Adjusting the Errors-In-Variables Model: Linearized Least-Squares vs. Nonlinear Total Least-Squares

Burkhard Schaffrin

Abstract

It has long been known that the Errors-In-Variables (EIV) Model is a special case of the nonlinear Gauss–Helmert Model (GHM) and can, therefore, be adjusted by standard least-squares techniques in iteratively linearized GH-Models, which is the approach by Helmert (Adjustment Computations Based on the Least-Squares Principle (in German), 1907) and – later – by Deming (Phil Mag 11:146–158, 1931; Phil Mag 17:804–829, 1934).

Apart from the fact that there are, at least, two other nonlinear models that are equivalent to the above GH-Model, thus allowing two more classical least-squares approaches based on iterative linearization, it was the seminal paper by Golub and van Loan (SIAM J Numer Anal 17:883–893, 1980) in which they proved that a purely nonlinear approach can be followed as well, thereby avoiding any model linearization. They called such an approach "Total Least-Squares adjustment" by which any normal equations may be replaced by a simple eigenvalue problem, as long as only diagonal dispersion matrices are involved.

Here, an attempt will be made to show the differences and parallels in various algorithms, even in the fully weighted case, which obviously all generate the same results, but without necessarily showing equal efficiency in doing so, as is well known since the publications by Schaffrin and Wieser (J Geodesy 82:415–421, 2008), Fang (Weighted Total Least-Squares solutions with applications in geodesy, 2011), and Mahboub (J Geodesy 86:359–367, 2012).

Keywords

Errors-In-Variables Models • Total Least-Squares • Equivalent nonlinear models • Linearized Least-Squares

1 Introduction

The Errors-In-Variables (EIV) Model has recently seen a lot of attention since, in accordance with Golub and van Loan (1980), it can be treated in its *nonlinear* form by a least-squares approach that they coined "*Total Least-Squares adjustment*". It eventually leads to a (generalized) eigenvalue problem that needs to be solved in lieu of the sequence of normal equations that would result from a traditional "*Least-Squares adjustment*" within *iteratively linearized* models. The latter approach dates, at least, back to Helmert (1907), but has as well been used by Deming (1931, 1934) for the approximation of curves and, more recently, by Neitzel (2010) to determine the parameters of a similarity transformation.

In contrast, the nonlinear Total Least-Squares (TLS) approach which, in its original formulation, could tolerate only "*element-wise weighting*" and thus only *diagonal* weight matrices, has since been generalized in several steps by Schaffrin and Wieser (2008), Fang (2011), and Mahboub (2012) to now accept any *positive-definite weight matrices*.

B. Schaffrin (✉)
Division of Geodetic Science, School of Earth Sciences, The Ohio State University, Columbus, OH, USA
e-mail: schaffrin.1@osu.edu

N. Sneeuw et al. (eds.), *VIII Hotine-Marussi Symposium on Mathematical Geodesy*, International Association of Geodesy Symposia 142, DOI 10.1007/1345_2015_61

This development will be presented in the following Sect. 2, thereby showing how the more specialized algorithms can be derived from the more general ones by simplification.

Moreover, it should be noted that progress has also been made towards the use of *positive-semidefinite dispersion matrices* in TLS adjustment, which may be handled as described by Schaffrin et al. (2014). These cases are quite relevant whenever the random error matrix needs to show a certain pattern or structure after the adjustment. Due to the limited space, these advanced methods will not be discussed below.

Instead, attention will be paid to a *triplet of classical nonlinear models* that all can be constructed to be *equivalent* to the EIV-Model and, furthermore, may undergo a sequence of Least-Squares adjustments via iterative linearization which, in the end, converge to the very same TLS solution. This will be the theme in Sect. 3 although many details have to be left out; for those, see Schaffrin (2015).

2 Nonlinear TLS Adjustment in an EIV-Model

2.1 Fang's Algorithm

Let the EIV-Model be defined by

$$y = \underbrace{(A - E_A)}_{n\times m}\xi + e_y, \qquad rkA = m < n, \tag{1a}$$

$$e := \begin{bmatrix} e_y \\ e_A := vecE_A \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \ \sigma_o^2 \begin{bmatrix} \underset{n\times n}{P_y^{-1}} & 0 \\ 0 & \underset{nm\times nm}{P_A^{-1}} \end{bmatrix} =: \sigma_o^2 P^{-1} \right) \tag{1b}$$

where

y is the $n \times 1$ observation vector;
A is the $n \times m$ (random) coefficient matrix with full column rank (aka "data matrix");
E_A is the $n \times m$ (unknown) random error matrix associated with A;
ξ is the $m \times 1$ (unknown) parameter vector;
e_y is the $n \times 1$ (unknown) random error vector associated with y;
e_A is the $nm \times 1$ vectorial form of the matrix E_A;
Q is the $n(m+1) \times n(m+1)$ block-diagonal pos.- def. cofactor matrix;
$P := Q^{-1}$ is the corresponding block-diagonal pos.- def. weight matrix;
σ_o^2 is the (unknown) variance component (unit- free);
$\text{Cov}\left\{e_y, vecE_A\right\} = 0$ for the sake of simplicity.

The model generalizes the one used by Schaffrin and Wieser (2008) where a Kronecker product structure for

$$Q_A = P_A^{-1} = Q_o \otimes Q_x \tag{2}$$

was assumed, as well as the one used by Golub and von Loan (1980) who only allowed *diagonal* cofactor matrices with

$$Q_o := I_m, \quad Q_x := Q_y = Diag\left(p_1^{-1}, \ldots\ldots, p_n^{-1}\right) = P_y^{-1}. \tag{3}$$

The objectives of a *nonlinear Total Least-Squares (TLS) adjustment* are now based on the principle

$$e_y^T P_y e_y + e_A^T P_A e_A = \min . \quad \text{s.t.} \quad (1a)\,, \tag{4}$$

which can be given the equivalent form of a *Lagrange target function*, namely:

$$\begin{aligned}\phi\left(e_y, e_A, \xi, \lambda\right) &:= e_y^T P_y e_y + e_A^T P_A e_A + \\ &+ 2\lambda^T\left[y - A\xi - e_y + \left(\xi^T \otimes I_n\right) e_A\right] = \text{stationary}.\end{aligned} \tag{5}$$

Consequently, the Euler-Lagrange necessary conditions result in the following system of *nonlinear "normal equations"*:

$$\frac{1}{2}\frac{\partial\phi}{\partial e_y} = P_y\tilde{e}_y - \widehat{\lambda} \doteq 0 \tag{6a}$$

$$\frac{1}{2}\frac{\partial\phi}{\partial e_A} = P_A\tilde{e}_A + \left(\widehat{\xi}^T \otimes I_n\right)\widehat{\lambda} \doteq 0, \tag{6b}$$

$$\frac{1}{2}\frac{\partial\phi}{\partial \xi} = -\left(A - \tilde{E}_A\right)^T\widehat{\lambda} \doteq 0, \tag{6c}$$

$$\frac{1}{2}\frac{\partial\phi}{\partial \lambda} = y - A\widehat{\xi} - \tilde{e}_y + \left(\widehat{\xi}^T \otimes I_n\right)\tilde{e}_A \doteq 0, \tag{6d}$$

which still needs to be reduced by partial elimination since the sufficient condition is fulfilled as

$$\frac{1}{2}\frac{\partial^2\phi}{\partial\begin{bmatrix} e_y \\ e_A \end{bmatrix}\partial\left[\, e_y^T \,\middle|\, e_A^T \,\right]} = \begin{bmatrix} P_y & 0 \\ 0 & P_A \end{bmatrix} \text{ is pos.-def.} \tag{7}$$

Now, (6a, b) are transformed to provide the *residual vectors* through

$$\tilde{e}_y = Q_y\widehat{\lambda} \ \text{ and } \ \tilde{e}_A = -Q_A\left(\widehat{\xi} \otimes I_n\right)\widehat{\lambda} \tag{8a}$$

so that (6d) can be rewritten as

$$y - A\widehat{\xi} = \left[Q_y + \left(\widehat{\xi} \otimes I_n\right)^T Q_A \left(\widehat{\xi} \otimes I_n\right)\right] \cdot \widehat{\lambda} =: Q_1 \cdot \widehat{\lambda}, \tag{8b}$$

with $Q_1 = Q_1\left(\widehat{\xi}\right)$ being *nonsingular*, thus leading to

$$\widehat{\lambda} = Q_1^{-1}\left(y - A\widehat{\xi}\right) \tag{9}$$

and, together with (6c), to the system

$$\begin{bmatrix} Q_1 & (A - \tilde{E}_A) \\ (A - \tilde{E}_A)^T & 0 \end{bmatrix} \begin{bmatrix} \widehat{\lambda} \\ \widehat{\xi} \end{bmatrix} = \begin{bmatrix} y - \tilde{E}_A\widehat{\xi} \\ 0 \end{bmatrix} \tag{10}$$

Obviously, the estimated parameter vector is now obtained as in Fang (2011, p.27) via

$$\widehat{\xi} = \left[(A - \tilde{E}_A)^T Q_1^{-1}(A - \tilde{E}_A)\right]^{-1}(A - \tilde{E}_A)^T Q_1^{-1} \cdot \cdot\left(y - \tilde{E}_A\widehat{\xi}\right) \tag{11}$$

and allows updates for Q_1, $\widehat{\lambda}$, and $\tilde{e}_A$, from which a new estimate $\widehat{\xi}$ results.

The Total Sum of weighted Squared Residuals (TSSR) may now readily be computed from

$$\tilde{e}_y^T P_y \tilde{e}_y + \tilde{e}_A^T P_A \tilde{e}_A = \widehat{\lambda}^T \left[Q_y + \left(\widehat{\xi} \otimes I_n\right)^T Q_A \left(\widehat{\xi} \otimes I_n\right)\right]\widehat{\lambda} = = \widehat{\lambda}^T Q_1 \widehat{\lambda} = \widehat{\lambda}^T\left(y - A\widehat{\xi}\right) =: \text{TSSR} \tag{12}$$

so that a suitable variance component estimate may be obtained through

$$\widehat{\sigma}_o^2 = \widehat{\lambda}^T\left(y - A\widehat{\xi}\right)/(n-m) = \text{TSSR}/\text{(n-m)} \tag{13}$$

as the redundancy in model (1a, b) is still *n-m*.

Alternatively, system (10) can be given the *asymmetric* form

$$\begin{bmatrix} Q_1 & A \\ (A - \tilde{E}_A)^T & 0 \end{bmatrix} \begin{bmatrix} \widehat{\lambda} \\ \widehat{\xi} \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix} \tag{14}$$

which would then provide the estimated parameter vector through

$$\widehat{\xi} = \left[(A - \tilde{E}_A)^T Q_1^{-1} A\right]^{-1}(A - \tilde{E}_A)^T Q_1^{-1} y \tag{15}$$

and should lead to a similar iteration as before. Note that (15) also appears as formula (21) in Xu et al. (2012), but essentially represents a variant of Fang's algorithm; also, cf. Fang (2013) where further alternatives are presented.

2.2 Mahboub's Algorithm

On the other hand, combining (9) with (6c) leads to the following sequence of identities:

$$\begin{aligned} A^T Q_1^{-1}\left(y - A\widehat{\xi}\right) &= A^T\widehat{\lambda} = \tilde{E}_A^T\widehat{\lambda} = \left(\widehat{\lambda}^T \otimes I_m\right) vec\left(\tilde{E}_A^T\right) = \\ &= \left(\widehat{\lambda}^T \otimes I_m\right)\cdot(K\tilde{e}_A) = \left(I_m \otimes \widehat{\lambda}^T\right)\tilde{e}_A = \\ &= -\left[\left(I_m \otimes \widehat{\lambda}\right)^T Q_A\left(\widehat{\xi} \otimes I_n\right)\right]\widehat{\lambda} =: -R_1 \cdot \widehat{\lambda} = \\ &= -R_1 \cdot Q_1^{-1}\left(y - A\widehat{\xi}\right) \end{aligned} \tag{16}$$

where K denotes a $nm \times nm$ "*commutation matrix*" that is also known as "*vec-permutation matrix*"; for more details, see Magnus and Neudecker (2007).

Obviously, (16) translates into the estimated parameter vector

$$\widehat{\xi} = \left[\left(A^T + R_1\right)Q_1^{-1}A\right]^{-1}\left(A^T + R_1\right)Q_1^{-1}y \tag{17a}$$

with $R_1 = R_1\left(\widehat{\xi}, \widehat{\lambda}\right)$ and, from (16), with

$$R_1\widehat{\lambda} = -\tilde{E}_A^T\widehat{\lambda} \tag{17b}$$

without necessarily implying that $R_1 = -\tilde{E}_A^T$. Therefore, the sequence of solutions to (15) may differ from the sequence of solutions to (17a) when iteratively updating Q_1, $\widehat{\lambda}$, and R_1, before a new parameter vector estimate $\widehat{\xi}$ can be found; yet the ultimate convergence points will be the same.

Again, the TSSR can be computed from (12) which will lead to the variance component estimate in (13).

2.3 A New Variant of Mahboub's Algorithm

After giving (16) the form

$$A^T Q_1^{-1}\left(y - A\widehat{\xi}\right) = -\left[\left(I_m \otimes \widehat{\lambda}\right)^T Q_A\left(I_m \otimes \widehat{\lambda}\right)\right]\widehat{\xi}, \tag{18a}$$

the estimated parameter vector may as well be obtained from

$$\widehat{\xi} = \left[A^T Q_1^{-1} A - \left[\left(I_m \otimes \widehat{\lambda}\right)^T Q_A\left(I_m \otimes \widehat{\lambda}\right)\right]\right]^{-1} A^T Q_1^{-1} y \tag{18b}$$

thus allowing updates for Q_1 and $\widehat{\lambda}$. This algorithm will be further explored in the near future.

2.4 The Schaffrin–Wieser Algorithm

This algorithm was designed for the somewhat more special case where the cofactor matrix Q_A can be split into a *Kronecker product*, thereby indicating that all columns have cofactor matrices proportional to each other. This implies

$$Q_A = Q_o \otimes Q_x \Rightarrow Q_1 = Q_y + \left(\widehat{\xi}^T Q_o \widehat{\xi}\right) \cdot Q_x \tag{19}$$

and thus

$$A^T Q_1^{-1}\left(y - A\widehat{\xi}\right) = A^T\left[Q_y + \left(\widehat{\xi}^T Q_o\widehat{\xi}\right)\cdot Q_x\right]^{-1} \cdot \left(y - A\widehat{\xi}\right) = A^T\widehat{\lambda} = \tilde{E}_A^T\widehat{\lambda} \tag{20}$$

with

$$\tilde{e}_A = -\left(Q_o\widehat{\xi} \otimes Q_x\right)\widehat{\lambda} = -vec\left(Q_x\widehat{\lambda}\widehat{\xi}^T Q_o\right) = vec\tilde{E}_A. \tag{21}$$

(20) and (21) together generate the identity

$$A^T\left[Q_y + \left(\widehat{\xi}^T Q_o\widehat{\xi}\right)\cdot Q_x\right]^{-1}\left(y - A\widehat{\xi}\right) = = -Q_o\widehat{\xi}\cdot\left(\widehat{\lambda}^T Q_x\widehat{\lambda}\right) =: -Q_o\widehat{\xi}\cdot\widehat{\nu} \tag{22a}$$

suggesting the *iteration*

$$\widehat{\xi} = \left(A^T\left[Q_y + \left(\widehat{\xi}^T Q_o\widehat{\xi}\right)\cdot Q_x\right]^{-1} A - Q_o\widehat{\nu}\right)^{-1} \cdot A^T\left[Q_y + \left(\widehat{\xi}^T Q_o\widehat{\xi}\right)\cdot Q_x\right]^{-1} y \tag{22b}$$

with

$$\widehat{\nu} := \left(\widehat{\lambda}^T Q_x\widehat{\lambda}\right) \quad \text{and} \quad \widehat{\lambda} := \left[Q_y + \left(\widehat{\xi}^T Q_o\widehat{\xi}\right)\cdot Q_x\right]^{-1} \cdot \left(y - A\widehat{\xi}\right) \tag{22c}$$

while (12) and (13) generate first the TSSR and then a suitable variance component estimate.

2.5 The Golub-van-Loan Algorithm

Now, the condition (19) is further specialized to

$$Q_x := Q_y \Rightarrow Q_1 = \left(1 + \widehat{\xi}^T Q_o\widehat{\xi}\right)\cdot Q_y \tag{23a}$$

and

$$\widehat{\lambda} = \left(1 + \widehat{\xi}^T Q_o\widehat{\xi}\right)^{-1} Q_y^{-1}\left(y - A\widehat{\xi}\right) \tag{23b}$$

so that (22a) becomes

$$A^T Q_y^{-1}\left(y - A\widehat{\xi}\right) = -Q_o\widehat{\xi}\cdot\widehat{\nu}\left(1 + \widehat{\xi}^T Q_o\widehat{\xi}\right) =: -Q_o\widehat{\xi}\cdot\sigma^2_{\min} \tag{24a}$$

with

$$\sigma^2_{\min} = \left(\widehat{\lambda}^T Q_y\widehat{\lambda}\right)\left(1 + \widehat{\xi}^T Q_o\widehat{\xi}\right) = = \left(y - A\widehat{\xi}\right)^T Q_y^{-1}\left(y - A\widehat{\xi}\right) / \left(1 + \widehat{\xi}^T Q_o\widehat{\xi}\right), \tag{24b}$$

and this, from (24a, b), becomes

$$\sigma^2_{\min}\cdot\left(1 + \widehat{\xi}^T Q_o\widehat{\xi}\right) = y^T Q_y^{-1}\left(y - A\widehat{\xi}\right) + \left(\widehat{\xi}^T Q_o\widehat{\xi}\right)\cdot\sigma^2_{\min} \Rightarrow \Rightarrow \sigma^2_{\min} = y^T Q_y^{-1}\left(y - A\widehat{\xi}\right) = \text{TSSR} \tag{24c}$$

(24a) and (24c) allow the problem to be rephrased as a *generalized eigenvalue problem*, specifically as:

$$\begin{bmatrix} A^T Q_y^{-1} A & A^T Q_y^{-1} y \\ y^T Q_y^{-1} A & y^T Q_y^{-1} y \end{bmatrix} \begin{bmatrix} \widehat{\xi} \\ -1 \end{bmatrix} = \begin{bmatrix} Q_o & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \widehat{\xi} \\ -1 \end{bmatrix} \cdot \sigma^2_{\min} \tag{25}$$

with the variance component estimate

$$\widehat{\sigma}_o^2 = \sigma^2_{\min} / (n - m) \tag{26}$$

The original situation, treated by Golub and van Loan (1980), was characterized by the further specializations

$$Q_o := I_m \text{ and } Q_y := \text{Diag}\left(p_1^{-1}, \ldots\ldots\ldots, p_n^{-1}\right) = P^{-1} \tag{27}$$

which, in turn, lead to the *standard eigenvalue problem*

$$\begin{bmatrix} A^T PA & A^T Py \\ y^T PA & y^T Py \end{bmatrix} \begin{bmatrix} \widehat{\xi} \\ -1 \end{bmatrix} = \begin{bmatrix} \widehat{\xi} \\ -1 \end{bmatrix} \cdot \sigma^2_{\min} \tag{28}$$

whose solution provides the *Total Least-Squares Solution* (*TLSS*).

In the next section, a few *equivalent models* will be presented for which, traditionally, an identical weighted LEast-Squares Solution (LESS) would have been found after *iterative linearization*.

3 Traditional Models, Equivalent to the EIV-Model

3.1 The Nonlinear Gauss–Helmert Model

Here, the new vectors

$$Y := vec\,[y|\,A] \quad \text{and} \quad e := vec\left[e_y\right|E_A\right] \tag{29}$$

are introduced. Then,

$$\underset{\sim}{b}\left(\mu := \underbrace{Y-e}_{n(m+1)\times 1}, \xi\right) := (y-e_y) - (A-E_A)\,\xi = 0, \tag{30a}$$

$$e \sim \left(0, \sigma_o^2 \begin{bmatrix} Q_y & 0 \\ 0 & Q_A \end{bmatrix} = \sigma_o^2 Q\right), \tag{30b}$$

with the *nonlinear* vector-valued vector function

$$\underset{\sim}{b} : R^{(n+1)(m+1)-1} \to R^n, \tag{30c}$$

due to the term $E_A \cdot \xi$, forms an equivalent *Gauss–Helmert Model* that would traditionally be linearized for an iterative Least-Squares adjustment.

The truncated Taylor series, following Pope (1972), then reads:

$$0 = \underset{\sim}{b}(\mu,\xi) \approx \underset{\sim}{b}(\mu_o,\xi_o) + \left.\frac{\partial \underset{\sim}{b}(\mu,\xi)}{\partial\left[\mu^T|\,\xi^T\right]}\right|_{\mu_o,\xi_o} \cdot \begin{bmatrix} \mu-\mu_o \\ \xi-\xi_o \end{bmatrix} \tag{31}$$

with suitable approximations ξ_o and $\mu_o := Y - \underset{\sim}{0}$ where $\underset{\sim}{0}$ here denotes a "*stochastic zero vector*" of size $n\,(m+1)\times 1$. This leads first to

$$0 \approx \underset{\sim}{b}(\mu_o,\xi_o) + \left.\frac{\partial \underset{\sim}{b}}{\partial \mu^T}\right|_{\mu_o,\xi_o} \cdot \left(\underset{\sim}{0} - e\right) + \left.\frac{\partial \underset{\sim}{b}}{\partial \xi^T}\right|_{\mu_o,\xi_o} \cdot (\xi-\xi_o)\,, \tag{32}$$

then to

$$\underset{\sim}{b}(\mu_o,\xi_o) + \left.\frac{\partial \underset{\sim}{b}}{\partial \mu^T}\right|_{\mu_o,\xi_o} \cdot \underset{\sim}{0} \approx \left.\frac{\partial \underset{\sim}{b}}{\partial \mu^T}\right|_{\mu_o,\xi_o} \cdot e - \left.\frac{\partial \underset{\sim}{b}}{\partial \xi^T}\right|_{\mu_o,\xi_o} \cdot (\xi-\xi_o)\,, \tag{33}$$

and eventually to the *linearized Gauss–Helmert Model:*

$$w_o := \underset{\sim}{b}\left(Y = \mu_o + \underset{\sim}{0}, \xi_o\right) \approx B^{(o)} \cdot e + A^{(o)} \cdot (\xi - \xi_o)\,, \tag{34a}$$

$$B^{(o)} := \left[I_n| - (\xi_o^T \otimes I_n)\right], \quad A^{(o)} := A - \underset{\sim}{0}, \tag{34b}$$

$$e \sim \left(0, \sigma_o^2 Q = \sigma_o^2 \begin{bmatrix} Q_y & 0 \\ 0 & Q_A \end{bmatrix}\right). \tag{34c}$$

Note that the weighted LEast-Squares Solution (LESS) is now being formed through the *normal equations*

$$\left[\left(A^{(o)}\right)^T \left(Q_1^{(o)}\right)^{-1} A^{(o)}\right] \cdot \widehat{\xi}^{(1)} = \left(A^{(o)}\right)^T \left(Q_1^{(o)}\right)^{-1} \cdot \left(y - \underset{\sim}{0} \cdot \xi_o\right) \tag{35a}$$

with

$$Q_1^{(o)} := B^{(o)} Q \left(B^{(o)}\right)^T = Q_y + (\xi_o \otimes I_n)^T Q_A (\xi_o \otimes I_n)\,, \tag{35b}$$

and the *residual vectors* through

$$\tilde{e}^{(1)} := \begin{bmatrix} \tilde{e}_y^{(1)} \\ \tilde{e}_A^{(1)} \end{bmatrix} = \begin{bmatrix} Q_y \\ -Q_A\left(\widehat{\xi}^{(1)} \otimes I_n\right) \end{bmatrix} \left(Q_1^{(o)}\right)^{-1} \cdot \left(y - \underset{\sim}{0} \cdot \xi_o - A^{(o)} \cdot \widehat{\xi}^{(1)}\right) \tag{35c}$$

Looking at the next and all the following iteration steps, it becomes clear that this represents one specific *iterative solver* of Fang's TLS normal equations (11).

For more details, see Fang (2011, ch. 4.4), Snow (2012, ch. 4), and the forthcoming OSU-Report by Schaffrin (2015), as well as Neitzel (2010) for a specific application.

3.2 The Nonlinear Gauss–Markov Model

In this case, the expectation of the data matrix A is introduced as a new $n \times m$ "*parameter matrix*"

$$\Xi_A := A - E_A \quad \text{with} \quad \xi_A := vec\,\Xi_A, \tag{36}$$

leading to the equivalent *Gauss-Markov Model*

$$y = (\xi \otimes I_n)^T \cdot \xi_A + e_y =: \underset{\sim}{a}(\xi, \xi_A) + e_y, \quad e_y \sim \left(0, \sigma_o^2 Q_y\right), \tag{37a}$$

with the *nonlinear* vector-valued vector function

$$\underset{\sim}{a} : R^{(n+1)m} \to R^n \tag{37b}$$

due to the term $\Xi_A \cdot \xi$. The linearization of model (37a, b) with respect to the approximations ξ_o and $\xi_A^{(o)} := vec\left(A^{(o)}\right) = vec\left(A - \underset{\sim}{0}\right)$, where $\underset{\sim}{0}$ now denotes a "*stochastic zero matrix*" of size $n \times m$, then leads first to

$$\xi_A - \xi_A^{(o)} = \underset{\sim}{0} - e_A, \quad e_A \sim \left(0, \sigma_o^2 Q_A\right), \tag{38a}$$

$$y - e_y \approx \underset{\sim}{a}\left(\xi_o, \xi_A^{(o)}\right) + \left.\frac{\partial \underset{\sim}{a}\left(\xi, \xi_A\right)}{\partial\left[\xi^T, \xi_A^T\right]}\right|_{\xi_o, \xi_A^{(o)}} \cdot \begin{bmatrix} \xi - \xi_o \\ \xi_A - \xi_A^{(o)} \end{bmatrix}$$
$$= A^{(o)} \cdot \xi_o + A^{(o)} \cdot (\xi - \xi_o) + (\xi_o \otimes I_n)^T \left(\xi_A - \xi_A^{(o)}\right), \tag{38b}$$

and finally to the *linearized Gauss–Markov Model*

$$\begin{bmatrix} y - A^{(o)} \cdot \xi_o \\ \underset{\sim}{0} \end{bmatrix} = \begin{bmatrix} A^{(o)} & (\xi_o \otimes I_n)^T \\ 0 & I_{nm} \end{bmatrix} \begin{bmatrix} \xi - \xi_o \\ \xi_A - \xi_A^{(o)} \end{bmatrix} + \begin{bmatrix} e_y \\ e_A \end{bmatrix}, \tag{39a}$$

$$\begin{bmatrix} e_y \\ e_A \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \ \sigma_o^2 \begin{bmatrix} Q_y & 0 \\ 0 & Q_A \end{bmatrix} = \sigma_o^2 Q \right). \tag{39b}$$

After a number of further manipulations, the weighted LESS for model (39a, b) can be shown to fulfill the "*normal equations*"

$$\left[\left(A^{(o)}\right)^T \left(Q_1^{(o)}\right)^{-1} A^{(o)}\right] \cdot \hat{\xi}^{(1)} = \left(A^{(o)}\right)^T \left(Q_1^{(o)}\right)^{-1} y \tag{40a}$$

with

$$Q_1^{(o)} := \left[Q_y^{-1} - Q_y^{-1}(\xi_o \otimes I_n)^T \cdot \left[Q_A^{-1} + \left(\xi_o \xi_o^T \otimes Q_y^{-1}\right)\right]^{-1} (\xi_o \otimes I_n)\, Q_y^{-1}\right]^{-1} \tag{40b}$$

$$= Q_y + (\xi_o \otimes I_n)^T Q_A (\xi_o \otimes I_n) \tag{40c}$$

which nicely corresponds to (35a, b). More details can be found in the forthcoming OSU-Report by Schaffrin (2015).

3.3 The Model of Direct Observations with Nonlinear Constraints

Now, the expectation of the observation vector y is introduced as just another parameter vector ξ_y of size $n \times 1$ so that the new model combines the *direct observation equations*

$$\begin{bmatrix} y \\ vec A \end{bmatrix} = \begin{bmatrix} \xi_y \\ \xi_A \end{bmatrix} + \begin{bmatrix} e_y \\ e_A \end{bmatrix}, \quad \begin{bmatrix} e_y \\ e_A \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \ \sigma_o^2 \begin{bmatrix} Q_y & 0 \\ 0 & Q_A \end{bmatrix} = \sigma_o^2 Q \right), \tag{41a}$$

with the *nonlinear constraints*

$$\xi_y - \Xi_A \cdot \xi = 0 \tag{41b}$$

which might be linearized into

$$\begin{bmatrix} A^{(o)} \,|\, -I_n \,|\, (\xi_o \otimes I_n)^T \end{bmatrix} \begin{bmatrix} \xi - \xi_o \\ \xi_y - \xi_y^{(o)} \\ \xi_A - \xi_A^{(o)} \end{bmatrix} = 0. \tag{42}$$

In the already mentioned OSU-Report by Schaffrin (2015), it will be shown how the resulting iterative LESS's do converge to the Total Least-Squares Solution.

For another take on this model, refer to Donevska et al. (2011) who stress the equivalence to orthogonal regression as applied by Deming (1931, 1934).

4 Conclusions

It has been clarified that the TLS approach towards the EIV-Model requires a *nonlinear treatment* of the *nonlinear model*. A number of different algorithms have been presented to generate the Total Least-Squares Solution from a certain set of nonlinear normal equations. A triplet of conventional nonlinear models has also been considered, suggesting that the LEast-Squares Solutions from iterative linearization do converge to the nonlinear TLS-Solution in all three cases. Most of the details, however, will be published in a forthcoming OSU-Report, due to the space restrictions for these Proceedings.

Acknowledgment This author is very much indebted to the thorough reading of the text by three anonymous reviewers. Their recommendations have improved the readibility quite substantially. Also appreciated are many discussions with his long-time collaborator Kyle Snow.

References

Deming WE (1931) The application of least squares. Phil Mag 11:146–158

Deming WE (1934) On the application of least squares II. Phil Mag 17:804–829

Donevska S, Fiśerova E, Horn K (2011) On the equivalence between orthogonal regression and linear model with type II constraints. Acta Univ Palacki Olomuc, Fac rer nat, Mathematica 50(2):19–27

Fang X (2011) Weighted Total Least-Squares solutions with applications in geodesy. Publ. No. 294, Department of Geodesy and Geoinformatics, Leibniz University, Hannover

Fang X (2013) Weighted Total Least-Squares: necessary and sufficient conditions, fixed and random parameters. J Geodesy 87:733–749

Golub GH, van Loan CF (1980) An analysis of the Total Least-Squares problem. SIAM J Numer Anal 17:883–893

Helmert FR (1907) Adjustment computations based on the Least-Squares Principle (in German), 2nd edn. Teubner, Leipzig

Magnus JR, Neudecker H (2007) Matrix differential calculus with applications in statistics and economics, 3rd edn. Wiley, Chichester

Mahboub V (2012) On weighted Total Least-Squares for geodetic transformations. J Geodesy 86:359–367

Neitzel F (2010) Generalization of Total Least-Squares on example of unweighted and weighted 2D similarity transformation. J Geodesy 84:751–762

Pope A (1972) Some pitfalls to be avoided in the iterative adjustment of nonlinear problems. In: Proceedings of the 38th Ann. ASPRS Meeting, Amer. Soc. of Photogrammetry: Falls Church, pp 449–477

Schaffrin B, Wieser A (2008) On weighted Total Least-Squares adjustment for linear regression. J Geodesy 82:415–421

Schaffrin B (2015) The Errors-In-Variables (EIV) model. Nonlinear Total Least-Squares (TLS) adjustment or iteratively linearized Least-Squares (LS) adjustment? OSU-Report, Division of Geodetic Science, School of Earth Sciences, The Ohio State University, Columbus

Schaffrin B, Snow K, Neitzel F (2014) On the Errors-In-Variables model with singular covariance matrices. J Geodetic Sci 4:28–36

Snow K (2012) Topics in Total Least-Squares adjustment within the Errors-In-Variables model: singular covariance matrices and prior information, Report No. 502, Division of Geodetic Science, School of Earth Sciences, The Ohio State University, Columbus

Xu P, Liu JN, Shi C (2012) Total Least-Squares adjustment in partial Errors-In-Variables models. Algorithms and statistical analysis. J Geodesy 86:661–675