_0 - \tilde{\boldsymbol{x}} \tag{5.39g}$$

$$\Omega = \hat{\nu} + \Big(\hat{\boldsymbol{\nu}}^0\Big)^T \cdot Q_0 \cdot \hat{\boldsymbol{\nu}}^0 \tag{5.39h}$$

▷ The variables $\tilde{\boldsymbol{x}}$, $\hat{\nu}$, and $\hat{\boldsymbol{\nu}}^0$ are shown without iteration superscripts in this step, implying that the variables from the final iteration of step 2 are to be used.

Chapter 6: Numerical Applications

The main purpose of this chapter is to show, in a straight-forward way, how to apply key formulas from the earlier chapters, where the material is somewhat abstract in nature. Thus, the problems presented here are kept relatively basic in order to facilitate this objective. Experiments in 2-D line-fitting and 2-D similarity transformations show how to apply the algorithms presented in Chapters 2, 3, and 5 to typical problems in geodetic science. Indeed, 2-D line-fitting (and regression analysis in general) is an applied problem that arises in practically every branch of science and engineering as well as in the social sciences and in economics.

6.1 2-D line-fitting

Algorithms 1, 2, and 3 were applied by fitting a 2-D line to data presented by Neri et al. [1989], listed here in Table 6.1. Following Schaffrin and Wieser [2008], the EIV model for fitting a 2-D line with slope parameter ξ_1 and intercept parameter ξ_2 to n measured coordinate pairs $(x_i,\, y_i)$ is given by

$$y_i - e_{y_i} = \xi_1 \cdot (x_i - e_{x_i}) + \xi_2, \quad i = 1, \dots, n, \tag{6.1a}$$

$$\underset{n\times 1}{\boldsymbol{e}_y} := [e_{y_i}] \sim (\boldsymbol{0}, \sigma_0^2 Q_y),\ \underset{n\times 1}{\boldsymbol{e}_x} := [e_{x_i}] \sim (\boldsymbol{0}, \sigma_0^2 Q_x),\ C\{\boldsymbol{e}_y, \boldsymbol{e}_x\} = \sigma_0^2 Q_{yx}, \tag{6.1b}$$

or, in matrix form,

$$\boldsymbol{y} - \boldsymbol{e}_y = \Big(\underset{=:A}{[\boldsymbol{x},\, \boldsymbol{1}]} - \underset{=:E_A}{[\boldsymbol{e}_x,\, \boldsymbol{0}]}\Big)\boldsymbol{\xi} = \underbrace{[A - E_A]}_{n\times 2}\boldsymbol{\xi}, \tag{6.2a}$$

$$\begin{bmatrix} \boldsymbol{e}_y \\ \boldsymbol{e}_A := \operatorname{vec} E_A \end{bmatrix} \sim (\boldsymbol{0}, \sigma_0^2 Q). \tag{6.2b}$$

With $m = 2$, the $n(m+1) \times n(m+1)$ cofactor matrix Q is defined as

$$\underset{3n\times 3n}{Q} := \begin{bmatrix} \underset{n\times n}{Q_y} & Q_{yA} \\ Q_{Ay} & Q_A \end{bmatrix}, \tag{6.3a}$$

$$\text{with} \quad \underset{2n\times 2n}{Q_A} := \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \otimes \underset{n\times n}{Q_x} \;\text{ and }\; \underset{n\times 2n}{Q_{yA}} = \begin{bmatrix} Q_{yx}, & 0 \end{bmatrix}, \tag{6.3b}$$

$$\Rightarrow Q = \begin{array}{c} \left[\begin{array}{c|cc} Q_y & Q_{yx} & 0 \\ \hline Q_{xy} & Q_x & 0 \\ 0 & 0 & 0 \end{array}\right] \\ \begin{array}{ccc} n & \quad n & \quad n \end{array} \end{array} \begin{array}{c} n \\ n \\ n \\ \end{array} . \tag{6.3c}$$

The $n \times n$ matrix Q_y is the cofactor matrix for the dependent variables (y_i-coordinates), and the $n \times n$ matrix Q_x is the cofactor matrix for the independent variables (x_i-coordinates), whereas $Q_{xy} = Q_{yx}^T$ reflects the correlation between the x_i- and y_i-coordinates. The column and row of zeros in Q are due to the y-intercept parameter ξ_2, which has no measurement-variables associated with it, as reflected by the column of zeros in E_A.

The independently measured coordinates (x_i, y_i) and their associated weights $(1/\sigma_{x_i}^2, 1/\sigma_{y_i}^2)$ are listed together in Table 6.1, where the data are given without units. The complete cofactor matrix is constructed from the inverse of the weights shown in Table 6.1 as

$$\begin{bmatrix} Q_y & Q_{yx} \\ Q_{xy} & Q_x \end{bmatrix} = \begin{bmatrix} Q_y & 0 \\ 0 & Q_x \end{bmatrix} = \operatorname{Diag}\Big(\big[\sigma_{y_1}^2, \ldots, \sigma_{y_n}^2, \sigma_{x_1}^2, \ldots, \sigma_{x_n}^2\big]\Big), \quad n = 10, \tag{6.4}$$

which furnishes the nonzero blocks of the matrix Q in (6.3c).

Apparently the data presented by Neri et al. [1989] have been in use for quite some time. The coordinates were originally presented by Pearson [1901], while the weights were introduced by York [1966], who acknowledged that these may seem like "somewhat extreme conditions of weighting" but claimed that similar weights were encountered in his particular work. Though the weights may indeed seem "somewhat extreme" for geodetic applications, they are acceptable for the purposes of this study.

Several cases were investigated in which correlation was added to the diagonal cofactor matrix shown in (6.4). In each case the TLS solution was compared to the least-squares solution within the Gauss-Helmert model (GHM) generated from Algorithm 4. For sake of space, only two cases are reported here. The first case treats cross-correlation between the non-singular cofactor matrices Q_x and Q_y such that the $n \times n$ matrix Q_1 (defined in (2.10b)) is also non-singular. The second case treats both cross-correlation and singularities in Q_x and Q_y such that Q_1 is also singular.

In both cases the initial parameter approximations are taken from the column labeled GLS in Table 2 of Schaffrin and Wieser [2008], which is the generalized least-squares solution obtained by treating the x-coordinates as errorless and using the

Table 6.1: Neri's Data for 2-D line-fitting: coordinate pairs (x_i, y_i) and corresponding weights $(1/\sigma^2_{x_i}, 1/\sigma^2_{y_i})$. All values are considered unitless.

Point No.	x_i	$1/\sigma^2_{x_i}$	y_i	$1/\sigma^2_{y_i}$
1	0.0	1.0×10^3	5.9	1.0
2	0.9	1.0×10^3	5.4	1.8
3	1.8	5.0×10^2	4.4	4.0
4	2.6	8.0×10^2	4.6	8.0
5	3.3	2.0×10^2	3.5	2.0×10
6	4.4	8.0×10	3.7	2.0×10
7	5.2	6.0×10	2.8	7.0×10
8	6.1	2.0×10	2.8	7.0×10
9	6.5	1.8	2.4	1.0×10^2
10	7.4	1.0	1.5	5.0×10^2

weights $1/\sigma^2_{y_i}$ for the y-coordinates. The resulting initial values are $\xi^0_1 = -0.611$ for the slope parameter and $\xi^0_2 = 6.100$ for the y-intercept. The convergence criterion δ, which is required for the last step of the algorithms, was set to $\delta = 10^{-10}$. A *model check* was made for both cases, which ensures that

$$\sum_{i=1}^{n} \Big(y_i - \tilde{e}_{y_i} - \hat{\xi}_1 \cdot (x_i - \tilde{e}_{x_i}) - \hat{\xi}_2 \Big)^2 \approx 0, \tag{6.5}$$

where the approximation sign reflects the finiteness of machine precision and the fact that the convergence criterion δ is nonzero.

6.1.1 Case 1 – nonsingular cofactor matrices

In the first case, (pseudo-)random correlations between individual (x_i, y_i)-pairs were introduced so that the submatrices Q_{xy} and Q_{yx} are diagonal, while the submatrices Q_x and Q_y remain unchanged. Randomness over the interval $[-1, 1]$ was imposed by use of the MATLAB function *rand*. The resulting ten correlation coefficients $\rho_{x_i y_i}$, associated with the ten respective (x_i, y_i)-pairs, are listed in Table 6.2. The relationship between the correlation coefficients $\rho_{x_i y_i}$ and the cofactors $\sigma_{x_i y_i}$ is given by

$$\rho_{x_i y_i} = \frac{\sigma_{x_i y_i}}{\sigma_{x_i} \sigma_{y_i}}, \quad i = 1, \dots, n, \tag{6.6}$$

where obviously $\sigma_{x_i y_i}$ is the i-th diagonal element of both $n \times n$ matrices Q_{xy} and Q_{yx} and is also the value of the element $\sigma_{(n+i)i}$ in the matrix $Q = [\sigma_{ij}]$ defined in (6.3c).

Table 6.2: Correlation coefficients for correlated xy-pairs

Variables	Numerical values				
$\rho_{x_1y_1} \ldots \rho_{x_5y_5}$	-0.165956	0.440649	-0.999771	-0.395335	-0.706488
$\rho_{x_6y_6} \ldots \rho_{x_{10}y_{10}}$	-0.815323	-0.627480	-0.308879	-0.206465	0.077633

In this case, the diagonal cofactor matrices Q_x and Q_y and the diagonal cross-cofactor matrices $Q_{xy} = Q_{yx}$ are all non-singular, and thus Algorithms 1 and 2 are employed. Both algorithms generated precisely the same solution, which is shown in Table 6.3. The only difference in their behavior was that Algorithm 1 required nine iterations to converge, whereas Algorithm 2 required 13. However, this does not imply that Algorithm 1 is more efficient than Algorithm 2 in general, as several other variations of the cofactor matrices revealed that sometimes Algorithm 1 took a few less iterations than Algorithm 2, and sometimes it was the other way around.

Table 6.3: Parameter estimates and TSSR for fitted 2-D line – Case 1

Parameter	Estimate
slope	-0.45922870
y-intercept	5.35727267
TSSR	16.72548303

The (weighted) TSSR shown in the last line of Table 6.3 was computed using (2.21c). The model check according to (6.5) was zero within machine precision. The solutions from both TLS algorithms also agree with the least-squares solution within the Gauss-Helmert model (LESS-GHM), computed using Algorithm 4, which converged in 16 iterations. Note that the number of digits reported for the parameter estimates is not meant to reflect their precisions but is merely shown for comparison purposes, in case the reader may want to experiment and compare with other estimators, for example.

The fitted line is plotted in Figure 6.1, where the projections of the data points onto the fitted line are shown by solid red lines labeled "residuals." Had the data been treated as iid, all of these projections would have been perpendicular to the fitted line (orthogonal regression). Had the x-coordinates been treated as errorless, and the errors in y-coordinates as iid, the projections would have been parallel to the y-axis (ordinary least-squares). Schaffrin and Wieser [2008] showed a similar graph for their fitted line, which was based on the uncorrelated weights listed in Table 6.1. The following quote from their paper (p. 419) applies here as well:

> *The plot clearly shows the impact of the WTLS adjustment where each observed point is projected onto the regression line along a direction resulting from the ratio of standard errors.*

See also Gerhold [1969, footnote number 2]. It is noted that the correlation coefficients will also play a role in the case studied here.

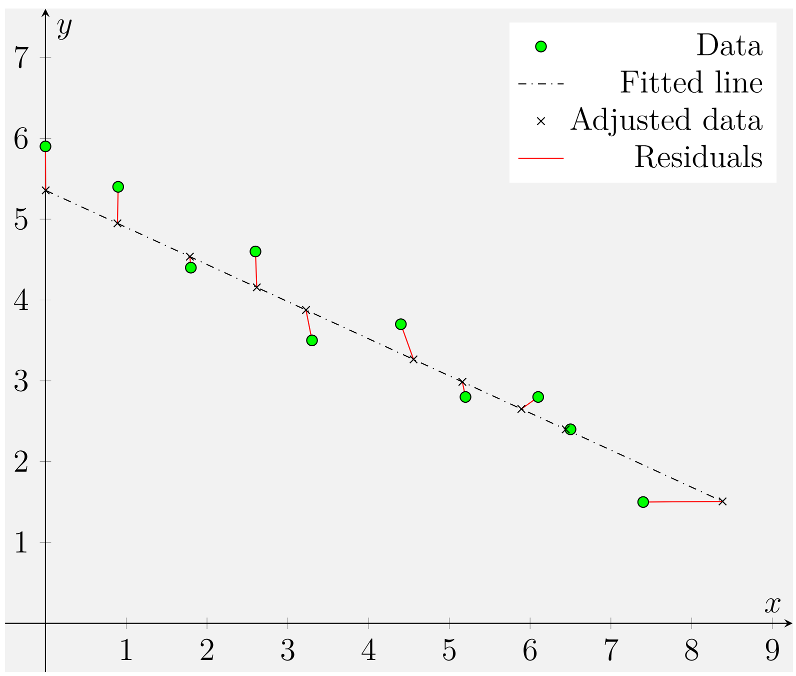


Figure 6.1: Fitted 2-D line – Case 1

6.1.2 Case 2 – singular cofactor matrices

To generate a cofactor matrix $\begin{bmatrix} Q_y & Q_{yx} \\ Q_{xy} & Q_x \end{bmatrix}$ that is both full and singular, thereby causing the matrix Q_1 (as defined in (2.10b)) to be singular, correlation coefficients were computed from the residual dispersion matrix (4.7c) of the least-squares solution within the Gauss-Helmert model when treating the coordinate-data as iid. Since this dispersion matrix turns out to be singular, using the correlation coefficients derived from it, together with the weights in table Table 6.1, will generate a singular cofactor matrix $\begin{bmatrix} Q_y & Q_{yx} \\ Q_{xy} & Q_x \end{bmatrix}$, resulting in a singular matrix Q_1, too. The numerical values for the correlation coefficients are listed in Appendix A.

Of the three algorithms given in Chapters 2 and 3, only Algorithm 3 can handle a singular matrix Q_1, provided the Neitzel/Schaffrin rank condition is satisfied. (See (3.6b) and Appendix C for details.) Owing to the recent work by Neitzel and Schaffrin [2012], the solution can also be compared to the least-squares solution within the Gauss-Helmert model with a singular dispersion matrix.

Table 6.4 shows the solution yielded by Algorithm 3, which converged in five iterations. The (weighted) TSSR was computed using (2.21c). The model check according to (6.5) was zero within machine precision. As noted in the previous section, the number of digits shown for the estimates are not all significant but are shown for comparison purposes, should anyone want to compare these results to solutions based on other methods. The TLS solution generated by Algorithm 3 agrees precisely with the least-squares solution within the Gauss-Helmert model with a singular dispersion matrix, which also converged in five iterations.

Table 6.4: Parameter estimates and TSSR for fitted 2-D line – Case 2

Parameter	Estimate
slope	-0.49317262
y-intercept	5.54275204
TSSR	6.04625953

The fitted line is plotted in Figure 6.2, where the projections of the data points onto the fitted line are shown by solid red lines labeled "residuals." See the last paragraph of the preceding section for further discussion about the graph of the fitted line.

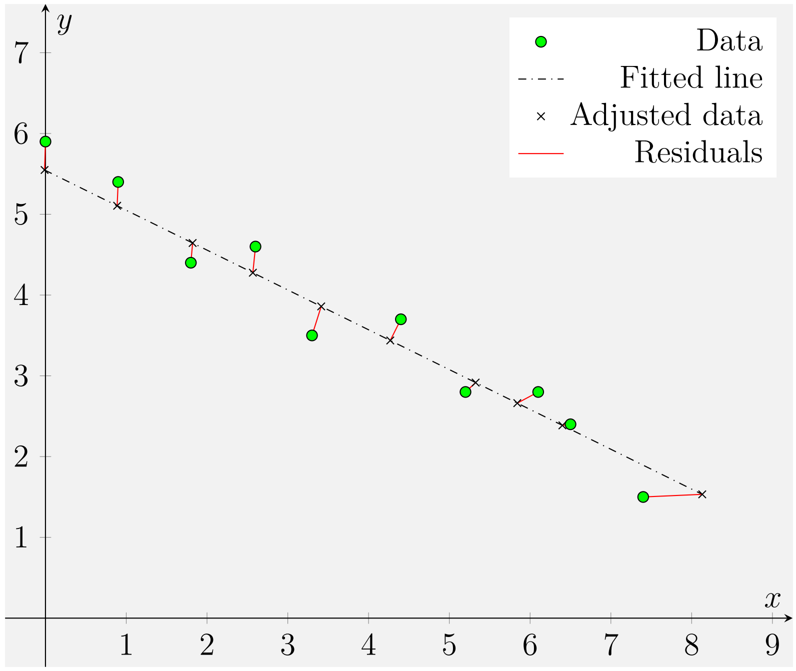


Figure 6.2: Fitted 2-D line – Case 2

6.1.3 Equivalence between Q_1 and BQB^T

Here the equivalence between BQB^T (LESS-GHM) and Q_1 (WTLSS-EIV) is shown in the context of 2-D line-fitting. Refer to Chapter 4 for a more general discussion.

The matrix Q_1 was defined earlier as

$$Q_1 := Q_y - Q_{yA}\big(\hat{\boldsymbol{\xi}} \otimes I_n\big) - \big(\hat{\boldsymbol{\xi}} \otimes I_n\big)^T Q_{Ay} + \big(\hat{\boldsymbol{\xi}} \otimes I_n\big)^T Q_A \big(\hat{\boldsymbol{\xi}} \otimes I_n\big).$$

Here

$$Q_A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \otimes Q_x \quad \text{and} \quad Q_{yA} = \begin{bmatrix} Q_{yx}, & 0 \end{bmatrix} = Q_{Ay}^T.$$

The vector $\hat{\boldsymbol{\xi}} \approx \boldsymbol{\xi}^0 = \begin{bmatrix} \xi_1^0, & \xi_2^0 \end{bmatrix}^T$ is used for an initial approximation of the slope and intercept values, respectively, which leads to

$$\begin{aligned} Q_1 = Q_y - \begin{bmatrix} Q_{yx}, & 0 \end{bmatrix} \begin{bmatrix} \xi_1^0 I_n \\ \xi_2^0 I_n \end{bmatrix} - \begin{bmatrix} \xi_1^0 I_n, & \xi_2^0 I_n \end{bmatrix} \begin{bmatrix} Q_{xy} \\ 0 \end{bmatrix} + \\ + \begin{bmatrix} \xi_1^0 I_n, & \xi_2^0 I_n \end{bmatrix} \begin{bmatrix} Q_x & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \xi_1^0 I_n \\ \xi_2^0 I_n \end{bmatrix} = \\ = Q_y - \xi_1^0 Q_{yx} - \xi_1^0 Q_{xy} + \xi_1^0 Q_x \xi_1^0. \end{aligned}$$

If the same initial approximation vector $\boldsymbol{\xi}^0$ is used in the iterated LESS within the GHM, it follows that

$$
\begin{aligned}
BQB^T &= \begin{bmatrix} -\xi_1^0 I_n, & I_n \end{bmatrix} \begin{bmatrix} Q_x & Q_{xy} \\ Q_{yx} & Q_y \end{bmatrix} \begin{bmatrix} -\xi_1^0 I_n \\ I_n \end{bmatrix} = \\
&= \xi_1^0 Q_x \xi_1^0 - \xi_1^0 Q_{yx} - \xi_1^0 Q_{xy} + Q_y.
\end{aligned}
$$

Obviously Q_1 and BQB^T are numerically equivalent when evaluated at the same approximate value ξ_1^0 for the slope parameter. This agrees with the more general development of §4.2. Note that the intercept parameter ξ_2^0 does not factor into the equations, which is as expected since there are no measurement variables associated with it.

6.2 2-D similarity transformation

In the 2-D similarity transformation problem, four parameters are estimated for the purpose of transforming coordinates from a source system (here labeled xy-coordinate system) to a target system (here labeled XY-coordinate system). The estimation requires redundant data consisting of observed (or previously estimated) coordinates in both systems at common reference points, together with the associated cofactor matrices Q_{xy} and Q_{XY} for the source and target systems, respectively. Since the observed coordinates and their cofactor matrices typically come from different sources, it is assumed here that there is no cross-correlation between them.

The four parameters for the 2-D similarity transformation are

$$
\begin{array}{cl}
\xi_0,\ \xi_1 & \text{for the translation of the coordinate frame,} \\
\alpha & \text{for the rotation angle,} \\
\omega & \text{for the scale factor.}
\end{array}
$$

To transform the problem into a (quasi)linear one, two additional intermediary parameters are defined as $\xi_2 := \omega \cos\alpha$ and $\xi_3 := \omega \sin\alpha$. The vector of unknown parameters to estimate is then $\boldsymbol{\xi} = \begin{bmatrix} \xi_0, & \xi_1, & \xi_2, & \xi_3 \end{bmatrix}^T$.

The EIV model for the 2-D similarity transformation, with $n/2$ pairs of observed points in both source and target coordinate systems, is given by

$$
\underset{n\times 1}{\boldsymbol{y}} := \begin{bmatrix} X_1 \\ Y_1 \\ \dots \\ X_{n/2} \\ Y_{n/2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & x_1 & -y_1 \\ 0 & 1 & y_1 & x_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & x_{n/2} & -y_{n/2} \\ 0 & 1 & y_{n/2} & x_{n/2} \end{bmatrix} \begin{bmatrix} \xi_0 \\ \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} - \begin{bmatrix} 0 & 0 & e_{x_1} & -e_{y_1} \\ 0 & 0 & e_{y_1} & e_{x_1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & e_{x_{n/2}} & -e_{y_{n/2}} \\ 0 & 0 & e_{y_{n/2}} & e_{x_{n/2}} \end{bmatrix} \begin{bmatrix} \xi_0 \\ \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} + \begin{bmatrix} e_{X_1} \\ e_{Y_1} \\ \vdots \\ e_{X_{n/2}} \\ e_{Y_{n/2}} \end{bmatrix} = \tag{6.7a}
$$

$$= \underset{n\times m}{\Big[A - E_A\Big]} \cdot \boldsymbol{\xi} + \boldsymbol{e}_y, \quad \text{with} \quad \operatorname{rk} A = m = 4. \tag{6.7b}$$

The random errors are distributed as

$$\begin{bmatrix}\boldsymbol{e}_y\\ \boldsymbol{e}_A\end{bmatrix} := \begin{bmatrix}\boldsymbol{e}_y\\ \operatorname{vec} E_A\end{bmatrix} \sim \left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \sigma_0^2\begin{bmatrix}Q_y & 0\\ 0 & Q_A\end{bmatrix}\right), \tag{6.7c}$$

with

$$\underset{n\times n}{Q_y} := Q_{XY} \quad \text{and} \quad \underset{nm\times nm}{Q_A} := \begin{bmatrix}0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & Q_{A_{33}} & Q_{A_{34}}\\ 0 & 0 & Q_{A_{43}} & Q_{A_{44}}\end{bmatrix}. \tag{6.7d}$$

The relationship between the nonzero blocks of Q_A and the cofactor matrix Q_{xy} from the source coordinate-system is determined as follows: define a 2×2 block-diagonal transformation matrix T of dimension $n\times n$ such that $\boldsymbol{a}_4 = T\boldsymbol{a}_3$, where $\boldsymbol{a}_3$ and $\boldsymbol{a}_4$ are the third and fourth columns, respectively, of the data matrix $A = \Big[\boldsymbol{a}_1,\ \boldsymbol{a}_2,\ \boldsymbol{a}_3,\ \boldsymbol{a}_4\Big]$. The matrix T is then given by

$$\underset{n\times n}{T} := \operatorname{Diag}(T', \ldots, T'), \quad \text{with} \quad T' := \begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}, \tag{6.8}$$

where the matrix T' obviously occurs $n/2$ times in the Diag argument. Note that T is orthonormal, and thus $T^T = T^{-1}$ and $(T^T)^{-1} = T$.

Applying the law of variance propagation leads to the following expressions for the non-zero blocks of Q_A in terms of Q_{xy}:

$$\underset{n\times n}{Q_{A_{33}}} = Q_{xy} = Q_{xy}^T,\ \underset{n\times n}{Q_{A_{34}}} = Q_{xy}T^T,\ \underset{n\times n}{Q_{A_{43}}} = TQ_{xy},\ \underset{n\times n}{Q_{A_{44}}} = TQ_{xy}T^T. \tag{6.9}$$

Now the matrix of combined cofactors

$$Q_1 = Q_y - Q_{yA}\Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big) - \Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)^T Q_{Ay} + \Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big)^T Q_A\Big(\hat{\boldsymbol{\xi}} \otimes I_n\Big) \tag{6.10}$$

can be readily expressed in terms of Q_{XY} and Q_{xy}. Note that due to the first two rows and columns of zeros in Q_A, the last matrix product in Q_1, evaluated at an approximate value $\boldsymbol{\xi}^0$, reduces to

$$\begin{aligned}
&\Big(\boldsymbol{\xi}^0 \otimes I_n\Big)^T Q_A \times \Big(\boldsymbol{\xi}^0 \otimes I_n\Big) = \\
&= \left(\Big[\xi_2^0,\quad \xi_3^0\Big] \otimes I_n\right)\begin{bmatrix}Q_{xy} & Q_{xy}T^T\\ TQ_{xy} & TQ_{xy}T^T\end{bmatrix}\left(\begin{bmatrix}\xi_2^0\\ \xi_3^0\end{bmatrix} \otimes I_n\right) = \\
&= \Big[\xi_2^0 I_n,\quad \xi_3^0 I_n\Big]\begin{bmatrix}I_n\\ T\end{bmatrix} Q_{xy}\Big[I_n,\quad T^T\Big]\begin{bmatrix}\xi_2^0 I_n\\ \xi_3^0 I_n\end{bmatrix} = \\
&= \Big(\xi_2^0 I_n + \xi_3^0 T\Big)Q_{xy}\Big(\xi_2^0 I_n + \xi_3^0 T^T\Big).
\end{aligned} \tag{6.11}$$

Thus we have

$$Q_1^0 = Q_{XY} + \xi_2^0 Q_{xy}\xi_2^0 + \xi_3^0 T Q_{xy}\xi_2^0 + \xi_2^0 Q_{xy}T^T\xi_3^0 + \xi_3^0 T Q_{xy} T^T \xi_3^0. \tag{6.12}$$

6.2.1 Equivalence between Q_1 and BQB^T

Here a comparison between Q_1 and the matrix product BQB^T from the LESS-GHM is made to verify the more general development in section 4.2. In the LESS-GHM, the matrices Q and BQB^T are defined as

$$\underset{2n\times 2n}{Q} := \begin{bmatrix} Q_{XY} & 0 \\ 0 & Q_{xy} \end{bmatrix}, \quad \text{and} \quad \underset{n\times 2n}{B^0} = \left[I_n, \quad -\left[\xi_2^0 I_n,\, \xi_3^0 I_n\right] \begin{bmatrix} I_n \\ T \end{bmatrix} \right],$$

so that

$$\underbrace{B^0 Q (B^0)^T}_{\text{LESS-GHM}} = Q_{XY} + \left(\xi_2^0 I_n + \xi_3^0 T\right) Q_{xy} \left(\xi_2^0 I_n + \xi_3^0 T^T\right) = \underbrace{Q_1^0}_{\text{WTLSS-EIV}},$$

provided that both matrices are evaluated at the same values for the approximate parameters. This agrees with the more general development of §4.2.

6.3 2-D similarity transformation with singular cofactor matrices

The data for the 2-D similarity transformation were provided by Professor Frank Neitzel. They are comprised of 2-D coordinates of five stations from both the source and target systems, together with their associated cofactor matrices, both of which are fully populated and singular. It is noted again that the source and target data are not correlated with each other. The coordinates are listed in Table 6.5, whereas the cofactor matrices are listed in Appendix A.2.

Table 6.5: Coordinate estimates in source and target systems

Point No.	x_i[m]	y_i[m]	X_i[m]	Y_i[m]
1	453.8001	137.6099	400.0040	100.0072
2	521.2865	350.7972	500.0019	299.9994
3	406.8728	433.9247	399.9925	399.9933
4	110.5545	386.9880	100.0059	400.0022
5	157.4861	90.6802	99.9956	99.9978

Both 10×10 source and target cofactor matrices Q_{xy} and Q_{XY}, respectively, have the rank of seven, and, when incorporated into the 10×10 matrix Q_1 of (6.12), the resulting matrix Q_1 is found to have rank eight. However, the Neitzel/Schaffrin rank condition is still satisfied. (See (3.6b) and Appendix C for details.) Thus the problem is solved with Algorithm 3, as Algorithms 1 and 2 cannot treat the singularity of Q_1.

Step 1 of Algorithm 3 requires an initial approximation for the parameter vector $\boldsymbol{\xi}$. As explained by Neitzel and Schaffrin [2012], initial approximations can be computed by a traditional approach (using ordinary least-squares) where Q_{xy} is replaced by 0 and Q_{XY} by I_n. The resulting approximate values are

$$\boldsymbol{\xi}^0 = [-69.73, 35.08, 0.988, -0.156]^T.$$

The above problem is also solved by the least-squares solution within the Gauss-Helmert model (LESS-GMH) for comparison purposes. The LESS-GHM permits computation of the variances of the parameter estimates. Note that the variances of the estimated shift parameters $\hat{\xi}_0$ and $\hat{\xi}_1$ can be taken directly from the dispersion matrix (4.7a) computed as a byproduct of the LESS-GHM, whereas the variances of the orientation and scale parameters are computed as a function of the variances of the estimated intermediate parameters $\hat{\xi}_2$ and $\hat{\xi}_3$. The variance formulas for the orientation and scale parameters are shown below, where for simplicity the variables are written in a generic form without the use of hat symbols.

Scale factor:

$$\omega = \sqrt{\xi_2^2 + \xi_3^2}$$
$$\Rightarrow \sigma_\omega^2 = \frac{\xi_2^2\sigma_{\xi_2}^2 + \xi_3^2\sigma_{\xi_3}^2 + 2\cdot\xi_2\xi_3\sigma_{\xi_2\xi_3}}{\xi_2^2 + \xi_3^2} \tag{6.13}$$

Orientation:

$$\alpha = \arctan(\xi_3/\xi_2)$$
$$\Rightarrow \sigma_\alpha^2 = \frac{1}{\omega^4}\Big(\xi_3^2\cdot\sigma_{\xi_2}^2 - 2\xi_2\xi_3\cdot\sigma_{\xi_2\xi_3} + \xi_2^2\cdot\sigma_{\xi_3}^2\Big) \tag{6.14}$$

Now using hats to represent estimates, the empirical standard deviations of the orientation and scale are computed by $\sqrt{\hat{\sigma}_0^2\cdot\sigma_\alpha^2}$ and $\sqrt{\hat{\sigma}_0^2\cdot\sigma_\omega^2}$, respectively, where $\hat{\sigma}_0^2$ is the estimated variance component computed according to (4.8).

Using the initial approximation $\boldsymbol{\xi}^0$ given above, and a convergence tolerance $\delta = 1.0\times 10^{-12}$ for (3.34g), the WTLSS-EIV converged in three iterations, whereas the LESS-GHM took four. Tables 6.6 and 6.7 show the WTLSS-EIV estimated parameters and predicted residuals, respectively. The standard deviations computed from the LESS-GHM are also shown. A TSSR of $\Omega = 6.1640345$ was computed using (2.21c). Considering the system redundancy of $r = 6$, the estimated variance component is then computed using (2.22) as $\hat{\sigma}_0^2 = \Omega/r = (1.0135774)^2 = 1.027339$.

It was found that the WTLSS-EIV generated from Algorithm 3 is the same as the LESS-GHM generated from Algorithm 4, to the precision shown in Table 6.6 and Table 6.7, for both the estimated parameters and the predicted errors.

Table 6.6: WTLS-EIV solution for the 2-D similarity transformation with standard deviations computed from the LESS-GHM

Parameter	Estimated value	Std dev
x-shift ξ_0	-69.726354 m	$\pm$4.090 mm
y-shift ξ_1	35.078215 m	$\pm$2.488 mm
$\xi_2 = \omega \cos \alpha$	0.98765502	$\pm 1.093 \times 10^{-5}$
$\xi_3 = \omega \sin \alpha$	-0.15642921	$\pm 1.730 \times 10^{-6}$
scale factor ω	0.99996626	$\pm 1.091 \times 10^{-5}$
rotation angle α	-10.00000154 gon	$\pm 1.427 \times 10^{-5}$ mgon
var. component σ_0^2	1.027339	$\pm \sigma_0^2 \sqrt{2/r}$

Table 6.7: 2-D similarity transformation residuals predicted by WTLS-EIV

Point	Target System		Source System	
	$\tilde{e}_Y$ [mm]	$\tilde{e}_X$ [mm]	$\tilde{e}_y$ [mm]	$\tilde{e}_x$ [mm]
1	0.8998	1.0204	-5.3231	-4.4026
2	-0.1634	0.3453	0.5454	-1.8617
3	-0.9923	-1.5805	6.2318	7.1386
4	1.2009	1.0399	-6.8490	-4.2616
5	-0.9450	-0.8250	5.3948	3.3873

In addition to the tabulated residuals in Table 6.7, the *total* predicted error matrix

$$\left[\ \tilde{\boldsymbol{e}}_y \mid \tilde{E}_A\ \right] = \left[\begin{array}{r|rr rr}
1.0204 & 0 & 0 & -4.4026 & 5.3231 \\
0.8998 & 0 & 0 & -5.3231 & -4.4026 \\
0.3453 & 0 & 0 & -1.8617 & -0.5454 \\
-0.1634 & 0 & 0 & 0.5454 & -1.8617 \\
-1.5805 & 0 & 0 & 7.1386 & -6.2318 \\
-0.9923 & 0 & 0 & 6.2318 & 7.1386 \\
1.0399 & 0 & 0 & -4.2616 & 6.8490 \\
1.2009 & 0 & 0 & -6.8490 & -4.2616 \\
-0.8250 & 0 & 0 & 3.3873 & -5.3948 \\
-0.9450 & 0 & 0 & 5.3948 & 3.3873
\end{array}\right] \text{mm}$$

reveals interesting features of the WTLS-EIV algorithm. A comparison between this matrix and equation (6.7a) shows that the structure of the data matrix A has been replicated exactly in the residual matrix $\tilde{E}_A$. The first two columns of both matrices contain only zeros. The structure of the last two columns of $\tilde{E}_A$ is highlighted by drawing a box around the first two rows. This replication of structure in the residual matrix had already been pointed out by Fang [2011], Mahboub [2012], and Schaffrin et al. [2012a] for EIV models with cofactor matrices having full rank. The property holds here as well in the new estimator that handles rank-deficient cofactor matrices.

The points are plotted in a 2-D map in Figure 6.3, where the dotted lines represent a grid for the adjusted coordinates $(\boldsymbol{x}-\tilde{\boldsymbol{e}}_x, \boldsymbol{y}-\tilde{\boldsymbol{e}}_y)$ in the source system, and the dash-dotted lines represent a (rotated) grid for the adjusted coordinates $(\boldsymbol{X}-\tilde{\boldsymbol{e}}_X, \boldsymbol{Y}-\tilde{\boldsymbol{e}}_Y)$ in the target system. The origin of the source system has coordinates $(\hat{\xi}_0, \hat{\xi}_1)$ in the target system.

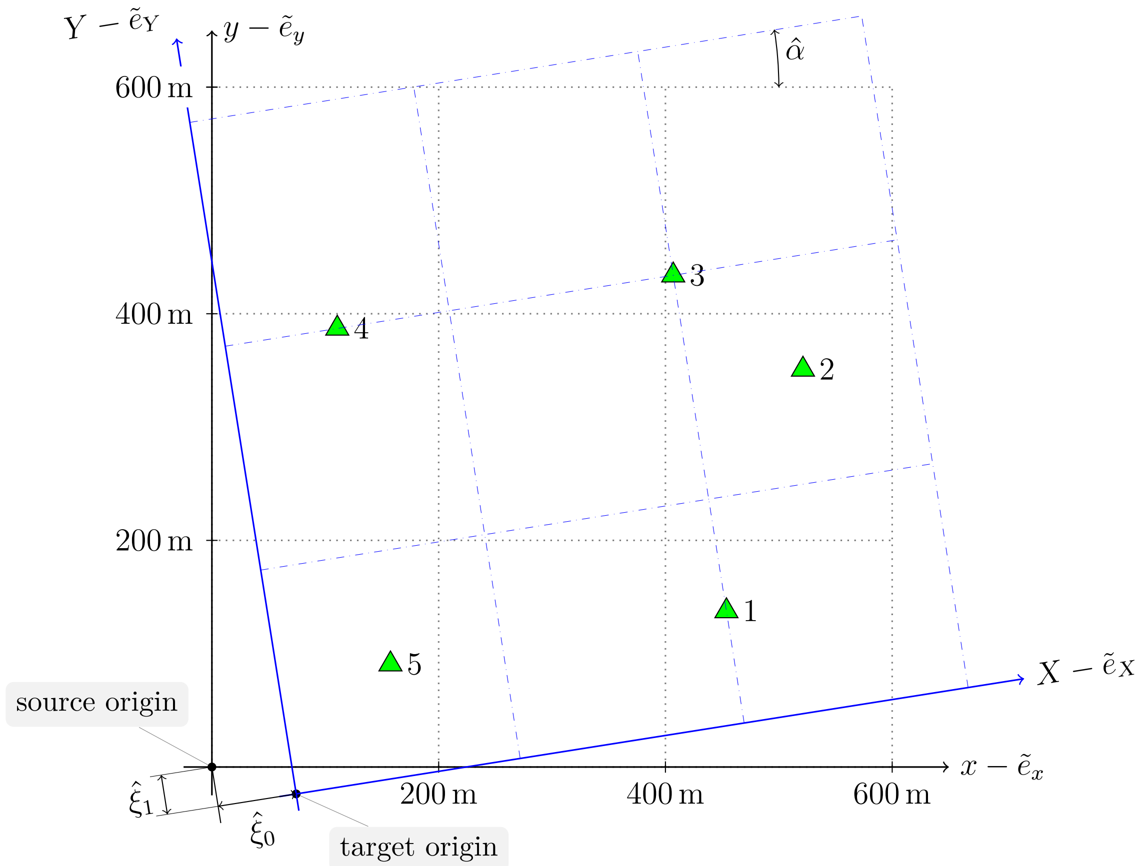


Figure 6.3: Map view of the five data points in the source and target systems. The dotted lines represent a grid for the adjusted coordinates $(\boldsymbol{x} - \tilde{\boldsymbol{e}}_x, \boldsymbol{y} - \tilde{\boldsymbol{e}}_y)$ in the source system, and the dash-dotted lines represent a (rotated) grid for the adjusted coordinates $(\boldsymbol{X} - \tilde{\boldsymbol{e}}_X, \boldsymbol{Y} - \tilde{\boldsymbol{e}}_Y)$ in the target system. The origin of the source system has coordinates $(\hat{\xi}_0, \hat{\xi}_1)$ in the target system. The grid interval for both grids is 200 meters.

6.4 2-D line-fitting by WTLS collocation

In this section, prior information is introduced for the parameters of a 2-D line in order to experiment with WTLSC within the EIV-REM model developed in Chapter 5. The data from §6.1 are used, and the errors are treated as heteroscadastic by adopting the weights from Table 6.1 without adding correlation. Thus, the matrix Q_1 of equation (6.12) is nonsingular.

In the absence of actual prior information, the information must be generated somehow for experimental purposes. In practice, prior information often comes from a previous estimation task based on previously observed data. Then, often after the passage of some time, new observations are made that need to be integrated with the old solution, which now becomes the "prior information." For this experiment, the estimates, and their covariance matrix, from the generalized least-squares estimator (GLSE) are adopted as prior information. Here, as in Rao et al. [2008], GLSE means the least-squares estimator within the Gauss-Markov model (GMM) that treats only the y-variables of Table 6.1 as random (making use of the weights $1/\sigma_{y_i}^2$ from Table 6.1) and considers the x-variables as errorless. The key equations for generating the prior information are summarized below.

Formulas for generating simulated prior information: GLSE within the GMM

$$\boldsymbol{y} = \underset{n\times m}{A}\,\boldsymbol{\xi} + \boldsymbol{e}_y,\ \operatorname{rk} A = m, \text{ and} \qquad \boldsymbol{e}_y \sim (\boldsymbol{0}, \sigma_0^2 P^{-1} = \sigma_0^2 Q_y) \tag{6.15a}$$

$$[N,\, \boldsymbol{c}] = A^T P[A, \boldsymbol{y}] \tag{6.15b}$$

$$\hat{\boldsymbol{\xi}} = N^{-1}\boldsymbol{c} \tag{6.15c}$$

$$\tilde{\boldsymbol{e}}_y = \boldsymbol{y} - A\hat{\boldsymbol{\xi}} \tag{6.15d}$$

$$\hat{\sigma}_0^2 = (n-m)^{-1} \cdot (\tilde{\boldsymbol{e}}_y^T P \tilde{\boldsymbol{e}}_y) \tag{6.15e}$$

$$\hat{D}\{\hat{\boldsymbol{\xi}}\} = \hat{\sigma}_0^2 N^{-1} \tag{6.15f}$$

The estimates computed from (6.15c) are assigned to $\boldsymbol{\beta}_0$, and the estimated dispersion matrix computed from (6.15f) is assigned to Q_0. The numerical values are:

$$\beta_0 = \begin{bmatrix} -0.610812957 \\ 6.100109317 \end{bmatrix},$$

$$Q_0 = (2.07199202153158)^2 \cdot \begin{bmatrix} 0.000905254577531 & -0.006064590624825 \\ -0.006064590624825 & 0.041886814963206 \end{bmatrix}.$$

The large number of digits is shown only for the sake of reproducing the experiment without significant loss of computing precision. The solution is in agreement with the GLS solution of Schaffrin and Wieser [2008, see Table 2].

6.4.1 Experiments with scaled heteroscedastic weights

It would be interesting to vary the influence of the prior information from an extreme of having virtually no effect to an extreme of completely dominating the solution, which can be done by introducing a scale factor s for the matrix Q_0, such that $Q_0 \to s \cdot Q_0$. The results based on a range of values from $s = 1.0 \times 10^{-10}$ to $s = 1.0 \times 10^{10}$, with an increase by a factor of 10 at each step, are shown in Table 6.8. Of course, such a variation of Q_0 has no practical value but does serve as a certain level of validation for the predictor derived in Chapter 5. The value of $s = 1$ represents the solution that would be adopted in practice.

As expected, the scaling of the cofactor matrix Q_0 by an extremely small number, thereby greatly magnifying the precision of the prior information, results in a solution that reproduces precisely the prior information. See the first row of Table 6.8. On the other hand, applying an extremely large scale factor to Q_0 completely eliminates the effect of the prior information, which is made apparent by comparing the last row of Table 6.8 to the WTLS solution of Schaffrin and Wieser [2008, see Table 2] or to the results of Neri et al. [1989, see Table 2]. Figure 6.4 portrays the transition of the slope predictions across the range of scale factors s. Most of the change occurs between $s = 1.0 \times 10^{-2}$ and $s = 1.0 \times 10^{2}$, which is apparent also from the table. Finally, it is noted that the total sum of squared residuals (TSSR) is smallest when the prior information is effectively eliminated. The TSSR was computed by use of (5.26a).

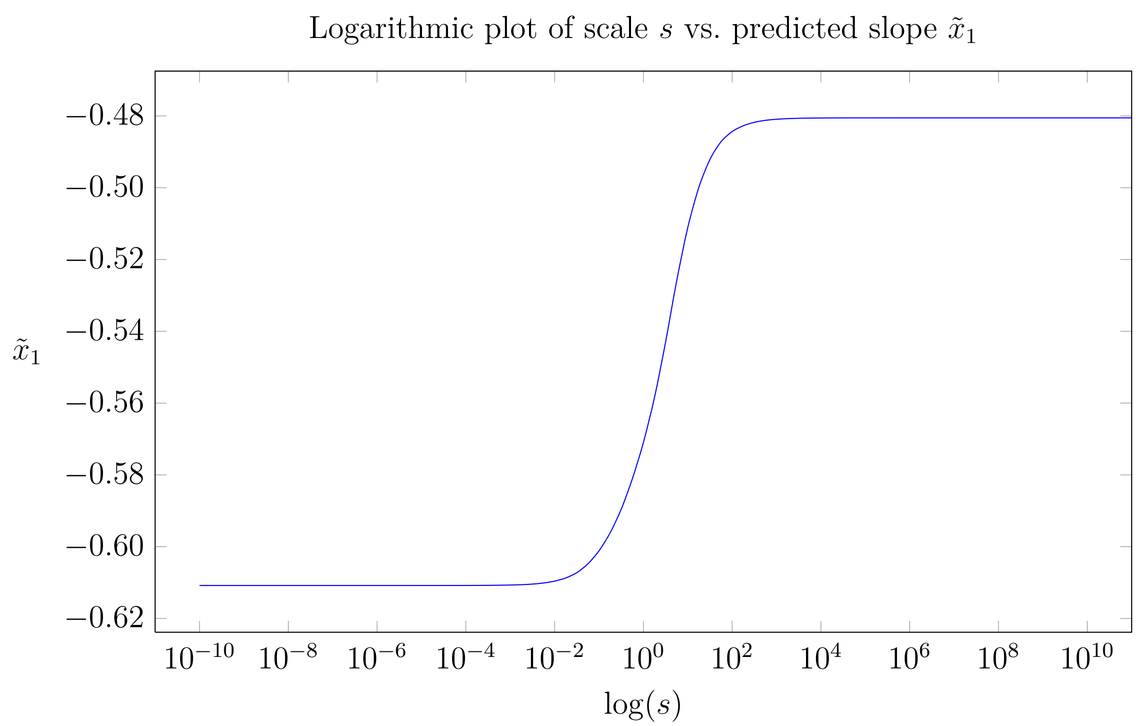


Figure 6.4: Predicted slopes as a function of the scaled cofactor matrix $Q_0 \to s \cdot Q_0$

Another experiment that demonstrates the versatility of the WTLSC predictor is conducted by scaling the cofactor matrix Q_0 of the a-priori information by a large

Table 6.8: WTLSC predictions of slope and intercept parameters with various cofactor matrices $Q_0 \to s \cdot Q_0$. N is the number of iterations. The formal model redundancy is $r = 10$.

s	N	slope	intercept	$TSSR = \Omega$	$\sqrt{\hat{\sigma}_0^2} = \sqrt{\Omega/r}$
1.0×10^{-10}	2	-0.610812957	6.100109317	16.285265299	1.276137348
1.0×10^{-9}	3	-0.610812956	6.100109316	16.285265296	1.276137347
1.0×10^{-8}	3	-0.610812955	6.100109308	16.285265259	1.276137346
1.0×10^{-7}	3	-0.610812944	6.100109233	16.285264895	1.276137332
1.0×10^{-6}	3	-0.610812831	6.100108475	16.285261255	1.276137189
1.0×10^{-5}	3	-0.610811703	6.100100904	16.285224851	1.276135763
1.0×10^{-4}	4	-0.610800423	6.100025213	16.284860928	1.276121504
1.0×10^{-3}	5	-0.610687971	6.099270801	16.281232946	1.275979347
1.0×10^{-2}	6	-0.609596967	6.091968014	16.246038438	1.274599484
1.0×10^{-1}	8	-0.601147853	6.036638197	15.973709326	1.263871407
1.0	10	-0.570982196	5.868134044	15.007054498	1.225032836
1.0×10^{1}	12	-0.511074244	5.606976283	13.028815960	1.141438389
1.0×10^{2}	12	-0.484355500	5.495727556	12.022301584	1.096462566
1.0×10^{3}	13	-0.480924351	5.481527013	11.882471781	1.090067511
1.0×10^{4}	13	-0.480572590	5.480072256	11.867970447	1.089402150
1.0×10^{5}	13	-0.480537327	5.479926431	11.866514973	1.089335347
1.0×10^{6}	13	-0.480533799	5.479911845	11.866369373	1.089328664
1.0×10^{7}	13	-0.480533447	5.479910386	11.866354812	1.089327995
1.0×10^{8}	13	-0.480533411	5.479910240	11.866353356	1.089327928
1.0×10^{9}	13	-0.480533408	5.479910226	11.866353210	1.089327922
1.0×10^{10}	13	-0.480533407	5.479910224	11.866353196	1.089327921

value and then computing solutions over a range of scaled matrices Q_x, being the cofactor matrix for the independent variables x_i (x-coordinates). This effectively eliminates the influence of the prior information while allowing the weights $1/\sigma_x^2$ to range from having no influence to having their expected influence. In the former case, we obtain the generalized least-squares solution with weights $1/\sigma_x^2$ having no effect and weights $1/\sigma_y^2$ having their full effect. In the latter case, we obtain the WTLSS of Schaffrin and Wieser [2008]. These results are tabulated in Table 6.9, where the first and last rows can be compared to the GLS, resp. WTLSS columns of Table 2 in Schaffrin and Wieser [2008]. Row 1 can also be compared to the generalized least-squares solution computed at the beginning of this section.

The TSSR was computed by use of (5.26a). As expected, the smallest TSSR occurs when errors in all coordinates are minimized, which is shown in the last row of Table 6.9. A plot of slopes versus scale factor s is shown in Figure 6.5.

Table 6.9: WTLSC predictions of slope and intercept parameters with $Q_0 \to 1.0 \times 10^{12} \cdot Q_0$ and various cofactor matrices $Q_x \to s \cdot Q_x$. The formal model redundancy is $r = 10$.

s	slope	intercept	$TSSR = \Omega$	$\sqrt{\hat{\sigma}_0^2} = \sqrt{\Omega/r}$
1.0×10^{-11}	-0.610812957	6.100109317	34.345207491	1.853246003
1.0×10^{-10}	-0.610812956	6.100109315	34.345207422	1.853246002
1.0×10^{-9}	-0.610812953	6.100109300	34.345206731	1.853245983
1.0×10^{-8}	-0.610812919	6.100109154	34.345199827	1.853245797
1.0×10^{-7}	-0.610812585	6.100107691	34.345130789	1.853243934
1.0×10^{-6}	-0.610809238	6.100093057	34.344440439	1.853225308
1.0×10^{-5}	-0.610775789	6.099946774	34.337540506	1.853039139
1.0×10^{-4}	-0.610442812	6.098489654	34.268895335	1.851185980
1.0×10^{-3}	-0.607255779	6.084461488	33.615765206	1.833460259
1.0×10^{-2}	-0.584516082	5.980894722	29.170989937	1.707951695
1.0×10^{-1}	-0.522030153	5.681869418	19.079158770	1.381273281
1.0	-0.480533407	5.479910224	11.866353194	1.089327921

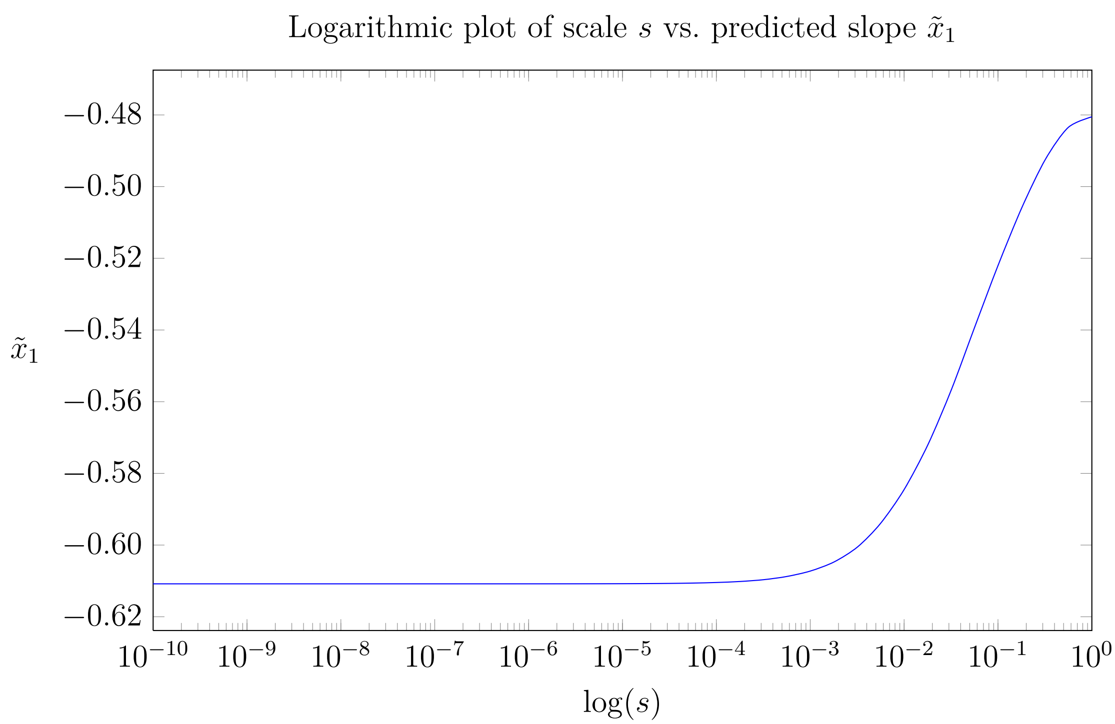


Figure 6.5: Predicted slopes as a function of the scaled cofactor matrix $Q_x \to s \cdot Q_x$

Chapter 7: Conclusions and Outlook

This work has reviewed the origin and progression of the EIV model with particular emphasis on the types of admissible cofactor, resp. weight matrices. It is indeed interesting to see the progression from somewhat restricted weighting possibilities to very general weighting possibilities that allow for full variance-covariance matrices for both the observation vector and the data matrix, as well as cross-correlation matrices to account for correlations between their errors. An important step towards this end was the work of Schaffrin and Wieser [2008], with even more general developments due to the work of Mahboub [2012] and Fang [2011], the latter representing a complete maturation of the EIV model in terms of accommodating traditional variance-covariance matrices (i.e., symmetric positive-definite) for the data.

Though recent extensions to the EIV model open up a much wider range for its use, estimators within these models were not available if the required combination of cofactor matrices (i.e., matrix Q_1 herein) turned out to be singular. Of course, in this case one could try using the least-squares estimator within the Gauss-Helmert model as provided by Neitzel and Schaffrin [2012], which would generate a unique solution as long as the Neitzel/Schaffrin rank condition was satisfied. However, this still leaves somewhat of a deficiency in the EIV model, which has now been addressed in Chapter 3 of this work.

The TLS estimators of Chapter 3 now accommodate singular cofactor matrices in the EIV model, while guaranteeing a unique solution provided that a certain rank condition is satisfied. This development opens new possibilities for use of the EIV model. Such use includes cases where high correlation between errors in the observation vector and errors in the data matrix gives rise to a singular Q_1 matrix. In other cases the cofactor matrix may have been derived from a rank-deficient least-squares adjustment problem, for example the 2-D similarity transformation problem of §6.2 or the 2-D line-fitting problem of §6.1.2. Surely many other examples exist.

Chapter 4 follows up on the work of Fang [2011] in making analytical comparisons between respective TLS estimators with the EIV and Gauss-Helmert models. This work also complements the comparisons made by Neitzel and Petrovic [2008] at the adjustment level. The reader will also find, in Chapter 4, a useful reminder about how to avoid the pernicious pitfalls in solving non-linear, iterative least-squares problems (including a detailed algorithm), which were pointed out by Pope [1972] quite some time ago. Nevertheless, the reminder still needs to be made, as several examples can be found in textbooks on least-squares adjustments that completely ignore Pope's advice, even though they were written after his work was published.

In some TLS problems the researcher or practitioner will have not only cofactor matrices for the observed variables but also for the vector of parameters, as well as prior information on the parameter values. In this case, the parameter vector is no longer one of type *fixed effects*, as in the usual EIV models, but rather becomes a vector of type *random* effects, for which the EIV with random effects model (EIV-REM) is required. Schaffrin [2009] already presented the predictor for this model in the case that the data were iid. In Chapter 5 of this work the iid restriction has been removed. Here a TLS predictor has been derived within the EIV-REM having very general cofactor matrices.

There is certainly more work that can be done to extend the EIV model even further and/or to develop different types of estimators with the EIV model. A couple of suggestions for further research are mentioned below.

The first suggestion is to include multiple variance components into the EIV model. As a start, one could include separate variance components for the cofactor matrix Q_y, of the observation vector $\boldsymbol{y}$, and the cofactor matrix Q_A, of the data matrix A. This would be critical in the case where the variables in $\boldsymbol{y}$ where observed using a different technique or instrument than was used for the variables in A and when the relative precision of the differences is not well known.

In the case of the EIV-REM, one could consider introducing a separate variance component for the prior information (perhaps then having three variance components in total). It was already assumed in Chapter 5 that the observational data and the prior information come from different sources and thus are uncorrelated. Perhaps in some cases the reality of this assumption also means that it is hard to know the relative precision between the observational data and the prior information, which would certainly be a strong enough motivation for considering an additional variance component for the prior information.

The second suggestion for further research is the incorporation of regularization (i.e., of type Tikhonov) into the TLS estimator within the EIV model. This would require modifying the target function used to derive the estimator, while the EIV model itself would remain unchanged. Schaffrin and Snow [2010] already derived such an estimator within the GHM, where they showed that the EIV model was a type of GHM. However, it would still be worthwhile to develop Tikhonov regularization within the extended EIV model of (2.1), with the objective of obtaining an estimator with certain properties of optimality.

Appendix A: Cofactor and Correlation Matrices

A.1 Correlation coefficients for 2-D line-fitting

Table A.1 lists the correlation coefficients used for the 2-D line-fitting in §6.1.2, with a slightly different ordering that must be considered. In order to use these coefficients to generate the matrix $\begin{bmatrix} Q_x & Q_{xy} \\ Q_{yx} & Q_y \end{bmatrix}$, construct a symmetric matrix $R = [\rho_{ij}]$ with ones on the diagonal and the off-diagonal elements taken from the values for ρ_{ij} in Table A.1. Then let Q_0 be a diagonal matrix comprised of the inverses of the weights given in Table 6.1, i.e., $Q_0 := \text{Diag}\left(\left[1/\sigma^2_{x_1}, \ldots, 1/\sigma^2_{x_{10}}, 1/\sigma^2_{y_1}, \ldots, 1/\sigma^2_{y_{10}}\right]\right)$. Note the ordering of the elements! Then compute

$$\begin{bmatrix} Q_x & Q_{xy} \\ Q_{yx} & Q_y \end{bmatrix} = Q_0^{1/2} R Q_0^{1/2}.$$

If it is the matrix $\begin{bmatrix} Q_y & Q_{yx} \\ Q_{xy} & Q_x \end{bmatrix}$ that is required, it can be generated by

$$\begin{bmatrix} Q_y & Q_{yx} \\ Q_{xy} & Q_x \end{bmatrix} = \begin{bmatrix} 0 & I_n \\ I_n & 0 \end{bmatrix} \begin{bmatrix} Q_x & Q_{xy} \\ Q_{yx} & Q_y \end{bmatrix} \begin{bmatrix} 0 & I_n \\ I_n & 0 \end{bmatrix}.$$

Table A.1: Correlation coefficients for 2-D line-fitting – Case 2

$R = \{\rho_{ij}\}$	Numerical values of $R = \{\rho_{ij}\}$				
$\rho_{01,02} \cdots \rho_{01,20}$	−0.10729581873004	−0.13036772607508	−0.14543901792983	−0.17728539923803	−0.07425424945810
	−0.00031909202598	0.07868531508505	0.05545515048479	0.11888722238149	1.00000000000000
	−0.10729581873004	−0.13036772607508	−0.14543901792983	−0.17728539923803	−0.07425424945810
	−0.00031909202598	0.07868531508505	0.05545515048479	0.11888722238149	
$\rho_{02,03} \cdots \rho_{02,20}$	−0.14802155668504	−0.16631815753526	−0.20504857968775	−0.08997554009936	−0.01579733299168
	0.07433378391878	0.05677853351729	0.12793755162244	−0.10729581873004	1.00000000000000
	−0.14802155668504	−0.16631815753526	−0.20504857968775	−0.08997554009936	−0.01579733299168
	0.07433378391878	0.05677853351729	0.12793755162244		
$\rho_{03,04} \cdots \rho_{03,20}$	−0.20677760277889	−0.25913321746667	−0.12106477055257	−0.04769547503133	0.06397035186675
	0.05852381274748	0.14448582847873	−0.13036772607508	−0.14802155668504	1.00000000000000
	−0.20677760277889	−0.25913321746667	−0.12106477055257	−0.04769547503133	0.06397035186675
	0.05852381274748	0.14448582847873			
$\rho_{04,05} \cdots \rho_{04,20}$	−0.30519893292947	−0.15398805118467	−0.09915289932191	0.02889396723543	0.04842194382799
	0.14351829731912	−0.14543901792983	−0.16631815753526	−0.20677760277889	1.00000000000000
	−0.30519893292947	−0.15398805118467	−0.09915289932191	0.02889396723543	0.04842194382799
	0.14351829731912				
$\rho_{05,06} \cdots \rho_{05,20}$	−0.22465607987009	−0.21055387807942	−0.04769138812163	0.02609634246228	0.14044733149058
	−0.17728539923803	−0.20504857968775	−0.25913321746667	−0.30519893292947	1.00000000000000
	−0.22465607987009	−0.21055387807943	−0.04769138812163	0.02609634246228	0.14044733149058
$\rho_{06,07} \cdots \rho_{06,20}$	−0.24695374215615	−0.16674302304054	−0.04735842255881	−0.00223986777699	−0.07425424945810
	−0.08997554009936	−0.12106477055257	−0.15398805118467	−0.22465607987009	1.00000000000000
	−0.24695374215615	−0.16674302304054	−0.04735842255881	−0.00223986777699	
$\rho_{07,08} \cdots \rho_{07,20}$	−0.55332321941763	−0.21967022592978	−0.22992934105249	−0.00031909202598	−0.01579733299168
	−0.04769547503133	−0.09915289932191	−0.21055387807943	−0.24695374215615	1.00000000000000
	−0.55332321941763	−0.21967022592978	−0.22992934105249		
$\rho_{08,09} \cdots \rho_{08,20}$	−0.24903718461920	−0.31109981507291	0.07868531508505	0.07433378391878	0.06397035186675
	0.02889396723543	−0.04769138812163	−0.16674302304054	−0.55332321941763	1.00000000000000
	−0.24903718461920	−0.31109981507291			

Table A.1–Continued

$R = \{\rho_{ij}\}$	Numerical values of $R = \{\rho_{ij}\}$				
$\rho_{09,10} \cdots \rho_{09,20}$	−0.15375435693872	0.05545515048479	0.05677853351729	0.05852381274748	0.04842194382799
	0.02609634246228	−0.04735842255881	−0.21967022592978	−0.24903718461920	1.00000000000000
	−0.15375435693872				
$\rho_{10,11} \cdots \rho_{10,20}$	0.11888722238149	0.12793755162244	0.14448582847873	0.14351829731912	0.14044733149058
	−0.00223986777699	−0.22992934105249	−0.31109981507291	−0.15375435693872	1.00000000000000
$\rho_{11,12} \cdots \rho_{11,20}$	−0.10729581873004	−0.13036772607508	−0.14543901792983	−0.17728539923803	−0.07425424945810
	−0.00031909202598	0.07868531508505	0.05545515048479	0.11888722238149	
$\rho_{12,13} \cdots \rho_{12,20}$	−0.14802155668504	−0.16631815753526	−0.20504857968775	−0.08997554009936	−0.01579733299168
	0.07433378391878	0.05677853351729	0.12793755162244		
$\rho_{13,14} \cdots \rho_{13,20}$	−0.20677760277889	−0.25913321746667	−0.12106477055257	−0.04769547503133	0.06397035186675
	0.05852381274748	0.14448582847873			
$\rho_{14,15} \cdots \rho_{14,20}$	−0.30519893292947	−0.15398805118467	−0.09915289932191	0.02889396723543	0.04842194382799
	0.14351829731912				
$\rho_{15,16} \cdots \rho_{15,20}$	−0.22465607987009	−0.21055387807943	−0.04769138812163	0.02609634246228	0.14044733149058
$\rho_{16,17} \cdots \rho_{16,20}$	−0.24695374215615	−0.16674302304054	−0.04735842255881	−0.00223986777699	
$\rho_{17,18} \cdots \rho_{17,20}$	−0.55332321941763	−0.21967022592978	−0.22992934105249		
$\rho_{18,19}, \rho_{18,20}$	−0.24903718461920	−0.31109981507291			
$\rho_{19,20}$	−0.15375435693872				

A.2 Cofactor matrices for the 2-D similarity transformation

Listed below are the cofactor matrices for the 2-D similarity transformation problem of §6.3.

The 10×10 cofactor Matrix Q_{xy} associated with the source system:

$$
Q_{xy} = \left[\begin{array}{rrrrr}
36.370281457026799 & -5.470847531049374 & -10.717856095227500 & -4.151968395749720 & -8.652980417908310 \\
-5.470847531049374 & 29.082186568548600 & 5.848221379655184 & -12.471413471703499 & 4.362573210487724 \\
-10.717856095227500 & 5.848221379655184 & 31.714850184591501 & 3.083310192383325 & -9.754412805230141 \\
-4.151968395749720 & -12.471413471703499 & 3.083310192383325 & 30.958038019578300 & 3.061227368299260 \\
-8.652980417908310 & 4.362573210487724 & -9.754412805230141 & 3.061227368299260 & 29.490003562291800 \\
-6.309575380973525 & -6.363232761238930 & -3.437514270957705 & -17.720699070083000 & 6.203628967652265 \\
-3.008722473688215 & -0.456888271105760 & -7.170133519153175 & 5.116298153814835 & -9.264714951570751 \\
5.061541661693445 & -2.835610062891940 & -5.891999249139355 & -0.556828023714815 & -10.232648328119300 \\
-13.990722470202799 & -4.283058787987756 & -4.072447764980720 & -7.108867318747670 & -1.817895387582575 \\
10.870849646079201 & -7.411930272714205 & 0.397981948058545 & -0.209097454076965 & -3.394781218319955
\end{array}\right. \cdots
$$

$$
\left.\begin{array}{rrrrr}
-6.309575380973525 & -3.008722473688215 & 5.061541661693445 & -13.990722470202799 & 10.870849646079201 \\
-6.363232761238930 & -0.456888271105760 & -2.835610062891940 & -4.283058787987756 & -7.411930272714205 \\
-3.437514270957705 & -7.170133519153175 & -5.891999249139355 & -4.072447764980720 & 0.397981948058545 \\
-17.720699070083000 & 5.116298153814835 & -0.556828023714815 & -7.108867318747670 & -0.209097454076965 \\
6.203628967652265 & -9.264714951570751 & -10.232648328119300 & -1.817895387582575 & -3.394781218319955 \\
38.734832754164600 & 2.979330360651055 & -10.531447467909800 & 0.564130323627930 & -4.119453454932780 \\
2.979330360651055 & 26.031496206858101 & -1.237264484444125 & -6.587925262445999 & -6.401475758915990 \\
-10.531447467909800 & -1.237264484444125 & 32.063142073587599 & 12.300370400009299 & -18.139256519071100 \\
0.564130323627930 & -6.587925262445999 & 12.300370400009299 & 26.468990885212101 & -1.472574616901790 \\
-4.119453454932780 & -6.401475758915990 & -18.139256519071100 & -1.472574616901790 & 29.879737700795101
\end{array}\right] [\mathrm{mm}^2]
$$

The 10×10 cofactor Matrix Q_{XY} associated with the target system:

$$
Q_{XY} = \left[\begin{matrix}
6.794487571800590 & -1.246373797420340 & -2.067492652515560 & -0.879050061336250 & -1.752371987000400 \\
-1.246373797420340 & 6.094989272208480 & 1.090128341653400 & -2.499242822649745 & 0.938081379301460 \\
-2.067492652515560 & 1.090128341653400 & 6.429318039548140 & 0.554510615436640 & -1.970512300406600 \\
-0.879050061336250 & -2.499242822649745 & 0.554510615436640 & 5.912760395723920 & 0.362129743332930 \\
-1.752371987000400 & 0.938081379301460 & -1.970512300406600 & 0.362129743332930 & 6.229399700497530 \\
-1.163579541268595 & -1.204799645049460 & -0.917558590359990 & -3.440142089389230 & 1.443374807628205 \\
-0.451547525907080 & -0.106812374044770 & -1.403742194865920 & 1.212589988419975 & -2.051290948290685 \\
0.979701730455350 & -0.699300070526570 & -0.955341882018735 & -0.117912645153740 & -2.018715860691035 \\
-2.523075406377545 & -0.675023549489755 & -0.987570891760060 & -1.250180285853295 & -0.455224464799860 \\
2.309301669569825 & -1.691646733982720 & 0.228261515288690 & 0.144537161468785 & -0.724870069571550
\end{matrix} \right. \cdots
$$

$$
\left. \begin{matrix}
-1.163579541268595 & -0.451547525907080 & 0.979701730455350 & -2.523075406377545 & 2.309301669569825 \\
-1.204799645049460 & -0.106812374044770 & -0.699300070526570 & -0.675023549489755 & -1.691646733982720 \\
-0.917558590359990 & -1.403742194865920 & -0.955341882018735 & -0.987570891760060 & 0.228261515288690 \\
-3.440142089389230 & 1.212589988419975 & -0.117912645153740 & -1.250180285853295 & 0.144537161468785 \\
1.443374807628205 & -2.051290948290685 & -2.018715860691035 & -0.455224464799860 & -0.724870069571550 \\
7.206462860853270 & 0.582948103761090 & -1.847407593285470 & 0.054815220239305 & -0.714113533129125 \\
0.582948103761090 & 5.079972517721780 & -0.048153307431655 & -1.173391848658100 & -1.640572410704635 \\
-1.847407593285470 & -0.048153307431655 & 6.360602565647050 & 2.042509319686065 & -3.695982256681265 \\
0.054815220239305 & -1.173391848658100 & 2.042509319686065 & 5.139262611595570 & -0.172120704582330 \\
-0.714113533129125 & -1.640572410704635 & -3.695982256681265 & -0.172120704582330 & 5.957205362324320
\end{matrix} \right] [\mathrm{mm}^2]
$$

Appendix B: Connections between the Work of Pearson and that of Schaffrin

In this section, a connection is made between the characteristic equation of Pearson [1901, eq. (10)] and of the characteristic equation of Schaffrin's normal-equation matrix [Schaffrin, 2007, eq. (1.15)] for the case of 2-D line-fitting. The least root of the former is claimed to be the total sum of squared residuals (TSSR) divided by the number of observations n, whereas the least root of the latter is the minimum eigenvalue of the normal-equation matrix, which is equivalent to the TSSR. Indeed it is shown below that the least root of the two characteristic equations differ exactly by a factor of n, implying that Pearson effectively solved a minimum eigenvalue problem.

Assume two n-vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, representing measured coordinate pairs in 2-D space, both having random errors distributed with zero-mean expectation and $\sigma_0^2 I_n$ dispersion. Let $\bar{x}$ and $\bar{y}$ denote the mean values of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively, so that

$$\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix} \sim (\begin{bmatrix}\bar{x}\\ \bar{y}\end{bmatrix} \otimes \mathbf{1}, \sigma_0^2 \begin{bmatrix}\sigma_x^2 & \sigma_{xy}\\ \sigma_{xy} & \sigma_y^2\end{bmatrix} \otimes I_n), \tag{B.1}$$

where $\mathbf{1}$ is an $n \times 1$ vector of ones, and σ_0^2 is an unknown variance component. The following empirical definitions for the variances σ_x^2 and σ_y^2, the covariance σ_{xy}, and the correlation coefficient ρ_{xy} are well known:

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = (\boldsymbol{x}^T\boldsymbol{x})/n - \bar{x}^2, \tag{B.2a}$$

$$\sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = (\boldsymbol{y}^T\boldsymbol{y})/n - \bar{y}^2, \tag{B.2b}$$

$$\sigma_{xy} = \sigma_{yx} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = (\boldsymbol{x}^T\boldsymbol{y})/n - \bar{x}\bar{y}, \tag{B.2c}$$

$$\rho_{xy} = \sigma_{xy}/\sqrt{\sigma_x^2\sigma_y^2} = \rho_{yx}. \tag{B.2d}$$

Write the upper 2×2 block of Pearson's characteristic equation (10), here using ρ rather than his r and Ω' rather than his Σ^2:

$$\begin{vmatrix} 1 - \frac{\Omega'}{\sigma_x^2} & \rho_{xy} \\ \rho_{yx} & 1 - \frac{\Omega'}{\sigma_y^2} \end{vmatrix} = 0 \Rightarrow \tag{B.3a}$$

$$\left(1 - \Omega'/\sigma_x^2\right)\left(1 - \Omega'/\sigma_y^2\right) - \rho_{xy}^2 = 0 \Rightarrow \tag{B.3b}$$

$$\Omega'_{\text{min}} = \frac{1}{2}(\sigma_x^2 + \sigma_y^2) - \frac{1}{2}\sqrt{(\sigma_x^2 + \sigma_y^2)^2 - 4\sigma_x^2\sigma_y^2(1 - \rho_{xy}^2)} = \tag{B.3c}$$

$$\Omega'_{\text{min}} = \frac{1}{2}(\sigma_x^2 + \sigma_y^2) - \frac{1}{2}\sqrt{(\sigma_x^2 - \sigma_y^2)^2 + 4\sigma_x^2\sigma_y^2\rho_{xy}^2} \tag{B.3d}$$

$$\Omega'_{\text{min}} \triangleq n^{-1} \cdot (\text{TSSR})_{\text{Pearson}}. \tag{B.3e}$$

Write the normal-equation matrix of Schaffrin:

$$\begin{bmatrix} \boldsymbol{x}^T\boldsymbol{x} - n \cdot \bar{x}^2 & \boldsymbol{x}^T\boldsymbol{y} - n \cdot \bar{x}\bar{y} \\ \boldsymbol{x}^T\boldsymbol{y} - n \cdot \bar{x}\bar{y} & \boldsymbol{y}^T\boldsymbol{y} - n \cdot \bar{y}^2 \end{bmatrix} = n \begin{bmatrix} \sigma_x^2 & \rho_{xy} \\ \rho_{xy} & \sigma_y^2 \end{bmatrix}. \tag{B.4}$$

Now form the characteristic equation for the matrix, using λ to denote an eigenvalue and factoring the scalar n outside of the determinant operator. The latter step is appropriate for computing eigenvalues of a matrix but not for computing its determinant, which would require a squaring of n.

$$n\begin{vmatrix} \sigma_x^2 - \lambda & \rho_{xy} \\ \rho_{xy} & \sigma_y^2 - \lambda \end{vmatrix} = 0 \Rightarrow \tag{B.5a}$$

$$n\Big[\lambda^2 - (\sigma_x^2 + \sigma_y^2)\lambda + \rho_{xy}^2\Big] = 0 \Rightarrow \tag{B.5b}$$

$$\lambda_{\text{min}} = n \cdot \left[\frac{1}{2}(\sigma_x^2 + \sigma_y^2) - \frac{1}{2}\sqrt{(\sigma_x^2 + \sigma_y^2)^2 + 4\rho_{xy}^2}\right] \Rightarrow \tag{B.5c}$$

$$\lambda_{\text{min}} = n \cdot \left[\frac{1}{2}(\sigma_x^2 + \sigma_y^2) - \frac{1}{2}\sqrt{(\sigma_x^2 - \sigma_y^2)^2 + 4\rho_{xy}^2\sigma_x^2\sigma_y^2}\right] \tag{B.5d}$$

$$\lambda_{\text{min}} \triangleq (\text{TSSR})_{\text{Schaffrin}} \tag{B.5e}$$

Finally, comparing the least root of Pearson to the minimum eigenvalue of Schaffrin gives

$$\underbrace{n \cdot \Omega'_{\text{min}}}_{\text{Pearson}} = \underbrace{\lambda_{\text{min}}}_{\text{Schaffrin}} \triangleq \text{TSSR}, \tag{B.6}$$

which shows the connection between the developments in Pearson's work of 1901 and Schaffrin's work of 2007.

Appendix C: The Neitzel/Schaffrin Criterion for Uniqueness

The following statement of the Neitzel/Schaffrin uniqueness condition (or rank condition) is based on their work from 2009, submitted for publication in 2012, which presented a unique least-squares estimator within the Gauss-Helmert model with a rank-deficient dispersion matrix (or, equivalently, cofactor matrix).

Considering the EIV-Model

$$\boldsymbol{y} - \boldsymbol{e}_y - \underbrace{(A - E_A)}_{n\times m}\boldsymbol{\xi} = \boldsymbol{0} \tag{C.1}$$

as a special case of the nonlinear Gauss-Helmert model

$$\boldsymbol{b}(\boldsymbol{y} - \boldsymbol{e}_y, A - E_A, \boldsymbol{\xi}) = \boldsymbol{0}, \tag{C.2}$$

the Neitzel/Schaffrin condition for uniqueness of the least-squares solution in the presence of a singular dispersion matrix, i.e.,

$$Q = \begin{bmatrix} \underset{n\times n}{Q_y} & \underset{n\times nm}{Q_{yA}} \\ \underset{nm\times n}{Q_{Ay}} & \underset{nm\times nm}{Q_A} \end{bmatrix} \tag{C.3}$$

symmetric positive-semidefinite, reads

$$\operatorname{rk}\Big[BQ \;\Big|\; A\Big] = n, \tag{C.4}$$

$$\text{where } B := \frac{\partial \boldsymbol{b}}{\partial[(\boldsymbol{y}-\boldsymbol{e})^T,\ (\operatorname{vec} A - \boldsymbol{e}_A)^T]} = \underbrace{\Big[I_n,\quad -\Big(\boldsymbol{\xi}^T \otimes I_n\Big)\Big]}_{n\times(m+1)n}, \tag{C.5}$$

or, equivalently,

$$\begin{aligned} &\operatorname{rk}\Bigg[\Big[Q_y - \Big(\boldsymbol{\xi}\otimes I_n\Big)^T Q_{Ay},\quad Q_{yA} - \Big(\boldsymbol{\xi}\otimes I_n\Big)^T Q_A\Big] \;\Bigg|\; A\Bigg] = \\ &= \operatorname{rk}\Big(Q_1 + ASA^T\Big) = n, \end{aligned} \tag{C.6}$$

for any symmetric positive-definite $m \times m$ matrix S and

$$Q_1 := Q_y - Q_{yA}\Big(\boldsymbol{\xi}\otimes I_n\Big) - \Big(\boldsymbol{\xi}\otimes I_n\Big)^T Q_{Ay} + \Big(\boldsymbol{\xi}\otimes I_n\Big)^T Q_A\Big(\boldsymbol{\xi}\otimes I_n\Big). \tag{C.7}$$

Appendix D: Fang's WTLS Estimator as a Generalization of Schaffrin and Wieser's Estimator

Here it is shown that the WTLS estimators derived by Fang [2011, eqs. (4.22, 4.25, 426)] and Mahboub [2012, eq. (24)] are generalizations of the WTLS estimator by Schaffrin and Wieser [2008, eq. (17)]. The purpose in showing these relations is merely to facilitate comparison of these authors' works.

The EIV model presented by Schaffrin and Wieser reads

$$\boldsymbol{y} = (\underset{n\times m}{A} - E_A) \cdot \boldsymbol{\xi} + \boldsymbol{e}_y, \tag{D.1a}$$

$$\begin{bmatrix} \boldsymbol{e}_y \\ \boldsymbol{e}_A \end{bmatrix} := \begin{bmatrix} \boldsymbol{e}_y \\ \operatorname{vec} E_A \end{bmatrix} \sim (\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \sigma_0^2 \begin{bmatrix} Q_y & 0 \\ 0 & Q_A \end{bmatrix}), \tag{D.1b}$$

where the terms are defined as in Chapter 2, except that Q_A has the special form $Q_A := Q_0 \otimes Q_x$, with Q_0 having dimension $m \times m$ and Q_x being size $n \times n$.

Now, under this model, the matrix Q_1 defined in (2.10b) reduces to

$$\begin{aligned} Q_1 = \Big[Q_y + \big(\hat{\boldsymbol{\xi}}^T \otimes I_n\big)\big(Q_0 \otimes Q_x\big)\big(\hat{\boldsymbol{\xi}} \otimes I_n\big)\Big] = \\ = Q_y + \big(\hat{\boldsymbol{\xi}}^T Q_0 \hat{\boldsymbol{\xi}}\big) Q_x. \end{aligned} \tag{D.2}$$

Then Schaffrin and Wieser's normal equations (17) could be revised to read

$$\big(A^T Q_1^{-1} A - \hat{\boldsymbol{\nu}} \cdot Q_0\big)\hat{\boldsymbol{\xi}} = A^T Q_1^{-1} \boldsymbol{y}. \tag{D.3}$$

Now substituting their equation (16)

$$-A^T \hat{\boldsymbol{\lambda}} = -\tilde{E}_A^T \hat{\boldsymbol{\lambda}} = Q_0 \hat{\boldsymbol{\xi}} \cdot \hat{\nu} \tag{D.4}$$

and performing some algebraic manipulation leads to

$$\begin{aligned} \big(A^T Q_1^{-1} A\big)\hat{\boldsymbol{\xi}} &= A^T Q_1^{-1} \boldsymbol{y} - \tilde{E}_A^T \hat{\boldsymbol{\lambda}} \\ \Rightarrow \big(A^T Q_1^{-1} A\big)\hat{\boldsymbol{\xi}} &= A^T Q_1^{-1} \boldsymbol{y} - \operatorname{vec}\big(\tilde{E}_A^T \hat{\boldsymbol{\lambda}}\big) \\ \Rightarrow \big(A^T Q_1^{-1} A\big)\hat{\boldsymbol{\xi}} &= A^T Q_1^{-1} \boldsymbol{y} - \operatorname{vec}\big(\hat{\boldsymbol{\lambda}}^T \tilde{E}_A\big) \\ \Rightarrow \big(A^T Q_1^{-1} A\big)\hat{\boldsymbol{\xi}} &= A^T Q_1^{-1} \boldsymbol{y} - \big(I_m \otimes \hat{\boldsymbol{\lambda}}^T\big)\tilde{\boldsymbol{e}}_A \end{aligned}$$

Schaffrin and Wieser [2008]	Fang [2011]	Mahboub [2012]
m	u	m
$Q_y + (\hat{\boldsymbol{\xi}}^T Q_0 \hat{\boldsymbol{\xi}}) Q_x = Q_1$	$(\hat{B} Q_{ll} \hat{B})$	R_1^{-1}
$\tilde{E}_A$	$-\tilde{V}_A$	$\tilde{E}_A$
$\tilde{\boldsymbol{e}}_A$	$\tilde{\boldsymbol{v}}_A$	$\tilde{\boldsymbol{e}}_A$

Table D.1: Differences in notation between authors. The first column is consistent with the notation used in this dissertation.

$$\boxed{\Rightarrow \hat{\boldsymbol{\xi}} = \left(A^T Q_1^{-1} A\right)^{-1} \left(A^T Q_1^{-1} \boldsymbol{y} - \left(I_m \otimes \hat{\boldsymbol{\lambda}}^T\right) \tilde{\boldsymbol{e}}_A\right)}. \tag{D.5}$$

Going further by subtracting a term in $\tilde{E}_A^T$ from each side of the preceding equations, and using the relation $Q_1 \hat{\boldsymbol{\lambda}} = (\boldsymbol{y} - A\hat{\boldsymbol{\xi}})$ from (2.10a), leads to

$$\begin{aligned}
\left[\left(A - \tilde{E}_A\right)^T Q_1^{-1} A\right] \hat{\boldsymbol{\xi}} &= A^T Q_1^{-1} \boldsymbol{y} - \left(I_m \otimes \hat{\boldsymbol{\lambda}}^T\right) \tilde{\boldsymbol{e}}_A - \tilde{E}_A^T Q_1^{-1} A \hat{\boldsymbol{\xi}} = \\
&= A^T Q_1^{-1} \boldsymbol{y} - \tilde{E}_A^T Q_1^{-1} \left(A \hat{\boldsymbol{\xi}} + Q_1 \hat{\boldsymbol{\lambda}}\right) = \\
&= A^T Q_1^{-1} \boldsymbol{y} - \tilde{E}_A^T Q_1^{-1} \left(A \hat{\boldsymbol{\xi}} + \boldsymbol{y} - A \hat{\boldsymbol{\xi}}\right)
\end{aligned}$$

$$\boxed{\Rightarrow \hat{\boldsymbol{\xi}} = \left[\left(A - \tilde{E}_A\right)^T Q_1^{-1} A\right]^{-1} \left(A - \tilde{E}_A\right)^T Q_1^{-1} \boldsymbol{y}}. \tag{D.6}$$

It may be desirable to have rather a symmetric matrix to invert in (D.6). This can readily be obtained by subtracting $(A - \tilde{E}_A)^T Q_1^{-1} \tilde{E}_A \hat{\boldsymbol{\xi}}$ from both sides of the preceding equations to arrive at

$$\begin{aligned}
\left[\left(A - \tilde{E}_A\right)^T Q_1^{-1} \left(A - \tilde{E}_A\right)\right] \hat{\boldsymbol{\xi}} &= \left(A - \tilde{E}_A\right)^T Q_1^{-1} \boldsymbol{y} - \left(A - \tilde{E}_A\right)^T Q_1^{-1} \tilde{E}_A \hat{\boldsymbol{\xi}} = \\
&= \left(A - \tilde{E}_A\right)^T Q_1^{-1} \left(\boldsymbol{y} - \tilde{E}_A \hat{\boldsymbol{\xi}}\right)
\end{aligned}$$

$$\boxed{\Rightarrow \hat{\boldsymbol{\xi}} = \left[\left(A - \tilde{E}_A\right)^T Q_1^{-1} \left(A - \tilde{E}_A\right)\right]^{-1} \left(A - \tilde{E}_A\right)^T Q_1^{-1} \left(\boldsymbol{y} - \tilde{E}_A \hat{\boldsymbol{\xi}}\right)}. \tag{D.7}$$

The notation of Fang and Schaffrin/Wieser is similar except for the differences shown in the following table:

Now considering the notation differences listed in Table D.1, it is apparent that equations (D.5), (D.6), and (D.7) are algebraically equivalent to Fang's equations (4.22), (4.25), and (4.26), respectively, which are also associated with his algorithms 1, 2, 3, respectively.

Now turning to the work by Mahboub [2012], define

$$R_2 := \left(I_m \otimes \hat{\boldsymbol{\lambda}}^T\right) Q_A \left(\hat{\boldsymbol{\xi}} \otimes I_n\right) Q_1^{-1}, \tag{D.8}$$

which is identical to his equation (23), except that he substitutes R_1 for Q_1^{-1}. Note also that Mahboub's Q_1 takes the form of (3.1d), as he has not treated the case of nonzero Q_{Ay}. Since

$$\begin{aligned} R_2(\boldsymbol{y} - A\hat{\boldsymbol{\xi}}) = \left(I_m \otimes \hat{\boldsymbol{\lambda}}^T\right) Q_A \left(\hat{\boldsymbol{\xi}} \otimes I_n\right) Q_1^{-1} (\boldsymbol{y} - A\hat{\boldsymbol{\xi}}) = \\ = \left(I_m \otimes \hat{\boldsymbol{\lambda}}^T\right) Q_A \left(\hat{\boldsymbol{\xi}} \otimes I_n\right) \hat{\boldsymbol{\lambda}} = \\ = -\left(I_m \otimes \hat{\boldsymbol{\lambda}}^T\right) \tilde{\boldsymbol{e}}_A, \end{aligned}$$

it is apparent that the estimator by Mahboub shown in his equation (24) as

$$\hat{\boldsymbol{\xi}} = \left(A^T R_1 A + R_2 A\right)^{-1} \left(A^T R_1 + R_2\right) \boldsymbol{y} \tag{D.9}$$

is algebraically equivalent to (D.5). It is also identical to the estimator shown in (2.14) when the notation differences of Table D.1 are considered.

In summary, the estimators presented by Schaffrin and Wieser [2008], Fang [2011], and Mahboub [2012] are algebraically equivalent up to the definition of the matrix Q_1. Schaffrin/Wieser define Q_1 as in (D.2); Fang defines Q_1 as in (2.10b); and Mahboub defines Q_1 as in (3.1d). However, one should not underestimate the significance in the differences in these definitions, as they represent different levels of generality in the admission of weights into the EIV model. It could be said that, in terms of the admissible weight matrices, the differences vary from *fairly general* to *general* to *very general* between Schaffrin/Wieser, Mahboub, and Fang, respectively.

Appendix E: Some Properties of the vec Operator and the Kronecker Product of Matrices

The vec operator

Definition E.1. *Let A be an $n \times m$ matrix $A = [a_{ij}] = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m]$; then the vec operation* vec A *vertically stacks the successive $n \times 1$ columns of A, from the leftmost column $\boldsymbol{a}_1$ to the rightmost column $\boldsymbol{a}_m$, to form the $nm \times 1$ vector*

$$\underset{nm\times 1}{\operatorname{vec} A} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \\ a_{12} \\ \vdots \\ a_{n2} \\ a_{13} \\ \vdots \\ a_{nm} \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_m \end{bmatrix}.$$

The process is also called the vectorization of matrix A.

The Kronecker product

Definition E.2. *Let $G = [g_{ij}]$ be a $p \times q$ matrix and $H = [h_{ij}]$ be an $r \times s$ matrix; then*

$$G \otimes H := \left[g_{ij}H\right]$$

gives the Kronecker product (also called Kronecker-Zehfuss product) $G \otimes H$, which is of size $pr \times qs$.

Commutation matrix

Obviously the $mn \times 1$ vector $\operatorname{vec}(A^T)$ contains exactly the same elements as the $nm \times 1$ vector vec A, except in a different order. The difference in order of elements is revealed by a unique *commutation* matrix of size $nm \times nm$.

Definition E.3. *Given an $n \times m$ matrix A, the relationship between $\operatorname{vec} A$ and $\operatorname{vec}(A^T)$ is given by use of a unique $nm \times nm$ commutation (or vec-permutation) matrix K_{nm}, where*

$$K_{nm} \operatorname{vec} A = \operatorname{vec}(A^T).$$

K_{nm} is a permutation matrix and therefore satisfies the relations $K_{nm}^T = K_{nm}^{-1} = K_{mn}$. Also, it holds that $K_{n1} = K_{1n} = I_n$, and $K_{nn} = K_{nn}^T$ if $m = n$.

Definition E.4. *The $nm \times nm$ commutation matrix K_{nm} is determined by*

$$K_{nm} = \sum_{i=1}^{n} \sum_{j=1}^{m} \Big(H_{ij} \otimes H_{ij}^T \Big),$$

where $H_{ij} := \boldsymbol{\eta}_i \boldsymbol{\eta}_i^T$ is an $n \times m$ matrix with a 1 in its ij entry and zeros elsewhere [Magnus and Neudecker, 1979]. Here $\boldsymbol{\eta}_i$ and $\boldsymbol{\eta}_j$ denote the i-th, resp, j-th unit vector of size $n \times 1$, resp. $m \times 1$.

Computational rules

$$\operatorname{vec} ABC^T = (C \otimes A) \operatorname{vec} B \tag{E.1}$$

$$\operatorname{tr} ABC^T D^T = \operatorname{tr} D^T ABC^T = (\operatorname{vec} D)^T (C \otimes A) \operatorname{vec} B \tag{E.2}$$

$$(G \otimes H)^T = G^T \otimes H^T \tag{E.3}$$

$$(G \otimes H)^{-1} = G^{-1} \otimes H^{-1} \tag{E.4}$$

$$\alpha(G \otimes H) = \alpha G \otimes H = G \otimes \alpha H \text{ for } \alpha \in \mathbb{R} \tag{E.5}$$

$$(F + G) \otimes H = (F \otimes H) + (G \otimes H) \tag{E.6}$$

$$G \otimes (H + J) = (G \otimes H) + (G \otimes J) \tag{E.7}$$

$$(A \otimes B)(G \otimes H) = AG \otimes BH \tag{E.8}$$

Let H be of dimension $m \times n$ and G be of dimension $p \times q$.

$$(H \otimes G) = K_{pm}(G \otimes H) K_{nq} \tag{E.9}$$

$$K_{mp}(H \otimes G) = (G \otimes H) K_{nq} \tag{E.10}$$

$$K_{mp}(H \otimes \boldsymbol{g}) = \boldsymbol{g} \otimes H \quad \forall\, \boldsymbol{g} \in \mathbb{R}^{p \times 1} \tag{E.11}$$

$$\operatorname{tr}(G \otimes H) = \operatorname{tr} G \cdot \operatorname{tr} H \tag{E.12}$$

$$G \text{ and } H \text{ positive-(semi)definite } \Rightarrow G \otimes H \text{ positive-(semi)definite.} \tag{E.13}$$

Any of the following works can be consulted for further details on these linear algebra topics: Magnus and Neudecker [1979], Lütkepohl [1996], Horn and Johnson [1994], Harville [1997], or Magnus and Neudecker [2007].

References

R. J. Adcock. Note on the method of least squares. *The Analyst*, IV(6):183–184, 1877.

A. C. Aitken. On least squares and linear combinations of observations. *Proceedings of the Royal Society of Edinburgh*, 55:42–48, 1935.

A. Ben-Israel and T. N. E. Greville. *Generalized Inverses: Theory and Applications*. CMS Books in Mathematics. Springer, New York, second edition, 2003.

A. Björck. *Numerical Methods for Least Squares Problems*. SIAM, Philadelphia, PA, 1996.

J. C. Davis. *Statistics and Data Analysis in Geology*. John Wiley & Sons, New York, third edition, 2002.

W. E. Deming. *Statistical Adjustment of Data*. Dover Publications, New York, 1964. Unabridged and corrected republication of the 1943 publication by John Wiley & Sons, New York.

J. Demmel. Constrained total least squares problems and the smallest perturbation of a sub-matrix which lowers the rank. Technical Report 174, Computer Science Department, New York University, Sept. 1985.

G. W. Dunnington, J. Gray, and F.-E. Dohse. *Carl Friedrich Gauss: Titan of Science*. Spectrum Series. The Mathematical Association of America, Washington, DC, 2004. Reprint of 1955 publication by Hafner Publishing Company, New York.

X. Fang. *Weighted Total Least Squares Solutions for Applications in Geodesy*. PhD dissertation, Publ. No. 294, Dept. of Geodesy and Geoinformatics, Leibniz University Hannover, Germany, 2011.

Y. Felus and B. Schaffrin. On Total Least-Squares Adjustment with Constraints. In F. Sansò, editor, *A Window on the Future of Geodesy*, pages 417–421, Berlin/Heidelberg/New York, 2005. Springer.

W. Gander, G. H. Golub, and R. Strebel. Least-squares fitting of circles and ellipses. *BIT*, 34(4):558–578, Dec. 1994.

C. F. Gauss. *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium*. Perthes & Besser, Hamburg, 1809.

C. F. Gauss. *Theoria combinationis observationum erroribus minimis obnoxiae: Pars Prior, Pars Posterior, Supplementum*. H. Dieterich, Göttingen, 1823.

C. F. Gauss. *Theory of the Motion of the Heavenly Bodies Moving About the Sun in Conic Sections*. Little, Brown and Company, Boston, 1857. Translated by C. H. Davis.

C. F. Gauss. *Theory of the Combination of Observations Least Subject to Errors: Part One, Part Two, Supplement*. Classics in Applied Mathmatics. SIAM, Philadelphia, PA, 1995. Translated by G. W. Stewart.

G. A. Gerhold. Least-squares adjustment of weighted data to a general linear equation. *American Journal of Physics*, 37(2):156–161, Feb. 1969.

G. Golub and C. Van Loan. Total least squares. In T. Gasser and M. Rosenblatt, editors, *Smoothing Techniques for Curve Estimation*, volume 757 of *Lecture Notes in Mathematics*, pages 69–76. Springer Berlin/Heidelberg, 1979.

G. H. Golub. Some modified matrix eigenvalue problems. *SIAM Review*, 15(2):318–334, Apr. 1973.

G. H. Golub and R. J. Plemmons. Large-Scale Geodetic Least-Squares Adjustment by Dissection and Orthogonal Decomposition. *Linear Algebra and its Applications*, 34:3–27, 1980.

G. H. Golub and C. Reinsch. Singular value decomposition and least squares solutions. *Numerische Mathematik*, 14:403–420, 1970.

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

G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in Mathematical Sciences. The Johns Hopkins University Press, Baltimore, MD, third edition, 1996.

D. Harville. Using ordinary least squares software to compute combined intra-interblock estimates of treatment contrasts. *The American Statistician*, 40(2): 153–157, 1986.

D. A. Harville. *Matrix Algebra From a Statistician's Perspective*. Springer, New York, 1997.

W. A. Heiskanen and H. Moritz. *Physical Geodesy*. W. H. Freeman and Company, San Francisco, first edition, 1967.

F. R. Helmert. *Adjustment Computations by the Method of Least Squares* (in German). Teubner, Leipzig, Germany, second edition, 1907.

R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, Cambridge etc., paperback edition, 1994.

K.-R. Koch. *Parameter Estimation and Hypothesis Testing in Linear Models*. Springer, Berlin, second edition, 1999.

T. Krarup. A contribution to the mathematical foundation of physical geodesy. Technical Report 44, Geodaetisk Institut, Kobenhavn, Denmark, 1969.

T. Krarup. A contribution to the mathematical foundation of physical geodesy. In K. Borre, editor, *Foundation of Geodesy Selected Papers of Torben Krarup*. Springer-Verlag, 2006.

Y. V. Linnik. *Method of least squares and principles of the theory of observations*. Pergamon Press, New York, 1961.

H. Lütkepohl. *Handbook of Matrices*. John Wiley & Sons, Chichester, England, 1996.

J. R. Magnus and H. Neudecker. The Commutation Matrix: Some Properties and Applications. *The Annals of Statistics*, 7(2):381–394, Mar. 1979.

J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester, England, third edition, 2007.

V. Mahboub. On weighted total least-squares for geodetic transformations. *Journal of Geodesy*, 86(5):359–367, 2012.

I. Markovsky, M. L. Rastello, A. Premoli, A. Kukush, and S. Van Huffel. The element-wise weighted total least-squares problem. *Computational Statistics & Data Analysis*, 50(1):181–209, Jan. 2006.

I. Markovsky, D. M. Sima, and S. Van Huffel. Total least squares methods. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(2):212–217, Mar. 2010.

H. Moritz. A generalized least-squares model. *Studia Geophysica et Geodaetica*, 14 (2):353–362, 1970.

H. Moritz. Least-squares collocation. *Rev. Geophys. Space Phys*, 16(3):421–430, 1978.

F. Neitzel and S. Petrovic. Total Least-Squares in the context of least-squares adjustment of a straight line (in German). *Z. für Vermessungswesen*, 133:141–148, 2008.

F. Neitzel and B. Schaffrin. On the Gauss-Helmert Model with a singular dispersion matrix: where rk BQ is smaller than rk B. *Studia Geophysica et Geodaetica*, 2012. Submitted in July, 2012.

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

K. Pearson. LIII. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*, 2(11):559–572, Nov. 1901.

R. L. Plackett. A historical note on the method of least squares. *Biometrika*, 36(3): 458–460, 1949.

A. J. Pope. Some pitfalls to be avoided in the iterative adjustment of nonlinear problems. In *Proceedings of the 38th Annual Meeting of the ASPRS*, pages 449–477, Falls Church, VA, 1972. American Society of Photogrammetry.

C. R. Rao, H. Toutenburg, Shalabh, and C. Heumann. *Linear Models and Generalizations: Least Squares and Alternatives.* Springer Series in Statistics. Springer, Berlin/Heidelberg, third edition, 2008.

J. Reinking. Comment on "On the determination of unique transformation parameters" (in German). *Zeitschrift für Vermessungswesen*, 126(5):295–296, 2001.

B. Schaffrin. Softly unbiased prediction. Part 2: The random effects model. *Bollettino di Geodesia e Scienze Affini*, 60(1):49–62, 2001.

B. Schaffrin. A note on Constrained Total Least-Squares estimation. *Linear Algebra and its Applications*, 417(1):245–258, 2006.

B. Schaffrin. Connecting the dots: The straight-line case revisited. *Zeitschrift für Vermessungswesen*, 132(1):385–394, 2007.

B. Schaffrin. TLS-Collocation: The Total-Least Squares Approach to EIV-Models with Stochastic Prior Information. Presented at the 18th Intl. Workshop on Matrices and Statistics, Smolenice Castle, Slovakia, June 2009.

B. Schaffrin and K. Snow. Total Least-Squares regularization of Tykhonov type and an ancient racetrack in Corinth. *Linear Algebra and its Applications*, 432(8):2061–2076, Apr. 2010.

B. Schaffrin and S. Uzun. *Errors-In-Variables for mobile mapping algorithms in the presence of outliers*, volume 22 of *Archives of Photogrammetry, Cartography and Remote Sensing*, pages 377–387. Polish Society for Photogrammetry and Remote Sensing, Cracow, 2011.

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

B. Schaffrin and A. Wieser. Total least-squares adjustment of condition equations. *Studia Geophysica et Geodaetica*, 55(3):529–536, 2011.

B. Schaffrin, I. Lee, Y. Choi, and Y. Felus. Total Least-Squares (TLS) for geodetic straight-line and plane adjustment. *Bollettino di Geodesia e Scienze Affini*, 65(3): 141–168, 2006.

B. Schaffrin, F. Neitzel, S. Uzun, and V. Mahboub. Modifying Cadzow's algorithm to generate the optimal TLS-solution for the structured EIV-model of a similarity transformation. *Journal of Geodetic Science*, 2(2):98–106, 2012a.

B. Schaffrin, K. Snow, and F. Neitzel. On the Errors-In-Variables Model with singular covariance matrices. Presented at the 21st Intl. Workshop on Matrices and Statistics, Będlewo, Poland, July 2012b.

K. Snow and B. Schaffrin. Weighted Total Least-Squares Collocation with Geodetic Applications. Presented at the 2012 SIAM Conference on Applied Linear Algebra, Valencia, Spain, June 2012.

G. Strang. *Linear Algebra and Its Applications*. Harcourt Brace Javanovich, Orlando, FL, third edition, 1988.

G. Strang and K. Borre. *Linear Algebra, Geodesy, and GPS*. Wellesley College, Wellesley, MA, 1997.

S. Van Huffel, editor. *Recent advances in total least squares techniques and errors in variables modeling*. SIAM, Philadelphia, PA, 1997.

S. Van Huffel. Total Least Squares and Errors-in-Variables Modeling: Bridging the Gap between Statistics, Computational Mathematics and Engineering. In S. Van Huffel and P. Lemmerling, editors, *Total Least Squares and Errors-In-Variables Modeling: Analysis, Algorithms and Applications*, pages 1–11. Kluwer, Dordrecht, NL, 2004.

S. Van Huffel and J. Vandewalle. Analysis and properties of the generalized total least-squares problem $AX \sim B$ when some or all columns of A are subject to error. *SIAM Journal on Matrix Analysis and Applications*, 10(3):294–315, 1989.

S. Van Huffel and J. Vandewalle. *The Total Least Squares Problem: Computational Aspects and Analysis*. Frontiers in Applied Mathematics. SIAM, Philadelphia, PA, 1991.

A. Wald. The fitting of straight lines if both variables are subject to error. *The Annals of Mathematical Statistics*, 11(3):284–300, 1940.

D. York. Least-squares fitting of a straight line. *Canadian Journal of Physics*, 44: 1079–1086, 1966.

F. Zhang, editor. *The Schur Complement and Its Applications*, volume 4 of *Numerical Methods and Algorithms*. Springer-Verlag, New York, 2005.

Author Index