TOWARDS A MORE RIGOROUS ERROR PROPAGATION WITHIN THE ERRORS-IN-VARIABLES MODEL FOR APPLICATIONS IN GEODETIC NETWORKS

Burkhard Schaffrin[1], Kyle Snow[1,2]

[1]Geodetic Science Program, School of Earth Sciences, The Ohio State University, Columbus, Ohio, USA (aschaffrin@earthlink.net)
[2]Topcon Positioning Systems, Inc., Columbus, Ohio, USA (ksnow6378137@gmail.com)

Key words: Errors-In-Variables model; Total Least-Squares adjustments; error propagation; geodetic network analysis

ABSTRACT

The use of the Errors-In-Variables (EIV) model for geodetic applications, along with the corresponding Total Least-Squares (TLS) adjustment, has been around for over a decade; see Schaffrin and Wieser (2008) who pioneered this method in geodetic science along with Schaffrin and Felus (2005). But, beyond the successful derivation of nonlinear formulas for TLS estimates of both the parameters and the variance component, a similarly elegant formula for the variance-covariance matrix of the estimated parameters is still missing. Moreover, a first attempt for an approximate representation by Amiri-Simkooei and Jazaeri (2012) had to be dismissed by Schaffrin and Snow (2014) shortly thereafter as being non-satisfactory due to the neglection of various sources of random errors in the nonlinear relationship between the data and the parameter estimates.

Here, we shall try to compensate for some of these neglections, while still relying on a linearized approach, as our attempts for fully nonlinear error propagation were unsuccessful so far. An example from geodetic network analysis will illustrate the new, more rigorous formulas.

INTRODUCTION

In the following, we present four different models for handling unknown, fixed parameters with random errors in both dependent and independent data variables. The first three models are well represented in the classic geodetic and deformation-monitoring literature, while the fourth one arrived on the scene more recently. The models treated in turn are (1) the nonlinear Gauss-Helmert model, (2) the nonlinear Gauss-Markov model, (3) the model of direct observations with nonlinear constraints, and (4) the Errors-In-Variables (EIV) model.

In so far as the models can treat the same unknown parameters and data variables, and the LEast-Squares Solutions (LESSs) within the models generate the same vectors of estimated parameters and predicted residuals, it can be said that they are equivalent. However, the dispersion (variances-covariance) matrix for the estimated parameters turns out to be somewhat different within the EIV model than within the other three, if the formula that we derived for it is used.

One motivation for presenting all four models is our view that seeing a problem approached from different angles can, for the attentive reader, serve to deepen insight into the problem. However, our primary purpose is to contrast the dispersion matrix for the estimated parameters computed within the nonlinear EIV model with the one that is computed from any of the other three (which are equivalent to each other) after iterative linearization.

In Section 1, we present the four models, providing precise definitions of all terms involved. In Section 2, we present the least-squares parameter estimates within each model, as well as the predicted residual vectors. In Section 3, a numerical example involving a 1-D similarity transformation is shown for illustrative purposes. Finally, some conclusions and an outlook for future work are given in Section 4.

1. MODEL DEFINITIONS

The function components of four equivalent data models under consideration are enumerated below. They all have the same stochastic properties, expressed in (1b).

1. The nonlinear Gauss-Helmert model is written as

$$\underset{n\times 1}{\boldsymbol{y}} = \underset{n\times m}{A}\,\boldsymbol{\xi} + \left[I_n \;\middle|\; -(\boldsymbol{\xi}\otimes I_n)^T\right] \begin{bmatrix} \underset{n\times 1}{\boldsymbol{e}_y} \\ \underset{nm\times 1}{\boldsymbol{e}_A} \end{bmatrix}, \tag{1a}$$

$$\boldsymbol{e} = \begin{bmatrix} \boldsymbol{e}_y \\ \boldsymbol{e}_A \end{bmatrix} \sim \left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \sigma_0^2 \begin{bmatrix} Q_y & 0 \\ 0 & Q_A \end{bmatrix} = \sigma_0^2 Q\right). \tag{1b}$$

2. The nonlinear Gauss-Markov model is written as

$$\underset{n\times m}{\Xi_A} := A - E_A \Rightarrow \tag{2a}$$

$$\Rightarrow \begin{cases} \underset{n\times 1}{\boldsymbol{y}} = (\boldsymbol{\xi}^T \otimes I_n)\ \underset{nm\times 1}{\boldsymbol{\xi}_A} + \boldsymbol{e}_y, \\ \operatorname{vec} A = \boldsymbol{\xi}_A + \boldsymbol{e}_A,\ \ \boldsymbol{\xi}_A = \operatorname{vec}\Xi_A, \\ \boldsymbol{e}_A = \operatorname{vec} E_A, \end{cases} \tag{2b}$$

with (1b) as its stochastic component.

3. The model of direct observations with non-linear constraints is written as

$$\underset{n\times 1}{\boldsymbol{\mu}_y} := \boldsymbol{y} - \boldsymbol{e}_y \Rightarrow \tag{3a}$$

$$\Rightarrow \begin{cases} \boldsymbol{y} = \boldsymbol{\mu}_y + \boldsymbol{e}_y, \\ \text{vec}\, A = \boldsymbol{\xi}_A + \boldsymbol{e}_A, \\ \boldsymbol{\mu}_y - \Xi_A \cdot \boldsymbol{\xi} = \boldsymbol{0}, \end{cases} \tag{3b}$$

with (1b) as its stochastic component.

4. The Errors-In-Variables (EIV) model is written as

$$\boldsymbol{y} - \boldsymbol{e}_y = (A - E_A)\boldsymbol{\xi}, \tag{4}$$

with (1b) as its stochastic component.

The terms of the models, and their sizes, are defined as follows:

$\boldsymbol{y}$ is a given $n \times 1$ vector of (incremental) observations.

A is a given $n \times m$ matrix of random coefficients ("data matrix") that has full column-rank.

E_A is an $n \times m$ matrix of unknown random errors associated with A.

$\boldsymbol{\xi}$ is an $m \times 1$ vector of unknown (incremental) parameters.

$\boldsymbol{e}_y$ is an $n \times 1$ vector of unknown random errors associated with the observation vector $\boldsymbol{y}$.

$\boldsymbol{e}_A$ is an $nm \times 1$ vector of unknown random errors, being the vectorization of matrix E_A.

$\boldsymbol{\mu}_y$ is an $n \times 1$ vector of true observables corresponding to $\boldsymbol{y}$.

Ξ_A is an unknown $n \times m$ matrix of true observables corresponding to matrix A, and, like A, it has full column-rank.

$\boldsymbol{\xi}_A$ is an unknown $nm \times 1$ vector, being the vectorization of matrix Ξ_A.

σ_0^2 is an unknown variance component (being unit free).

Q_y is a given positive-definite cofactor matrix of size $n \times n$ associated with $\boldsymbol{e}_y$.

Q_A is a given positive-definite cofactor matrix of size $nm \times nm$ associated with $\boldsymbol{e}_A$.

2. Estimated quantities

2.1. Estimates within the Gauss-Helmert model

The Gauss-Helmert model is given in detail as

$$\underset{n\times m}{A}\,\boldsymbol{\xi} + B\boldsymbol{e}, \tag{5a}$$

$$\underset{n\times n(m+1)}{B} := \left[I_n \;\middle|\; -(\boldsymbol{\xi}_0 \otimes I_n)^T\right], \tag{5b}$$

with (1b) as its stochastic component.

Minimization of the random errors in the model, based upon the method of least-squares adjustment, leads to the following estimated and predicted quantities.

The $m \times 1$ vector of estimated parameters is given by

$$\hat{\boldsymbol{\xi}} = \left[A^T(BQB^T)^{-1}A\right]^{-1}A^T(BQB^T)^{-1}\boldsymbol{y} \tag{6a}$$

with its $m \times m$ dispersion matrix provided by

$$D\{\hat{\boldsymbol{\xi}}\} = \sigma_0^2\left[A^T(BQB^T)^{-1}A\right]^{-1} = \tag{6b}$$

$$= \sigma_0^2\left\{A^T[Q_y + (\boldsymbol{\xi}_0 \otimes I_n)^T Q_A(\boldsymbol{\xi}_0 \otimes I_n)]^{-1}A\right\}^{-1}. \tag{6c}$$

Here, the nonrandom vector $\boldsymbol{\xi}_0$ is numerically equivalent to the random estimate $\hat{\boldsymbol{\xi}}$, with randomness having been formally stripped from the former via the relationship

$$\boldsymbol{\xi}_0 = \hat{\boldsymbol{\xi}} - \underset{\sim}{\boldsymbol{0}}, \tag{6d}$$

where $\underset{\sim}{\boldsymbol{0}}$ is called a "*random zero-vector*" after Harville (1986).

In addition, the unknown variance component can be estimated by

$$\hat{\sigma}_0^2 = (B\tilde{\boldsymbol{e}})^T(BQB^T)^{-1}(B\tilde{\boldsymbol{e}})/(n-m), \tag{7a}$$

which depends on the predicted residual vector

$$\tilde{\boldsymbol{e}} = QB^T(BQB^T)^{-1}(\boldsymbol{y} - A\hat{\boldsymbol{\xi}}). \tag{7b}$$

2.2. Estimates within the Gauss-Markov model

The Gauss-Markov model reads

$$\begin{bmatrix} \boldsymbol{y} \\ \text{vec}\, A \end{bmatrix} = \begin{bmatrix} \boldsymbol{\xi}^T \otimes I_n \\ I_{nm} \end{bmatrix} \boldsymbol{\xi}_A + \begin{bmatrix} \boldsymbol{e}_y \\ \boldsymbol{e}_A \end{bmatrix}, \tag{8a}$$

$$\boldsymbol{\xi}_A = \text{vec}\,\Xi_A = \text{vec}(A - E_A), \tag{8b}$$

$$\boldsymbol{y} = \Xi_A \cdot \boldsymbol{\xi} + \boldsymbol{e}_y, \tag{8c}$$

with (1b) as its stochastic component.

Linearization about the approximate points $\boldsymbol{\xi}_0$ and Ξ_A^0 leads to

$$\begin{aligned} \boldsymbol{y} = {} & \Xi_A^0 \cdot \boldsymbol{\xi}_0 + \Xi_A^0 \cdot (\boldsymbol{\xi} - \boldsymbol{\xi}_0) + \\ & + (\boldsymbol{\xi}_0^T \otimes I_n)(\boldsymbol{\xi}_A - \boldsymbol{\xi}_A^0) + \cdots + \boldsymbol{e}_y, \end{aligned} \tag{9a}$$

implying that

$$\begin{aligned} & \begin{bmatrix} \boldsymbol{y} - \Xi_A^0 \cdot \boldsymbol{\xi}_0 \\ \text{vec}(A - \Xi_A^0) \end{bmatrix} \approx \\ & \approx \left[\begin{array}{c|c} \Xi_A^0 & \boldsymbol{\xi}_0^T \otimes I_n \\ 0 & I_{nm} \end{array}\right] \begin{bmatrix} \boldsymbol{\xi} - \boldsymbol{\xi}_0 \\ \boldsymbol{\xi}_A - \boldsymbol{\xi}_A^0 \end{bmatrix} + \begin{bmatrix} \boldsymbol{e}_y \\ \boldsymbol{e}_A \end{bmatrix}. \end{aligned} \tag{9b}$$

Then, it follows that the unknown parameters (minus the approximate points) are estimated by

$$\begin{aligned} \begin{bmatrix} \hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0 \\ \hat{\boldsymbol{\xi}}_A - \boldsymbol{\xi}_A^0 \end{bmatrix} = {} & \left(\left[\begin{array}{c|c} (\Xi_A^0)^T & 0 \\ \boldsymbol{\xi}_0 \otimes I_n & I_{nm} \end{array}\right] \begin{bmatrix} Q_y^{-1} & 0 \\ 0 & Q_A^{-1} \end{bmatrix} \cdot \right. \\ & \left. \cdot \left[\begin{array}{c|c} \Xi_A^0 & \boldsymbol{\xi}_0^T \otimes I_n \\ 0 & I_{nm} \end{array}\right] \right)^{-1} \cdot \\ & \cdot \left[\begin{array}{c|c} (\Xi_A^0)^T & 0 \\ \boldsymbol{\xi}_0 \otimes I_n & I_{nm} \end{array}\right] \begin{bmatrix} Q_y^{-1} & 0 \\ 0 & Q_A^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{y} - \Xi_A^0 \cdot \boldsymbol{\xi}_0 \\ \text{vec}(A - \Xi_A^0) \end{bmatrix}, \end{aligned} \tag{9c}$$

with their $m \times m$ dispersion matrix provided by

$$
\begin{aligned}
&D\{\begin{bmatrix}\hat{\boldsymbol{\xi}}-\boldsymbol{\xi}_0\\ \hat{\boldsymbol{\xi}}_A-\boldsymbol{\xi}_A^0\end{bmatrix}\}=\\
&=\sigma_0^2\left(\left[\begin{array}{c|c}(\Xi_A^0)^T & 0\\ \boldsymbol{\xi}_0\otimes I_n & I_{nm}\end{array}\right]\left[\begin{array}{c|c}Q_y^{-1}\Xi_A^0 & \boldsymbol{\xi}_0^T\otimes Q_y^{-1}\\ 0 & Q_A^{-1}\end{array}\right]\right)^{-1}\\
&=\sigma_0^2\left[\begin{array}{c|c}(\Xi_A^0)^TQ_y^{-1}\Xi_A^0 & \boldsymbol{\xi}_0^T\otimes(\Xi_A^0)^TQ_y^{-1}\\ \boldsymbol{\xi}_0\otimes Q_y^{-1}\Xi_A^0 & \boldsymbol{\xi}_0\boldsymbol{\xi}_0^T\otimes Q_y^{-1}+Q_A^{-1}\end{array}\right]^{-1},
\end{aligned}
\tag{9d}
$$

finally leading to the matrix

$$
\begin{aligned}
&D\{\hat{\boldsymbol{\xi}}\}=\sigma_0^2\big[(\Xi_A^0)^TQ_y^{-1}\Xi_A^0-(\boldsymbol{\xi}_0\otimes Q_y^{-1}\Xi_A^0)^T\cdot\\
&\cdot(\boldsymbol{\xi}_0\boldsymbol{\xi}_0^T\otimes Q_y^{-1}+Q_A^{-1})^{-1}(\boldsymbol{\xi}_0\otimes Q_y^{-1}\Xi_A^0)\big]^{-1}
\end{aligned}
\tag{9e}
$$

being the Schur complement of the lower principle submatrix of (9d).

The estimated variance component is provided by

$$
\hat{\sigma}_0^2=(\tilde{\boldsymbol{e}}_y^TQ_y^{-1}\tilde{\boldsymbol{e}}_y+\tilde{\boldsymbol{e}}_A^TQ_A^{-1}\tilde{\boldsymbol{e}}_A)/(n-m), \tag{9f}
$$

where

$$
\tilde{\boldsymbol{e}}_A=\operatorname{vec}A-\hat{\boldsymbol{\xi}}_A \text{ and } \tilde{\boldsymbol{e}}_y=\boldsymbol{y}-(A-\tilde{E}_A)\hat{\boldsymbol{\xi}} \tag{9g}
$$

are the respective vectors of predicted residuals.

2.3. Estimates within the model of direct observations with non-linear constraints

The model can be stated as

$$
\left.\begin{array}{l}\boldsymbol{y}=\boldsymbol{\mu}_y+\boldsymbol{e}_y\\ \operatorname{vec}A=\boldsymbol{\xi}_A+\boldsymbol{e}_A\end{array}\right\}\Rightarrow
\begin{bmatrix}\boldsymbol{y}\\ \operatorname{vec}A\end{bmatrix}=\left[\begin{array}{c|c}I_n & 0\\ 0 & I_{nm}\end{array}\right]\begin{bmatrix}\boldsymbol{\mu}_y\\ \boldsymbol{\xi}_A\end{bmatrix}+\begin{bmatrix}\boldsymbol{e}_y\\ \boldsymbol{e}_A\end{bmatrix},
\tag{10a}
$$

with

$$
\boldsymbol{\mu}_y-(\boldsymbol{\xi}^T\otimes I_n)\boldsymbol{\xi}_A=\mathbf{0}, \tag{10b}
$$

and with (1b) as its stochastic component.

Linearization allows formation of the equation

$$
\begin{aligned}
\underset{n\times 1}{\boldsymbol{\mu}_y}-\Xi_A^0\cdot\boldsymbol{\xi}_0&\approx\underset{m\times 1}{\Xi_A^0(\boldsymbol{\xi}-\boldsymbol{\xi}_0)}+\\
&+\underset{nm\times 1}{(\boldsymbol{\xi}_0^T\otimes I_n)(\boldsymbol{\xi}_A-\boldsymbol{\xi}_A^0)},
\end{aligned}
\tag{11a}
$$

which implies that

$$
\begin{aligned}
&\big[(\Xi_A^0)^TQ_y^{-1}\Xi_A^0\big]\cdot(\boldsymbol{\xi}-\boldsymbol{\xi}_0)=(\Xi_A^0)^TQ_y^{-1}\cdot\\
&\cdot(\boldsymbol{\mu}_y-\Xi_A^0\boldsymbol{\xi}_0)-(\boldsymbol{\xi}_0\otimes Q_y^{-1}\Xi_A^0)^T(\boldsymbol{\xi}_A-\boldsymbol{\xi}_A^0),
\end{aligned}
\tag{11b}
$$

leading to the estimated quantity

$$
\begin{aligned}
&\hat{\boldsymbol{\xi}}-\boldsymbol{\xi}_0=\\
&=\big[(\Xi_A^0)^TQ_y^{-1}\Xi_A^0\big]^{-1}(\Xi_A^0)^TQ_y^{-1}(\hat{\boldsymbol{\mu}}_y-\Xi_A^0\boldsymbol{\xi}_0)-\\
&-\big[\boldsymbol{\xi}_0^T\otimes\big[(\Xi_A^0)^TQ_y^{-1}\Xi_A^0\big]^{-1}(\Xi_A^0)^TQ_y^{-1}\big](\hat{\boldsymbol{\xi}}_A-\boldsymbol{\xi}_A^0).
\end{aligned}
\tag{11c}
$$

Note that the equation

$$
\begin{aligned}
&\big\{I_n-\Xi_A^0\big[(\Xi_A^0)^TQ_y^{-1}\Xi_A^0\big]^{-1}(\Xi_A^0)^TQ_y^{-1}\big\}\cdot\\
&\cdot(\hat{\boldsymbol{\mu}}_y-\Xi_A^0\boldsymbol{\xi}_0)-\\
&-\big(\boldsymbol{\xi}_0^T\otimes\big\{I_n-\Xi_A^0\big[(\Xi_A^0)^TQ_y^{-1}\Xi_A^0\big]^{-1}(\Xi_A^0)^TQ_y^{-1}\big\}\big)\cdot\\
&\cdot(\hat{\boldsymbol{\xi}}_A-\boldsymbol{\xi}_A^0)=\underset{n\times 1}{\mathbf{0}}
\end{aligned}
\tag{11d}
$$

contains $n-m$ *independent* constraints and can be written more succinctly as

$$
\boxed{K_y(\hat{\boldsymbol{\mu}}_y-\Xi_A^0\boldsymbol{\xi}_0)-K_A(\hat{\boldsymbol{\xi}}_A-\boldsymbol{\xi}_A^0)=\mathbf{0},} \tag{11e}
$$

with obvious definitions for matrices K_y of size $(n-m)\times n$ and K_A of size $(n-m)\times nm$. Moreover, by subtracting $\Xi_A^0\boldsymbol{\xi}_0$ and $\operatorname{vec}\Xi_A^0$ from the first and second lines of (10a), respectively, we arrive at

$$
\boldsymbol{y}-\Xi_A^0\boldsymbol{\xi}_0=(\boldsymbol{\mu}_y-\Xi_A^0\boldsymbol{\xi}_0)+\boldsymbol{e}_y \tag{11f}
$$

and

$$
\operatorname{vec}(A-\Xi_A^0)=(\boldsymbol{\xi}_A-\boldsymbol{\xi}_A^0)+\boldsymbol{e}_A. \tag{11g}
$$

The *unconstrained estimates* and their associated dispersion matrix are then given by

$$
\begin{bmatrix}\hat{\boldsymbol{\mu}}_y-\Xi_A^0\boldsymbol{\xi}_0\\ \hat{\boldsymbol{\xi}}_A-\boldsymbol{\xi}_A^0\end{bmatrix}_{\mathsf{u}}=\begin{bmatrix}\boldsymbol{y}-\Xi_A^0\boldsymbol{\xi}_0\\ \operatorname{vec}(A-\Xi_A^0)\end{bmatrix}, \tag{12a}
$$

$$
D\{\begin{bmatrix}\hat{\boldsymbol{\mu}}_y-\Xi_A^0\boldsymbol{\xi}_0\\ \hat{\boldsymbol{\xi}}_A-\boldsymbol{\xi}_A^0\end{bmatrix}_{\mathsf{u}}\}=\sigma_0^2\begin{bmatrix}Q_y & 0\\ 0 & Q_A\end{bmatrix}, \tag{12b}
$$

respectively. Here, the subscript u stands for unconstrained, and below the subscript c stands for constrained. The differences between the *constrained* and *unconstrained* estimates are given by

$$
\begin{aligned}
&\begin{bmatrix}\hat{\boldsymbol{\mu}}_y-\Xi_A^0\boldsymbol{\xi}_0\\ \hat{\boldsymbol{\xi}}_A-\boldsymbol{\xi}_A^0\end{bmatrix}_{\mathsf{c}}-\begin{bmatrix}\hat{\boldsymbol{\mu}}_y-\Xi_A^0\boldsymbol{\xi}_0\\ \hat{\boldsymbol{\xi}}_A-\boldsymbol{\xi}_A^0\end{bmatrix}_{\mathsf{u}}=\\
&=\begin{bmatrix}(\hat{\boldsymbol{\mu}}_y)_{\mathsf{c}}-(\hat{\boldsymbol{\mu}}_y)_{\mathsf{u}}\\ (\hat{\boldsymbol{\xi}}_A)_{\mathsf{c}}-(\hat{\boldsymbol{\xi}}_A)_{\mathsf{u}}\end{bmatrix}=\begin{bmatrix}(\hat{\boldsymbol{\mu}}_y)_{\mathsf{c}}-\boldsymbol{y}\\ (\hat{\boldsymbol{\xi}}_A)_{\mathsf{c}}-\operatorname{vec}A\end{bmatrix}=\\
&=\begin{bmatrix}Q_yK_y^T\\ -Q_AK_A^T\end{bmatrix}(K_yQ_yK_y^T+K_AQ_AK_A^T)^{-1}\cdot\\
&\cdot\big\{\mathbf{0}-K_y(\boldsymbol{y}-\Xi_A^0\cdot\boldsymbol{\xi}_0)+K_A\big[\operatorname{vec}(A-\Xi_A^0)\big]\big\},
\end{aligned}
\tag{13a}
$$

with the associated dispersion matrix

$$
\begin{aligned}
D\{\begin{bmatrix}(\hat{\boldsymbol{\mu}}_y)_{\mathsf{c}}-\boldsymbol{y}\\ (\hat{\boldsymbol{\xi}}_A)_{\mathsf{c}}-\operatorname{vec}A\end{bmatrix}\}&=\sigma_0^2\begin{bmatrix}Q_yK_y^T\\ -Q_AK_A^T\end{bmatrix}\cdot\\
&\cdot(K_yQ_yK_y^T+K_AQ_AK_A^T)^{-1}\cdot\\
&\cdot\big[K_yQ_y\ \big|\ -K_AQ_A\big].
\end{aligned}
\tag{13b}
$$

Now we proceed with the constrained solution, but drop the subscript c, in the following. Equation (13a)

implies

$$\begin{bmatrix}\hat{\boldsymbol{\mu}}_y\\ \hat{\boldsymbol{\xi}}_A\end{bmatrix} = -\begin{bmatrix}Q_yK_y^T\\ -Q_AK_A^T\end{bmatrix}(K_yQ_yK_y^T+K_AQ_AK_A^T)^{-1}\cdot \cdot\left[K_y \mid -K_A\right]\cdot\begin{bmatrix}\boldsymbol{y}-\Xi_A^0\cdot\boldsymbol{\xi}_0\\ \text{vec}(A-\Xi_A^0)\end{bmatrix}+\begin{bmatrix}\boldsymbol{y}\\ \text{vec}\,A\end{bmatrix}= = \left(I_{n(m+1)}-S\right)\begin{bmatrix}\boldsymbol{y}\\ \text{vec}\,A\end{bmatrix}+S\begin{bmatrix}\Xi_A^0\cdot\boldsymbol{\xi}_0\\ \text{vec}\,\Xi_A^0\end{bmatrix}, \tag{14a}$$

with the $n(m+1)\times n(m+1)$ matrix S defined as

$$S := \left[\begin{array}{c|c} Q_yK_y^T(K_yQ_yK_y^T+K_AQ_AK_A^T)^{-1}K_y & \\ -Q_AK_A^T(K_yQ_yK_y^T+K_AQ_AK_A^T)^{-1}K_y & \\ \hline & -Q_yK_y^T(K_yQ_yK_y^T+K_AQ_AK_A^T)^{-1}K_A\\ & Q_AK_A^T(K_yQ_yK_y^T+K_AQ_AK_A^T)^{-1}K_A\end{array}\right]. \tag{14b}$$

Now, substituting (14a) into (11c) allows the estimate $\hat{\boldsymbol{\xi}}$ to be written as an update to the (given) approximate vector $\boldsymbol{\xi}_0$ as

$$\hat{\boldsymbol{\xi}} = \boldsymbol{\xi}_0 + \left[(\Xi_A^0)^TQ_y^{-1}\Xi_A^0\right]^{-1}\cdot \cdot\left[(\Xi_A^0)^TQ_y^{-1} \mid -(\boldsymbol{\xi}_0\otimes Q_y^{-1}\Xi_A^0)^T\right]\cdot \cdot\left(\begin{bmatrix}I_n & 0\\ 0 & I_{nm}\end{bmatrix}-\begin{bmatrix}S_{11} & S_{12}\\ S_{21} & S_{22}\end{bmatrix}\right)\begin{bmatrix}\boldsymbol{y}-\Xi_A^0\cdot\boldsymbol{\xi}_0\\ \text{vec}\,A-\text{vec}\,\Xi_A^0\end{bmatrix}, \tag{15a}$$

with the obvious partitioning of matrix S into four submatrices. Clearly, the vector $\boldsymbol{\xi}_0$ and the matrix Ξ_A^0 are nonrandom, since they contain (given) approximate values, e.g., expansion points of a Taylor series. Thus, the only random terms in (15a) are the vector $\boldsymbol{y}$ and the matrix A, with respective dispersion matrices $D\{\boldsymbol{y}\}=\sigma_0^2Q_y$ and $D\{\text{vec}\,A\}=\sigma_0^2Q_A$. Hence, the dispersion of $\hat{\boldsymbol{\xi}}$ can be approximated by the law of error propagation, resulting in

$$D\{\hat{\boldsymbol{\xi}}\} = = \sigma_0^2\cdot\left[(\Xi_A^0)^TQ_y^{-1}\Xi_A^0\right]^{-1}\begin{bmatrix}Q_y^{-1}(\Xi_A^0)\\ -(\boldsymbol{\xi}_0\otimes Q_y^{-1}\Xi_A^0)\end{bmatrix}^T\cdot \cdot\left(I_{n(m+1)}-S\right)\begin{bmatrix}Q_y & 0\\ 0 & Q_A\end{bmatrix}\left(I_{n(m+1)}-S\right)^T\cdot \cdot\begin{bmatrix}Q_y^{-1}(\Xi_A^0)\\ -(\boldsymbol{\xi}_0\otimes Q_y^{-1}\Xi_A^0)\end{bmatrix}\left[(\Xi_A^0)^TQ_y^{-1}\Xi_A^0\right]^{-1}. \tag{15b}$$

Finally, the respective predicted residual vectors are computed by

$$\tilde{\boldsymbol{e}}_y = \boldsymbol{y}-\hat{\boldsymbol{\mu}}_y \;\text{ and }\; \tilde{\boldsymbol{e}}_A = \text{vec}\,A-\hat{\boldsymbol{\xi}}_A. \tag{16}$$

2.4. Estimates within the Errors-In-Variables (EIV) model

The EIV model can be written as

$$\boldsymbol{y}-\boldsymbol{e}_y = (A-E_A)\boldsymbol{\xi}, \tag{17}$$

with (1b) as its stochastic component.

However, if (1b) is specialized so that $Q_A = Q_0\otimes Q_y$ (implying $Q_x=Q_y$), and Q_0 and Q_y are diagonal matrices, then the application of (total) least-squares minimization leads to the nonlinear system of normal equations

$$(N-\hat{\nu}Q_0)\hat{\boldsymbol{\xi}} = \boldsymbol{c}, \tag{18a}$$

with

$$\left[N \mid \boldsymbol{c}\right] := (\Xi_A^0)^TQ_y^{-1}\left[\Xi_A^0 \mid \boldsymbol{y}\right] \tag{18b}$$

and

$$\hat{\nu} = (\boldsymbol{y}-\Xi_A^0\boldsymbol{\xi}_0)^T(\boldsymbol{y}-\Xi_A^0\boldsymbol{\xi}_0)/(1+\boldsymbol{\xi}_0^TQ_0\boldsymbol{\xi}_0), \tag{18c}$$

where $\hat{\nu}$ turns out to be the *Total Sum of Squared Residuals* (TSSR) (see Schaffrin and Wieser (2008) and Schaffrin (2015)). Moreover, in the case where $Q_0=I_m$, the TSSR $\hat{\nu}$ is also the smallest eigenvalue of the matrix $\left[\begin{smallmatrix}N & \boldsymbol{c}\\ \boldsymbol{c}^T & \boldsymbol{y}^TQ_y^{-1}\boldsymbol{y}\end{smallmatrix}\right]$ in agreement with Golub and van Loan (1979).

The nonlinear system of equations (18a) can be linearized via a truncated Taylor series expansion, resulting in the "total differential"

$$N\cdot d\hat{\boldsymbol{\xi}} - Q_0d\hat{\boldsymbol{\xi}}\cdot\nu_0 - Q_0\boldsymbol{\xi}_0\cdot d\hat{\nu} = (\Xi_A^0)^TQ_y^{-1}d\boldsymbol{y}, \tag{19a}$$

or

$$d\hat{\boldsymbol{\xi}} = (N-\nu_0Q_0)^{-1}\big(Q_0\boldsymbol{\xi}_0\cdot d\hat{\nu}+(\Xi_A^0)^TQ_y^{-1}d\boldsymbol{y}\big), \tag{19b}$$

where the expansion points $\boldsymbol{\xi}_0$, Ξ_A^0, and ν_0 are known approximately. Now, (19b) is linear in the random variables $\boldsymbol{y}$ and $d\hat{\nu}$ (also $D\{d\hat{\nu}\}=D\{\hat{\nu}\}$ etc.). Therefore the *law of variance (error) propagation* may be applied, leading to the $m\times m$ dispersion matrix

$$D\{\hat{\boldsymbol{\xi}}\} = (N-\nu_0Q_0)^{-1}\cdot \cdot(\sigma_0^2N+Q_0\boldsymbol{\xi}_0\cdot D\{\hat{\nu}\}\cdot\boldsymbol{\xi}_0^TQ_0)(N-\nu_0Q_0)^{-1}. \tag{19c}$$

Obviously, an expression for $D\{\hat{\nu}\}$ is required for (19c).

Since $\hat{\nu}$ provides the TSSR, it follows that

$$\hat{\sigma}_0^2 = \hat{\nu}/(n-m) \tag{20a}$$

provides an estimate of the unknown variance component σ_0^2 appearing in (1b). It is well known that the estimate can also be derived from the principle of BIQUUE (Best Invariant Quadratic Uniformly Unbiased Estimate), and so can its dispersion, which is provided by

$$D\{\hat{\sigma}_0^2\} = 2(\sigma_0^2)^2/(n-m), \tag{20b}$$

implying that

$$D\{\hat{\nu}\} = (n-m)^2\cdot D\{\hat{\sigma}_0^2\} = 2(\sigma_0^2)^2(n-m) \tag{20c}$$

in this special case of diagonal cofactor matrices Q_0 and Q_y. Thus, we may rewrite (19c) as

$$D\{\hat{\boldsymbol{\xi}}\} = \sigma_0^2 \cdot (N - \nu_0 Q_0)^{-1} \cdot \\ \cdot (N + Q_0\boldsymbol{\xi}_0 \cdot 2\sigma_0^2(n-m) \cdot \boldsymbol{\xi}_0^T Q_0)(N - \nu_0 Q_0)^{-1}. \quad (20d)$$

Of course, (20d) cannot actually be computed due to the unknown term σ_0^2, but the usual remedy is to replace the unknown quantity by its estimate. Moreover, in any iteration scheme the value assigned to the quantity ν_0 may be taken from $\hat{\nu}_0$ upon convergence (and likewise $\hat{\boldsymbol{\xi}}$ for $\boldsymbol{\xi}_0$), thus justifying the following formula for practical problems:

$$\hat{D}\{\hat{\boldsymbol{\xi}}\} = (n-m)^{-1}\hat{\nu} \cdot (N - \hat{\nu}Q_0)^{-1} \cdot \\ \cdot (N + Q_0\hat{\boldsymbol{\xi}} \cdot 2\hat{\nu} \cdot \hat{\boldsymbol{\xi}}^T Q_0)(N - \hat{\nu}Q_0)^{-1}. \quad (21a)$$

or

$$\hat{D}\{\hat{\boldsymbol{\xi}}\} = (n-m)^{-1} \cdot (\hat{\nu}^{-1}N - Q_0)^{-1} \cdot \\ \cdot (\hat{\nu}^{-1}N + 2 \cdot Q_0\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}^T Q_0)(\hat{\nu}^{-1}N - Q_0)^{-1}. \quad (21b)$$

Obviously, if $Q_0 = I_m$, then (21a) reduces to

$$\hat{D}\{\hat{\boldsymbol{\xi}}\} = (n-m)^{-1}\hat{\nu} \cdot (N - \hat{\nu}I_m)^{-1} \cdot \\ \cdot (N + \hat{\boldsymbol{\xi}} \cdot 2\hat{\nu} \cdot \hat{\boldsymbol{\xi}}^T)(N - \hat{\nu}I_m)^{-1}. \quad (21c)$$

Finally, the predicted residual vector and matrix are computed by

$$\tilde{\boldsymbol{e}}_y = (1 + \hat{\boldsymbol{\xi}}^T Q_0 \hat{\boldsymbol{\xi}})^{-1} \cdot (\boldsymbol{y} - A\hat{\boldsymbol{\xi}}) \quad (22a)$$

and

$$\tilde{E}_A = -(1 + \hat{\boldsymbol{\xi}}^T Q_0 \hat{\boldsymbol{\xi}})^{-1} \cdot (\boldsymbol{y} - A\hat{\boldsymbol{\xi}})\hat{\boldsymbol{\xi}}^T Q_0, \quad (22b)$$

respectively, in the specialized case where $Q_x = Q_y$.

3. Numerical example

We have chosen the relative simple problem of a 1-D similarity transformation (or Helmert transformation) for our numerical example, as it serves primarily to illustrate the formulas presented in the preceding sections. Solving this transformation problem is tantamount to fitting measured coordinates to a 2-D line, which might be done in a variety of deformation-monitoring applications. It can also be used to estimate offset and scale parameters between two sets of heights estimated from different survey campaigns, a case of particular interest in monitoring the vertical stability of a man-made structure or characterizing surface subsidence due to subsurface water or mineral extraction.

Below are results from 2D line fitting. One data set is from Pearson (1901), being coordinates given without units. The other is from Haneberg (2004), where the x-variables are rain fall in mm and the y-variables are change in water level in cm. In both cases the data are considered iid with $Q_0 = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right]$, and $Q_y = Q_x = I_m$. The estimated slope is denoted by $\hat{\xi}_1$, while $\hat{\xi}_2$ refers to the estimated y-intercept.

The least-squares solutions within all four models described above yield precisely the same parameter estimates, sum of squared residuals (SSR), and residual vectors, which are listed in Tables 2 and 3. They are shown there at a greater precision than what is warranted, which would allow others to more precisely compare the output of their own algorithms.

Table 1: Two data sets for 2D line fitting

	Pearson		Haneberg	
No.	$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{x}$ mm	$\boldsymbol{y}$ cm
1	0	5.9	1.94	2.50
2	0.9	5.4	3.33	1.89
3	1.8	4.4	3.22	1.67
4	2.6	4.6	5.67	1.31
5	3.3	3.5	4.72	1.02
6	4.4	3.7	3.89	0.96
7	5.2	2.8	2.78	1.10
8	6.1	2.8	10.56	0.15
9	6.5	2.4	9.44	3.92
10	7.4	1.5	12.78	5.23
11			14.72	4.22
12			13.61	3.63
13			20.39	4.32
14			38.89	5.89

Table 2: Parameter estimates for 1D similarity transformation problems. The Pearson estimates are unitless. The units for Haneberg are cm/mm for slope $\hat{\xi}_1$ and cm for y-intercept $\hat{\xi}_1$.

Data set	n	TSSR	$\hat{\xi}_1$	$\hat{\xi}_2$
Pearson	10	0.6186	-0.5456	5.7840
Haneberg	14	17.9659	0.1396	1.2455

Table 3: 2D-line residuals computed for two data sets

Pearson		Haneberg	
$\tilde{\boldsymbol{e}}_x$	$\tilde{\boldsymbol{e}}_y$	$\tilde{\boldsymbol{e}}_x$ mm	$\tilde{\boldsymbol{e}}_y$ cm
-0.048751	0.089360	-0.134692	0.964862
-0.044969	0.082428	-0.024596	0.176191
0.169026	-0.309820	0.003426	-0.024541
-0.098554	0.180648	0.099551	-0.713135
0.203357	-0.372748	0.121101	-0.867509
-0.133034	0.243847	0.113452	-0.812712
0.061856	-0.113380	0.073065	-0.523398
-0.144576	0.265004	0.351859	-2.520541
-0.068153	0.124922	-0.185769	1.330756
0.103800	-0.190262	-0.301302	2.158375
		-0.125922	0.902040
		-0.066351	0.475308
		-0.031234	0.223745
		0.107411	-0.769440

In contrast, the dispersion matrix within the EIV model, based on (21a), deviated somewhat from the dispersion matrices computed within the other three models. The first pair of *empirical standard deviations* ($\hat{\sigma}_1$ and $\hat{\sigma}_2$) for each data set (labeled EIV in Table 4) were computed from (21a). The second pair were computed by the corresponding formulas within the

other three models, which were in precise agreement with each other. In all cases, the estimated variance component $\hat{\sigma}_0^2$ was substituted for the unknown, true one, hence the term "empirical standard deviation." Table 4 reveals that, for these data sets, the empirical standard deviations based on (21a) are less than or equal to those based on the formulas associated with the other three models, i.e., equations (6b), (9e), and (15b).

Table 4: Empirical standard deviations of estimated parameters in the 1D similarity transformation problems

Equations/	(21a) EIV		(6b), (9e), (15b)	
Data set	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$
Pearson	0.0376	0.1684	0.0422	0.1899
Haneberg	0.0348	0.4882	0.0346	0.4890

It is also interesting to compare the values in Table 4 with those computed from equation (23) in Schaffrin, Lee, et al. (2006), which reads

$$\hat{D}\{\hat{\boldsymbol{\xi}}\} \approx \hat{\sigma}_0^2 \cdot (N - \hat{\nu} I_m)^{-1} N (N - \hat{\nu} I_m)^{-1}.$$

These turn out to be $\hat{\sigma}_1 = 0.0375$ and $\hat{\sigma}_2 = 0.1680$ for the Pearson data and $\hat{\sigma}_1 = 0.0348$ and $\hat{\sigma}_2 = 0.4881$ for the Haneberg data. It is not surprising that they are slightly smaller than those computed from (21a) of this paper, since the middle term in that formula includes $\hat{\boldsymbol{\xi}} \cdot 2\hat{\nu} \cdot \hat{\boldsymbol{\xi}}^T$, which does not appear in the equation above.

4. Conclusions and outlook

In this contribution, we have shown how to generate equivalent least-squares solutions for estimated parameters and predicted residuals among four different data models. We have also shown that the dispersion (variance-covariance) matrix for the parameter estimates in the EIV model will vary slightly from those computed within the other three models if our proposed formula (21a) is used. We argue that (21a) represents a step forward in deriving an accurate formula for the dispersion matrix within the EIV model; however, more work is needed to bring the derivation to full maturity as noted below.

Equation (21a) was derived for an EIV model with stochastic properties $Q_A = Q_0 \otimes Q_y$, implying $Q_x = Q_y$. While this made the derivation tractable, it is a somewhat restrictive requirement. In future work, we hope to extend the formula in two steps. First, we want to remove the restriction that $Q_x = Q_y$. Second, we want to allow Q_A to be any symmetric, positive-(semi)definite cofactor matrix, thereby allowing for a much wider range of problems to be handled, including those that give rise to a structured design matrix A. In such models the same variable appears more than once in the design matrix. One example is a 2D similarity transformation commonly used in geodetic and deformation-monitoring problems.

References

Amiri-Simkooei, A. and S. Jazaeri (2012). "Weighted Total Least-Squares formulated by standard Least-Squares theory". In: *Journal of Geodetic Science* 2(2), pp. 113–124.

Golub, G. and C. van Loan (1979). "Total least squares". In: *Smoothing Techniques for Curve Estimation*. Ed. by T. Gasser and M. Rosenblatt. Vol. 757. Lecture Notes in Mathematics. Springer Berlin/Heidelberg, pp. 69–76.

Haneberg, W. (2004). *Computational Geosciences with Mathematica*. Springer Berlin Heidelberg.

Harville, D. A. (1986). "Using Ordinary Least Squares Software to Compute Combined Intra-Interblock Estimates of Treatment Contrasts." In: *The American Statistician* 40(2), pp. 153–157.

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

Schaffrin, B. and Y. Felus (2005). "On Total Least-Squares Adjustment with Constraints". In: *A Window on the Future of Geodesy*. Ed. by F. Sansò. IAG-Symp., Vol. 128, Springer: Berlin/Heidelberg/New York, pp. 417–421.

Schaffrin, B. (2015). "Adjusting the Errors-In-Variables Model: Linearized Least-Squares vs. Nonlinear Total Least-Squares". In: *VIII Hotine-Marussi Symposium on Mathematical Geodesy*. Ed. by N. Sneeuw, P. Novák, M. Crespi, and F. Sansò. Cham: Springer International Publishing, pp. 301–307.

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

Schaffrin, B. and K. Snow (2014). "The case of the Homogenous Errors-In-Variables Model". In: *J. of Geodetic Sci.* 4, pp. 166–173.

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