NOAA Technical Memorandum NOS NGS 74

On Least-Squares Adjustments within the Variance Component Model with Stochastic Constraints

Versions

Date	Changes
May 1, 2018	Original Release

NOAA Technical Memorandum
NOS NGS 74

On Least-Squares Adjustments Within the Variance Component Model with Stochastic Constraints

Dru Smith, Jacob Heck, Dan Gillins, Kyle Snow

1 Background

The estimation of unknown, but fixed, parameters from observations requires that an underlying mathematical model exists relating the parameters, the observations, and the nature of their random errors. When least-squares techniques are applied, the estimation process is called *Least-Squares Adjustment*.

In this memorandum, a new observational model called the *Variance Component Model (VCM) with Stochastic Constraints*[1] and a *LEast-Squares Solution (LESS)* within that model are introduced. This new model is needed to address the specific situation when prior information exists for the parameters, but it is presumed that the respective dispersion[2](covariance) matrices for the prior information and the observations do not share a common variance component. This implies that, while the weights among the observations and also those among the prior information are known accurately, there still remains an unknown scale difference between them.

The VCM with Stochastic Constraints is based on two other observational models, namely, the *Variance Component Model (VCM)* (Schaffrin and Snow (2017b)) and the *Gauss-Markov Model (GMM) with Stochastic Constraints* (Schaffrin and Snow (2017a).) The traditional VCM allows for multiple variance components, but it does not include prior information on the parameters. On the other hand, the Gauss-Markov Model with Stochastic Constraints allows for both observations and prior information on the parameters, but it has only one variance component.

[1]It is acknowledged that the model has also been presented in Schaffrin et al. (2018) and that we have openly shared ideas and numerical results with those authors.

[2]The terms dispersion matrix and covariance matrix are used synonymously throughout this document.

The motivation for developing the new model was driven by the pending release of the North American-Pacific Geopotential Datum of 2022 (NAPGD2022; NGS, 2017). It is expected that this release will give rise to combined GNSS and spirit leveling surveys, where surveyors would first utilize GNSS data to determine GNSS-derived orthometric heights. These heights would then be used as prior information on the parameters in a least-squares adjustment of geodetic leveling data. However, the new model is derived generically, and thus it can be applied to any sort of geodetic data where observations and prior information are provided and two (or more) variance components should be estimated.

An alternative approach for estimating height parameters in the scenario described above is the *Partial MInimum NOrm LEast-Squares Solution (Partial MINOLESS)*, which might be more suitable in cases where variances for the prior information are not available. In any case, it could be a useful check on the LESS within the new model, as one would not expect that parameters estimated from the two different approaches would vary significantly.

In the following, we review the GMM with Stochastic Constraints in Section 2, followed by a review of the traditional VCM in Section 3. In Section 4, we introduce the new model and then show the least-squares solution (LESS) within this model in Section 6, which follows the presentation of the LESS within the traditional VCM in Section 5. The Partial MINOLESS is introduced in Section 7. In Section 8, we present numerical results from a leveling network where prior information on the parameters was provided via a GNSS survey. Finally, in Section 9 we give a summary of our findings and make some recommendations for use of the presented models and least-squares adjustments.

2 The Gauss-Markov Model with Stochastic Constraints

Following Schaffrin and Snow (2017a), the (linearized) Gauss-Markov Model (GMM) with Stochastic Constraints can be written as

$$\boldsymbol{y} = A\boldsymbol{\xi} + \boldsymbol{e}, \tag{1}$$

$$\boldsymbol{z}_0 = K\boldsymbol{\xi} + \boldsymbol{e}_0, \tag{2}$$

$$\begin{bmatrix} \boldsymbol{e} \\ \boldsymbol{e}_0 \end{bmatrix} \sim \left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \sigma_0^2 \begin{bmatrix} P^{-1} & 0 \\ 0 & P_0^{-1} \end{bmatrix} \right). \tag{3}$$

The assumption that the GMM with Stochastic Constraints might be linearized is made to handle cases where the observations are a non-linear function of the unknown parameters. In those cases, it would be technically correct to speak of *incremental observations* and *incremental parameters*; however, for the sake of simplicity, the adjective "incremental" will not be used in this document. Each of the variables shown in (1)–(3) is described briefly in Table 1.

Assuming that $\text{rk}[A^T \mid K^T] = m$, the LESS within the GMM with Stochastic

Table 1: Variables used in the Gauss-Markov Model with Stochastic Constraints

Variable	Size	Description
$\boldsymbol{y}$	$n \times 1$	Vector of random observations
A	$n \times m$	Coefficient matrix relating observations to parameters
$\boldsymbol{\xi}$	$m \times 1$	Vector of unknown fixed parameters to be estimated
$\boldsymbol{e}$	$n \times 1$	Vector of unknown random errors in the observations
$\boldsymbol{z}_0$	$l \times 1$	Vector of prior information on the parameters
K	$l \times m$	Coefficient matrix for the prior information
$\boldsymbol{e}_0$	$l \times 1$	Vector of unknown random errors in the prior information
σ_0^2		Unknown variance component (scalar quantity)
P	$n \times n$	Invertible observational weight matrix
P_0	$l \times l$	Invertible weight matrix for the prior information
m		Number of unknown parameters to be estimated
n		Number of observations
l		Length of prior information vector and number of stochastic constraints

Constraints exists uniquely and is arrived at in a non-iterative fashion.[3] The variance component σ_0^2 is estimated (yielding $\hat{\sigma}_0^2$) as a byproduct of the LESS for the parameters (see (7)), which can then be used in the estimated dispersion (covariance) matrix for the estimated parameters $\hat{D}\{\hat{\boldsymbol{\xi}}\}$, as in (8).

The unknown variance component σ_0^2 in (3) (sometimes called variance of unit weight) implies that the dispersion matrices for the random error vectors $\boldsymbol{e}$ and $\boldsymbol{e}_0$ are only known up to a scale factor; i.e., $P = \sigma_0^2 \cdot [D\{\boldsymbol{e}\}]^{-1}$ and $P_0 = \sigma_0^2 \cdot [D\{\boldsymbol{e}_0\}]^{-1}$. Obviously, the variance component is common to both dispersion matrices, as reflected in (3).

Some of the more useful equations derived from the LESS are listed in (4)–(10) (*ibid*):

The estimated parameters:

$$\hat{\boldsymbol{\xi}} = (A^T P A + K^T P_0 K)^{-1} (A^T P \boldsymbol{y} + K^T P_0 \boldsymbol{z}_0). \tag{4}$$

[3]That is, no iteration is required to arrive at equations (4)–(10). However, if the Gauss-Markov Model with Stochastic Constraints was linearized, then equations (4)–(10) pertain to incremental parameters, and the solution may need to be iterated to arrive at estimates of the total parameters.

The adjusted observations:

$$\begin{bmatrix} \tilde{\boldsymbol{y}} \\ \tilde{\boldsymbol{z}}_0 \end{bmatrix} = \begin{bmatrix} A \\ K \end{bmatrix} \hat{\boldsymbol{\xi}}. \tag{5}$$

The predicted random errors (residuals):

$$\begin{bmatrix} \tilde{\boldsymbol{e}} \\ \tilde{\boldsymbol{e}}_0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{z}_0 \end{bmatrix} - \begin{bmatrix} \tilde{\boldsymbol{y}} \\ \tilde{\boldsymbol{z}}_0 \end{bmatrix}. \tag{6}$$

The estimated variance component:

$$\hat{\sigma}_0^2 = \frac{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} + \tilde{\boldsymbol{e}}_0^T P_0 \tilde{\boldsymbol{e}}_0}{n - m + l}. \tag{7}$$

Equation (4) provides the estimated parameters, (5) the adjusted observations, (6) the predicted random errors, or residuals, and (7) provides an estimate of the variance component. Note that the denominator in (7), $n-m+l$, denotes the redundancy of model (1)–(3). The estimated dispersion matrices for the estimated parameters, adjusted observations, and residuals are shown in (8)–(10), respectively. The dispersion (variance) of $\hat{\sigma}_0^2$ can also be estimated, under the assumption of normality, via $\hat{D}\{\hat{\sigma}_0^2\} = 2(\hat{\sigma}_0^2)^2[n - m + l]^{-1}$.

The estimated dispersion matrix for the estimated parameters:

$$\hat{D}\{\hat{\boldsymbol{\xi}}\} = \hat{\sigma}_0^2 (A^T P A + K^T P_0 K)^{-1}. \tag{8}$$

The estimated dispersion matrix for the adjusted observations:

$$\begin{aligned} \hat{D}\{\begin{bmatrix} \tilde{\boldsymbol{y}} \\ \tilde{\boldsymbol{z}}_0 \end{bmatrix}\} = \hat{D}\{\begin{bmatrix} A \\ K \end{bmatrix} \hat{\boldsymbol{\xi}}\} = \begin{bmatrix} A \\ K \end{bmatrix} \hat{D}\{\hat{\boldsymbol{\xi}}\} \begin{bmatrix} A \\ K \end{bmatrix}^T = \\ = \hat{\sigma}_0^2 \begin{bmatrix} A(A^T PA + K^T P_0 K)^{-1} A^T & A(A^T PA + K^T P_0 K)^{-1} K^T \\ K(A^T PA + K^T P_0 K)^{-1} A^T & K(A^T PA + K^T P_0 K)^{-1} K^T \end{bmatrix}. \end{aligned} \tag{9}$$

The estimated dispersion matrix for the residuals:

$$\begin{aligned} \hat{D}\{\begin{bmatrix} \tilde{\boldsymbol{e}} \\ \tilde{\boldsymbol{e}}_0 \end{bmatrix}\} = \hat{D}\{\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{z}_0 \end{bmatrix}\} - \hat{D}\{\begin{bmatrix} \tilde{\boldsymbol{y}} \\ \tilde{\boldsymbol{z}}_0 \end{bmatrix}\} = \hat{\sigma}_0^2 \begin{bmatrix} P^{-1} & 0 \\ 0 & P_0^{-1} \end{bmatrix} - \hat{D}\{\begin{bmatrix} \tilde{\boldsymbol{y}} \\ \tilde{\boldsymbol{z}}_0 \end{bmatrix}\} = \\ = \hat{\sigma}_0^2 \begin{bmatrix} P^{-1} - A(A^T PA + K^T P_0 K)^{-1} A^T & -A(A^T PA + K^T P_0 K)^{-1} K^T \\ -K(A^T PA + K^T P_0 K)^{-1} A^T & P_0^{-1} - K(A^T PA + K^T P_0 K)^{-1} K^T \end{bmatrix}. \end{aligned} \tag{10}$$

It is expected that cases will occur in which the observational weight matrix P and the prior information weight matrix P_0 differ from their true values by unique and *independent* scale factors, thereby invalidating the use of model (1)–(3). One model which allows for the estimation of two (or more) variance components, and therefore may be informative for such cases, is the Variance Component Model (VCM). Note, however, that the traditional VCM does not incorporate prior information on the parameters. Nevertheless, we present it in the next section as a preliminary step before introducing our new model.

3 The Variance Component Model

The (linearized) Variance Component Model (VCM) can be written as follows:[4]

$$\boldsymbol{y}' = A'\boldsymbol{\xi} + \boldsymbol{e}', \tag{11}$$

$$\boldsymbol{e}' \sim (\boldsymbol{0}, \sigma_1^2 Q_1 + \sigma_2^2 Q_2). \tag{12}$$

The terms used in (11) and (12) are described briefly in Table 2. Note that the matrices Q_1 and Q_2 are formally called cofactor matrices. They are known matrices, and when they are invertible (or when their non-zero submatrices are) the inverses are called weight matrices. The use of two variance components, σ_1^2 and σ_2^2, implies that the two dispersion matrices are only known up to unique and independent scale factors. Here we make a clarifying comment that, formally, it is the variance component times the cofactor matrix that is called the dispersion (covariance) matrix. For example, if $\Sigma_1 = \sigma_1^2 Q_1$, then Σ_1 is a dispersion matrix in this context.

Table 2: Variables used in the Variance Component Model

Variable	Size	Description
$\boldsymbol{y}'$	$n' \times 1$	Vector of random observations
A'	$n' \times m$	Coefficient matrix relating observations to parameters
$\boldsymbol{\xi}$	$m \times 1$	Vector of unknown fixed parameters to be estimated
$\boldsymbol{e}'$	$n' \times 1$	Vector of unknown random errors in the observations
σ_1^2		The first unknown variance component
σ_2^2		The second unknown variance component
Q_1	$n' \times n'$	The first observational cofactor matrix
Q_2	$n' \times n'$	The second observational cofactor matrix
n'		Number of observations
m		Number of parameters to be estimated

Now, it is expected that cases will occur in which the need for two variance components exists, but where one is associated with *observational* random errors and the other with random errors in the prior information, rather than both applying to the observations as in (11) and (12). It is this case that motivated our development of the new model described in the next section.

[4] The prime mark used here is non-traditional, but it is helpful to distinguish these variables from those in the GMM with Stochastic Constraints.

4 The Variance Component Model with Stochastic Constraints

The following linearized[5] Variance Component Model (VCM) with Stochastic Constraints looks very similar to the Gauss-Markov Model (GMM) with Stochastic Constraints (compare to (1)–(3)):

$$\boldsymbol{y} = A\boldsymbol{\xi} + \boldsymbol{e}, \tag{13}$$

$$\boldsymbol{z}_0 = K\boldsymbol{\xi} + \boldsymbol{e}_0, \tag{14}$$

$$\begin{bmatrix} \boldsymbol{e} \\ \boldsymbol{e}_0 \end{bmatrix} \sim \left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \sigma_1^2 P^{-1} & 0 \\ 0 & \sigma_2^2 P_0^{-1} \end{bmatrix} \right). \tag{15}$$

Note that the *only* change is the introduction of two unique variance components in (15), rather than one in (3). All other terms are identical to those in the GMM with Stochastic Constraints in (1)–(3).

The introduction of two variance components means that the formulas for the LESS within the GMM with Stochastic Constraints cannot be used within the VCM with Stochastic Constraints to estimate the unknown parameters $\boldsymbol{\xi}$. However, if (13)–(15) could be related to (11) and (12), then the iterative approach to the LESS within the VCM could be applied. This is the approach we take here, rather than derive the LESS from first principles. Specifically, note that the stochastic-constraint equations (14) have the same mathematical form as the observation equations (13). Thus, (13) and (14) may be combined into a single set of equations, which can be written as

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{z}_0 \end{bmatrix} = \begin{bmatrix} A \\ K \end{bmatrix} \boldsymbol{\xi} + \begin{bmatrix} \boldsymbol{e} \\ \boldsymbol{e}_0 \end{bmatrix}. \tag{16}$$

Furthermore, using appropriately sized blocks of zeros, (15) can be expressed equivalently as

$$\begin{bmatrix} \boldsymbol{e} \\ \boldsymbol{e}_0 \end{bmatrix} \sim \left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \sigma_1^2 \begin{bmatrix} P^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \sigma_2^2 \begin{bmatrix} 0 & 0 \\ 0 & P_0^{-1} \end{bmatrix} \right). \tag{17}$$

Now, (16) and (17) have the same form as (11) and (12), with the following useful equivalences:

$$\boldsymbol{e}' := \begin{bmatrix} \boldsymbol{e} \\ \boldsymbol{e}_0 \end{bmatrix}, \tag{18}$$

$$Q_1 := \begin{bmatrix} P^{-1} & 0 \\ 0 & 0 \end{bmatrix}, \tag{19}$$

$$Q_2 := \begin{bmatrix} 0 & 0 \\ 0 & P_0^{-1} \end{bmatrix}, \tag{20}$$

$$\underset{n' \times 1}{\boldsymbol{y}'} := \begin{bmatrix} \underset{n \times 1}{\boldsymbol{y}} \\ \underset{l \times 1}{\boldsymbol{z}_0} \end{bmatrix}, \tag{21}$$

$$\underset{n' \times m}{A'} := \begin{bmatrix} \underset{n \times m}{A} \\ \underset{l \times m}{K} \end{bmatrix}. \tag{22}$$

[5]See earlier comment about presumption that the mathematical model is already linearized.

Lastly, the size of the observation vectors are related by

$$n' = n + l, \tag{23}$$

and, for the uniqueness of the LESS, it is assumed that $\operatorname{rk} A' = m$.

The LESS within the VCM with Stochastic Constraints follows that of the LESS within the VCM itself, but it is then translated into the original form (13)–(15) using equations (18)–(23) as a guide. The LESS within the VCM is shown in Section 5, and the LESS within the VCM with Stochastic Constraints is shown in Section 6.

5 Least-Squares Solution within the Variance Component Model

Based on the derivation provided by Schaffrin and Snow (2017b), the following algorithm provides the (iterative) LESS within the VCM:

Let j be the iteration number. Define the vector of estimated variance components $\hat{\boldsymbol{\vartheta}}_j$ at iteration j as

$$\hat{\boldsymbol{\vartheta}}_j := \begin{bmatrix} (\hat{\sigma}_1^2)_j \\ (\hat{\sigma}_2^2)_j \end{bmatrix}. \tag{24}$$

Furthermore, define the estimated dispersion (covariance) matrix $\hat{\Sigma}_j$ as

$$\hat{\Sigma}_j := (\hat{\sigma}_1^2)_j \cdot Q_1 + (\hat{\sigma}_2^2)_j \cdot Q_2. \tag{25}$$

Then define the auxiliary matrix

$$\hat{W}_j := \hat{\Sigma}_{j-1}^{-1} - \hat{\Sigma}_{j-1}^{-1} A' (A'^T \hat{\Sigma}_{j-1}^{-1} A')^{-1} A'^T \hat{\Sigma}_{j-1}^{-1}. \tag{26}$$

Finally, the formula for the estimated variance components is given by (see Schaffrin and Snow (2017b) for details)

$$\hat{\boldsymbol{\vartheta}}_j := \begin{bmatrix} (\hat{\sigma}_1^2)_j \\ (\hat{\sigma}_2^2)_j \end{bmatrix} = \begin{bmatrix} \operatorname{tr}(\hat{W}_j Q_1 \hat{W}_j Q_1) & \operatorname{tr}(\hat{W}_j Q_1 \hat{W}_j Q_2) \\ \operatorname{tr}(\hat{W}_j Q_2 \hat{W}_j Q_1) & \operatorname{tr}(\hat{W}_j Q_2 \hat{W}_j Q_2) \end{bmatrix}^{-1} \cdot \begin{bmatrix} \boldsymbol{y}'^T \hat{W}_j Q_1 \hat{W}_j \boldsymbol{y}' \\ \boldsymbol{y}'^T \hat{W}_j Q_2 \hat{W}_j \boldsymbol{y}' \end{bmatrix}. \tag{27}$$

Note that, the inverse matrix, under the assumption of normality, may be taken as $\frac{1}{2}\hat{D}\{\hat{\boldsymbol{\vartheta}}\}$ after convergence.

So, at iteration $j = 0$ set

$$\hat{\boldsymbol{\vartheta}}_0 := \begin{bmatrix} (\hat{\sigma}_1^2)_0 \\ (\hat{\sigma}_2^2)_0 \end{bmatrix} \equiv \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \tag{28}$$

and thus compute

$$\hat{\Sigma}_0 = Q_1 + Q_2. \tag{29}$$

Continuing the algorithm for $j > 0$:

Set $j = 1$, and apply equations (26) and (27), using the $(j-1)$th solution to obtain the jth values of $\hat{W}_j$ and $\hat{\boldsymbol{\vartheta}}_j$.

Check the magnitude of $\hat{\boldsymbol{\vartheta}}_j - \hat{\boldsymbol{\vartheta}}_{j-1}$ at each iteration to see if

$$\delta = \|\hat{\boldsymbol{\vartheta}}_j - \hat{\boldsymbol{\vartheta}}_{j-1}\| < \epsilon \tag{30}$$

holds for some specified ϵ. If equation (30) is not true, then increment j and return to (25) for the next iteration. Repeat until (30) is satisfied, signifying that the algorithm has converged, thereby providing the estimated variance components.

Once iteration completes, the final estimated variance components are known in good approximation. Now dropping the subscript j for convenience, the estimated dispersion matrix of $\boldsymbol{y}'$ is provided by

$$\hat{\Sigma} = \hat{\sigma}_1^2 \cdot Q_1 + \hat{\sigma}_2^2 \cdot Q_2. \tag{31}$$

The estimated parameters themselves are computed by

$$\hat{\boldsymbol{\xi}} = (A'^T \hat{\Sigma}^{-1} A')^{-1} A'^T \hat{\Sigma}^{-1} \boldsymbol{y}', \tag{32}$$

with their estimated dispersion matrix provided by

$$\hat{D}\{\hat{\boldsymbol{\xi}}\} = (A'^T \hat{\Sigma}^{-1} A')^{-1} = [A'^T (\hat{\sigma}_1^2 \cdot Q_1 + \hat{\sigma}_2^2 \cdot Q_2)^{-1} A']^{-1}. \tag{33}$$

The adjusted observation and predicted random error (residual) vectors are, respectively,

$$\tilde{\boldsymbol{y}}' = A' \hat{\boldsymbol{\xi}}, \tag{34}$$

$$\tilde{\boldsymbol{e}}' = \boldsymbol{y}' - \tilde{\boldsymbol{y}}' = \boldsymbol{y}' - A' \hat{\boldsymbol{\xi}}. \tag{35}$$

The estimated dispersion matrix for the adjusted observations is given by

$$\hat{D}\{\tilde{\boldsymbol{y}}'\} = \hat{D}\{A' \hat{\boldsymbol{\xi}}\} = A' \hat{D}\{\hat{\boldsymbol{\xi}}\} A'^T = A' [A'^T (\hat{\sigma}_1^2 \cdot Q_1 + \hat{\sigma}_2^2 \cdot Q_2)^{-1} A']^{-1} A'^T. \tag{36}$$

And, finally, the estimated dispersion matrix for the predicted random errors (residuals) is

$$\hat{D}\{\tilde{\boldsymbol{e}}'\} = (\hat{\sigma}_1^2 \cdot Q_1 + \hat{\sigma}_2^2 \cdot Q_2) - A' [A'^T (\hat{\sigma}_1^2 \cdot Q_1 + \hat{\sigma}_2^2 \cdot Q_2)^{-1} A']^{-1} A'^T. \tag{37}$$

6 Translating back to the Variance Component Model with Stochastic Constraints

Using the LESS outlined in Section 5 (within the Variance Component Model (VCM)) and the relationships outlined in Section 4 (between the VCM and the VCM with Stochastic Constraints), the LESS within the VCM with Stochastic Constraints can be written as follows.

First, begin with the final, iterated estimates of the two variance components

$$\hat{\boldsymbol{\vartheta}} = \begin{bmatrix} \hat{\sigma}_1^2 \\ \hat{\sigma}_2^2 \end{bmatrix}. \tag{38}$$

Then, applying (19) and (20) to (31) yields the estimated dispersion (covariance) matrix $\hat{\Sigma}$ for the concatenated vector of observations $\boldsymbol{y}$ and prior information $\boldsymbol{z}_0$ as follows:

$$\hat{\Sigma} = \hat{\sigma}_1^2 \begin{bmatrix} P^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \hat{\sigma}_2^2 \begin{bmatrix} 0 & 0 \\ 0 & P_0^{-1} \end{bmatrix} = \begin{bmatrix} \hat{\sigma}_1^2 \cdot P^{-1} & 0 \\ 0 & \hat{\sigma}_2^2 \cdot P_0^{-1} \end{bmatrix}. \tag{39}$$

Parenthetically, we note that a block diagonal matrix is invertible if its diagonal blocks are invertible, and thus the inverse of the dispersion matrix in (39) is

$$\hat{\Sigma}^{-1} = \begin{bmatrix} (1/\hat{\sigma}_1^2) \cdot P & 0 \\ 0 & (1/\hat{\sigma}_2^2) \cdot P_0 \end{bmatrix}. \tag{40}$$

Next, applying (21), (22) and (40) to (32) yields the estimated parameters as a function of the observations $\boldsymbol{y}$, the prior information $\boldsymbol{z}_0$, and the two estimated variance components $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ as follows:

$$\begin{aligned} \hat{\boldsymbol{\xi}} &= \left(\begin{bmatrix} A \\ K \end{bmatrix}^T \hat{\Sigma}^{-1} \begin{bmatrix} A \\ K \end{bmatrix} \right)^{-1} \begin{bmatrix} A \\ K \end{bmatrix}^T \hat{\Sigma}^{-1} \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{z}_0 \end{bmatrix} = \\ &= \left[(1/\hat{\sigma}_1^2) \cdot A^T P A + (1/\hat{\sigma}_2^2) \cdot K^T P_0 K \right]^{-1} \left[(1/\hat{\sigma}_1^2) \cdot A^T P \boldsymbol{y} + (1/\hat{\sigma}_2^2) \cdot K^T P_0 \boldsymbol{z}_0 \right]. \end{aligned} \tag{41}$$

And, applying (22) and (40) to (33) yields the estimated dispersion matrix

$$\hat{D}\{\hat{\boldsymbol{\xi}}\} = \left(\begin{bmatrix} A^T & K^T \end{bmatrix} \hat{\Sigma}^{-1} \begin{bmatrix} A^T & K^T \end{bmatrix}^T \right)^{-1} = \left(\hat{\sigma}_1^{-2} \cdot A^T P A + \hat{\sigma}_2^{-2} \cdot K^T P_0 K \right)^{-1} \tag{42}$$

for the estimated parameters. Next, applying (21) and (22) to (34) yields

$$\begin{bmatrix} \tilde{\boldsymbol{y}} \\ \tilde{\boldsymbol{z}}_0 \end{bmatrix} = \begin{bmatrix} A \\ K \end{bmatrix} \hat{\boldsymbol{\xi}} \tag{43}$$

for the adjusted observation vector $\tilde{\boldsymbol{y}}$ and the adjusted prior information vector $\tilde{\boldsymbol{z}}_0$. Then, applying (21) and (22) to equation (35) yields the concatenated residual vector, comprised of $\tilde{\boldsymbol{e}}$ and $\tilde{\boldsymbol{e}}_0$,

$$\begin{bmatrix} \tilde{\boldsymbol{e}} \\ \tilde{\boldsymbol{e}}_0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{z}_0 \end{bmatrix} - \begin{bmatrix} \tilde{\boldsymbol{y}} \\ \tilde{\boldsymbol{z}}_0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{z}_0 \end{bmatrix} - \begin{bmatrix} A \\ K \end{bmatrix} \hat{\boldsymbol{\xi}}. \tag{44}$$

Next, we apply the law of error propagation (Koch, 1997, pp. 99, 100) to determine the estimated dispersion matrices for the adjusted observations and adjusted prior information.

$$\begin{aligned} \hat{D}\left\{\begin{bmatrix} \tilde{\boldsymbol{y}} \\ \tilde{\boldsymbol{z}}_0 \end{bmatrix}\right\} &= \hat{D}\left\{\begin{bmatrix} A \\ K \end{bmatrix} \hat{\boldsymbol{\xi}}\right\} = \begin{bmatrix} A \\ K \end{bmatrix} \left(\frac{1}{\hat{\sigma}_1^2} A^T P A + \frac{1}{\hat{\sigma}_2^2} K^T P_0 K \right)^{-1} \begin{bmatrix} A \\ K \end{bmatrix}^T = \\ &= \begin{bmatrix} A\left(\hat{\sigma}_1^{-2} \cdot A^T P A + \hat{\sigma}_2^{-2} \cdot K^T P_0 K\right)^{-1} A^T & A\left(\hat{\sigma}_1^{-2} \cdot A^T P A + \hat{\sigma}_2^{-2} \cdot K^T P_0 K\right)^{-1} K^T \\ K\left(\hat{\sigma}_1^{-2} \cdot A^T P A + \hat{\sigma}_2^{-2} \cdot K^T P_0 K\right)^{-1} A^T & K\left(\hat{\sigma}_1^{-2} \cdot A^T P A + \hat{\sigma}_2^{-2} \cdot K^T P_0 K\right)^{-1} K^T \end{bmatrix} \end{aligned} \tag{45}$$

Finally, the equation to compute the estimated dispersion matrix for the predicted random errors (residuals) is

$$\hat{D}\{\begin{bmatrix}\tilde{\boldsymbol{e}}\\ \tilde{\boldsymbol{e}}_0\end{bmatrix}\} = \left[\begin{matrix}\hat{\sigma}_1^2 P^{-1} - A\big(\hat{\sigma}_1^{-2}\cdot A^T P A + \hat{\sigma}_2^{-2}\cdot K^T P_0 K\big)^{-1} A^T \\ -K\big(\hat{\sigma}_1^{-2}\cdot A^T P A + \hat{\sigma}_2^{-2}\cdot K^T P_0 K\big)^{-1} A^T\end{matrix}\right| \;\ldots$$

$$\ldots\;\left|\begin{matrix}-A\big(\hat{\sigma}_1^{-2}\cdot A^T P A + \hat{\sigma}_2^{-2}\cdot K^T P_0 K\big)^{-1} K^T \\ \hat{\sigma}_2^2 P_0^{-1} - K\big(\hat{\sigma}_1^{-2}\cdot A^T P A + \hat{\sigma}_2^{-2}\cdot K^T P_0 K\big)^{-1} K^T\end{matrix}\right]. \tag{46}$$

7 Alternative solution: Partial-MINOLESS

In some cases it may be required to estimate the unknown parameters such that they turn out to be as close as possible to certain specified (a priori) values, while still guaranteeing that the weighted sum of squared observation residuals is minimum; i.e., the specified parameter values should not affect the adjustment of the observations. This could be the case for all the parameters or only a subset of them. For example, suppose the heights of certain network stations were determined by GNSS and that subsequent leveling observations between all the network stations needed to be adjusted. Furthermore, suppose that the GNSS-derived heights should have no affect on the adjusted leveling observations and that the corresponding estimated heights should be as close to the GNSS-obtained heights as possible for those certain stations. This example could be applied to the test network described in Section 8 below.

Network datum deficiency

The preceding example represents a case of a *network datum deficiency*, which occurs when the observations do not contain enough information to uniquely estimate the parameters of the model; i.e., $\text{rk}[A^T \mid K^T] = m > \text{rk}\, A^T = \text{rk}\, A$. The concept of datum deficiency is easy to understand by use of examples. In a geodetic or surveying network where 3D coordinates of monuments are the unknown parameters to be estimated, the coordinates represent a realization of an underlying network datum having seven parameters: scale (1), orientation (3), and origin (3). However, the coordinates themselves are rarely observed; but, rather, it is more common to observe azimuths, directions, zenith angles, and slant distances between the monuments. The angular measurements carry information about the orientation of the network, while the slant distances provide scale information. However, none of these observations provides information about the origin of the datum. Thus, the datum deficiency is three (one for each unknown origin parameter). Obviously, a 2D analog occurs in the case where only 2D coordinates are to be estimated (e.g., in the horizontal plane). Our leveling example is the 1D analog, since the parameters of interest are heights of monuments, but the leveling observations only carry information about height differences between the monuments.

From a least-squares adjustment perspective, the datum deficiency corresponds to a rank deficiency, of size d, in the least-squares normal equations, implying that the equations cannot be solved without imposing certain constraints on the parameters. At the level of the observation equations, only $m - d$ of the m columns of the coefficient matrix A are linearly independent. Note that, in our example, the rank deficiency is $d = 1$ regardless of how many heights must be estimated.

Surveyors often encounter such datum deficiencies in their work and may handle them by holding the coordinates of one station “fixed” in a network adjustment. This results in a *minimally constrained* adjustment, so called because only a minimum number of constraints are imposed upon the parameters to allow the system of equations to be solved. Such an adjustment yields unique and unbiased observation residuals, but not unbiased parameter estimates. In fact, it can be shown mathematically that the choice of datum constraint will always bias the estimated parameters of a network adjustment (whether 3D, 2D, or 1D). Thus, the adjustment that will result in a *minimum bias* among all adjustments of type minimally constrained is of particular interest.

It turns out that there is a minimally constrained adjustment that satisfies the minimum bias condition uniformly while also minimizing the changes between a priori and estimated parameters. It is called MInimum NOrm LEast-Squares Solution (*MINOLESS*), reflecting in its name the property of minimum parameter changes. The minimum bias property is proved from a statistical derivation of the equivalent Best Linear Uniformly Minimum Biased Estimate (*BLUMBE*), which also guarantees the desirable property of minimum trace of dispersion (covariance) matrix for the estimated parameters (see Schaffrin and Snow (2017b)).

We may summarize the properties of MINOLESS (and equivalent BLUMBE) by stating that it yields:

1. A unique and unbiased residual vector.
2. A minimum (weighted) sum of squared residuals.
3. A minimum change between a priori and estimated parameters in terms of L_2-norm.
4. A minimum trace of dispersion matrix for the estimated parameters.
5. A minimum bias of estimated parameters in the class of minimally constrained least-squares adjustments.

The *Partial-MINOLESS* is used for the case where changes in parameters should be minimized for only some of the unknown parameters. It is actually the more general case, since it employs a selection matrix S used to specify which parameters should be targeted for minimization of change, and, by setting S to the identity matrix, MINOLESS itself is obtained.

According to Schaffrin et al. (2018), Partial-MINOLESS can be obtained by minimizing the S-weighted parameter vector $\boldsymbol{\xi}$ subject to the condition $N\boldsymbol{\xi} = \boldsymbol{c}$, stated mathematically by the objective

$$(\boldsymbol{z}_0' - \boldsymbol{\xi})^T S(\boldsymbol{z}_0' - \boldsymbol{\xi}) = \min_{\boldsymbol{\xi}} \text{ such that } N(\boldsymbol{\xi} - \boldsymbol{z}_0') = \boldsymbol{c} - N\boldsymbol{z}_0'. \tag{47}$$

Here we have introduced $[N, \boldsymbol{c}] \coloneqq A^T P[A, \boldsymbol{y}]$, with A and $\boldsymbol{y}$ as originally defined in Section 2. We stress that the $n \times m$ matrix A is rank deficient and define its rank as

$$q \coloneqq \operatorname{rk} A < m < n. \tag{48}$$

The $m \times 1$ vector $\boldsymbol{z}_0'$ is defined by

$$\boldsymbol{z}_0' \coloneqq \begin{bmatrix} \boldsymbol{z}_0 \\ \mathbf{0} \end{bmatrix}, \tag{49}$$

where the vector of prior information $\boldsymbol{z}_0$ was defined in Section 2. In theory, the $m \times m$ matrix S can be any symmetric positive-semidefinite matrix. In its standard form, it is a diagonal matrix with ones corresponding to selected parameters and zero elsewhere. In harmony with the preceding sections, it may also be defined by

$$S \coloneqq K^T P_0 K = \begin{bmatrix} P_0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \underset{l \times m}{K} \coloneqq \begin{bmatrix} I_l \mid 0 \end{bmatrix}, \tag{50}$$

as it was in Schaffrin et al. (2018), where the $l \times m$ matrix K is rectangular (when $l < m$) with ones on its main diagonal and zeros elsewhere. Note that matrix K can always be constructed as such, though a reordering of the parameters so that the "selected" ones appear first in the vector $\boldsymbol{\xi}$ might be required.

Because of the rank deficiency, matrix N is not invertible, but as long as the matrix sum $S + N$ is invertible the Partial-MINOLESS can be represented by

$$\hat{\boldsymbol{\xi}} = \boldsymbol{z}_0' + (S + N)^{-1} N [N(S + N)^{-1} N]^- (\boldsymbol{c} - N\boldsymbol{z}_0'), \tag{51}$$

with the estimated dispersion matrix

$$\begin{aligned} \hat{D}\{\hat{\boldsymbol{\xi}}\} = \hat{\sigma}_0^2 (S + N)^{-1} N [N(S + N)^{-1} N]^- N [N(S + N)^{-1} N]^- N (S + N)^{-1} + \\ + \{I_m - (S + N)^{-1} N [N(S + N)^{-1} N]^- N\} \cdot \begin{bmatrix} \hat{\Sigma}_0 & 0 \\ 0 & 0 \end{bmatrix} \cdot \\ \cdot \{I_m - (S + N)^{-1} N [N(S + N)^{-1} N]^- N\}^T. \end{aligned} \tag{52}$$

Here, the raised minus sign stands for a generalized inverse, e.g., G^- stands for a generalized inverse of matrix G. Note that any generalized inverse of the matrix product $N(S + N)^{-1} N$ in (51) and (52) will suffice.

A perhaps somewhat simpler expression for Partial-MINOLESS can be derived from a set of minimum constraints

$$(ES)(\boldsymbol{z}_0' - \boldsymbol{\xi}) = \mathbf{0}, \quad \text{where} \quad NE^T = \mathbf{0}, \tag{53}$$

with the same matrix S and vector $\boldsymbol{z}_0'$ as before, and E being a suitable $d \times m$ matrix with full row rank. It is unique if and only if

$$\operatorname{rk}(ES) = \operatorname{rk} E = m - q = d, \tag{54}$$

where, as noted above, q is the rank of matrix A, and d is the network datum deficiency. The solution may then be represented by

$$\hat{\boldsymbol{\xi}} = \boldsymbol{z}_0' + (N + SE^T ES)^{-1}(\boldsymbol{c} - N\boldsymbol{z}_0'), \tag{55}$$

with the estimated dispersion matrix

$$\begin{aligned} \hat{D}\{\hat{\boldsymbol{\xi}}\} = \hat{\sigma}_0^2 (N + SE^T ES)^{-1} N (N + SE^T ES)^{-1} + \\ + [I_m - (N + SE^T ES)^{-1} N] \cdot \begin{bmatrix} \hat{\Sigma}_0 & 0 \\ 0 & 0 \end{bmatrix} \cdot [I_m - (N + SE^T ES)^{-1} N]^T. \end{aligned} \tag{56}$$

See Snow and Schaffrin (2007) for a full derivation. It is noted that, while the coefficient matrix A does not have full column rank due to the datum deficiency, the combined matrix $[A^T, E^T]$ has full row rank and thus spans $\mathbb{R}^m$. The form of matrix E is well known for a variety of datum deficient problems in geodesy. For the example we treat in the next section, it is simply a $1 \times m$ matrix containing only ones.

Note that matrix triple products appearing after the first lines in equations (52) and (56) are due to the uncertainty in the prior information vector $\boldsymbol{z}_0'$ appearing in (51) and (55), respectively, as expressed in the weight matrix $P_0 = \sigma_0^2 \cdot \Sigma_0^{-1}$. In the case that the prior information is given without uncertainly (i.e., the weight matrix P_0 is not provided), then $\boldsymbol{z}_0'$ is treated as a constant in variance propagation, and equations (52) and (56) are reduced to their respective first lines. Obviously, in that case, P_0 would be replaced by I_l in (50).

Regardless of whether (51) or (55) is used to compute the estimated parameters, the (unique) residual vector is computed in the usual way by

$$\tilde{\boldsymbol{e}} = \boldsymbol{y} - A\hat{\boldsymbol{\xi}}, \tag{57}$$

and the estimated variance component is provided by

$$\hat{\sigma}_0^2 = \frac{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}}}{n - q}. \tag{58}$$

Finally, we note that in some of the papers referenced in this section, Partial-MINOLESS (also called S-weighted MINOLESS) appears without the term $\boldsymbol{z}_0'$. That is because in those works it had been assumed that the solution $\hat{\boldsymbol{\xi}}$ represented an incremental change to a vector of initial values, whatever they may be. Here, we include the term $\boldsymbol{z}_0'$ explicitly for easy comparison to the estimators shown in the preceding sections and note that $\hat{\boldsymbol{\xi}}$ is updated iteratively in numerical algorithms, while the vector $\boldsymbol{z}_0'$ remains unchanged at each iteration.

8 Testing the Variance Component Model with Stochastic Constraints with GPS and spirit leveling data from a small network in Corbin, Virginia

A small network was established at NGS's Training Center in Corbin, Virginia, consisting of seven control points (Figure 1). All seven points were observed using Trimble GNSS receivers with Trimble Zephyr Geodetic antennas (TRM41249.00) for three 24-hour sessions on days of year 218–220 of 2016. However, for the purposes of this example, only GPS[6] data from the outermost three points (1–3) were used for this test. Additionally, first-order class I spirit leveling connected the seven points in 12 double-run sections using a Leica DNA03 digital automatic level. The lengths of each run are shown in Figure 1.

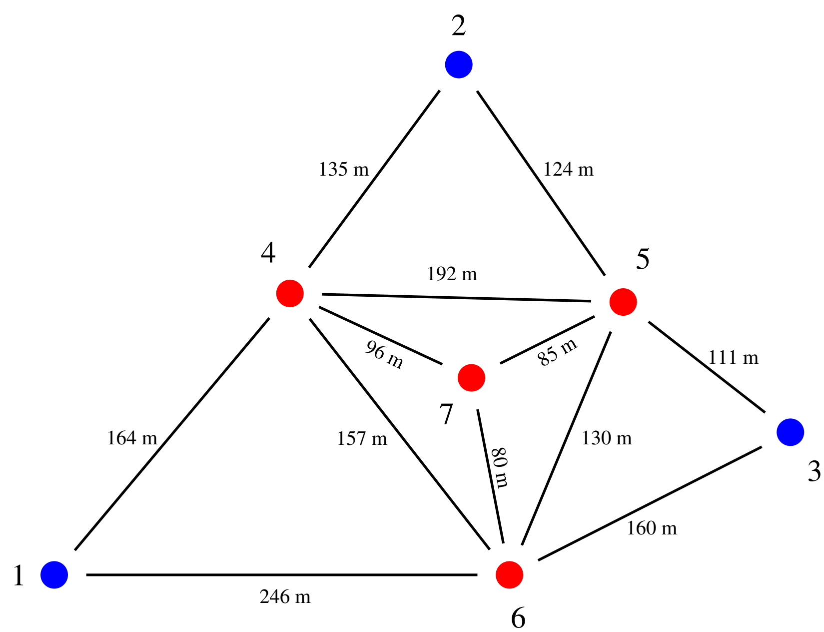


Figure 1: Test network at Corbin, Virginia. Points 1, 2, and 3 appear as stations MTCH, QAD2, and P150, respectively in Figure 2b.

In order to test the Variance Component Model with Stochastic Constraints using GPS data spanning collection times more typical of surveying practices,

[6]From here on we switch from the use of GNSS (Global Navigation Satellite System) to GPS (Global Positioning System), since only GPS data were processed for this investigation.

three 5-hour GPS sessions were extracted from the total GPS data collected at points 1–3. The first session was from 0:00 to 05:00 (GPS time) on the first day of the survey, the second session was from 02:00 to 07:00 on the second day, and the third session was from 04:00 to 09:00 on the third day.

The data were post-processed using OPUS-Projects, version 2.7.2. In addition to the GPS data from points 1–3 (Figure 2b), 24-hour GPS data files at six Continuously Operating Reference Stations (CORS) (stations BREW, LOY8, LOYB, LOYJ, LOYM, LOYO) were added to each session (Figure 2a).

The nearest CORS, LOY8, is within 12 km of the Corbin test network and was designated as the hub station[7] for all three sessions. One of the CORS, BREW, was more than 1,000 km from the hub and was only included for minimizing tropospheric modeling errors in the GPS baseline processing.

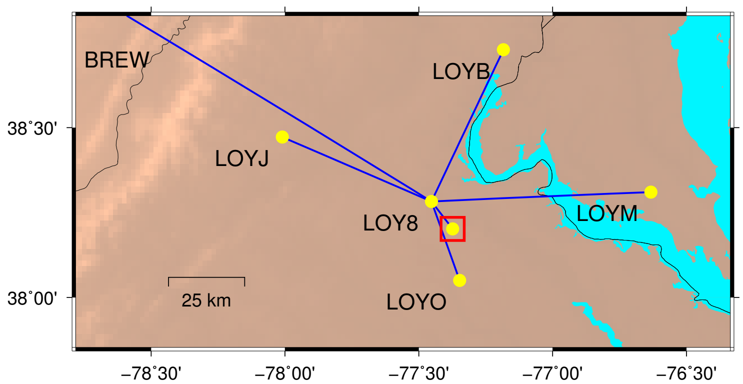
During session baseline processing, the coordinates for all of the CORS, with the exception of the hub, were constrained with *normal*[8] constraint weights. The outputs of all three processing sessions were used as observations in a GPS network adjustment, where the coordinates of all of the CORS, except for station BREW, were constrained with normal constraint weights.

The final estimates of the ellipsoid heights, as well as the estimated dispersion (covariance) matrix, were extracted from the output ascii file in Solution INdepedent EXchange Format (SINEX). Next, the NGS xGEOID16B[9] gravimetric geoid undulation model was applied to the estimated ellipsoid heights to yield estimated orthometric heights. As the purpose of this exercise was mostly to test the feasibility of the Variance Component Model with Stochastic Constraints, the geoid model random errors were not incorporated into this test. As such, the estimated dispersion matrix of the estimated ellipsoid heights was considered identical to the estimated dispersion matrix of the estimated orthometric heights. However, to be more rigorous, geoid model random errors should be propagated into the dispersion matrix of the orthometric heights. The estimated orthometric heights and the estimated dispersion matrix of the estimated orthometric heights became the 3×1 vector $\boldsymbol{z}_0$ and the (full) 3×3 cofactor submatrix P_0^{-1}, respectively, in the data model (cf. (14) and (15)).

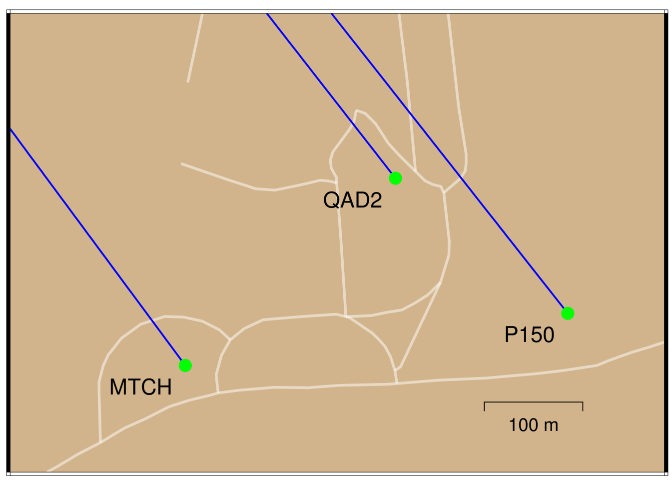
Each of the spirit leveling runs was processed using NGS leveling programs TRANSLEV, REDUC6, and ASTA. TRANSLEV was used to convert the raw leveling data into a data format that is suitable for entry in REDUC6 (i.e., Bluebooking format). REDUC6 was then used to apply rod scale, rod temperature, refraction, and astronomic corrections to each run to reduce systematic errors. Magnetic corrections were not necessary, and level corrections had already been applied in the field prior to processing. Surface gravity values at each point were

[7]NGS recommends using a *hub network* design when performing session baseline processing in OPUS-Projects (Armstrong et al. 2015; Gillins and Eddy 2017). In this design, a station that is preferably a CORS within approximately 100 km of all other observed stations in a session is designated as the hub. Then, baselines are processed such that every other observed station in the session is directly connected to the hub.

[8]In OPUS-Projects, users can apply *tight*, *normal*, or *loose* constraint weights which are meant to restrict the solution to within 0.1 mm, 1 cm, and 1 m of the coordinates of the control, respectively. Normal constraint weights are recommended, as they allow shifts on the same order as the typical accuracy of the published coordinates of the CORS.



(a) GPS CORS with red box enclosing Corbin test network. All baselines emanate from hub station LOY8.



(b) GPS stations in the Corbin test network with baselines connected to hub station LOY8. Stations MTCH, QAD2, and P150 correspond to points 1, 2, and 3, respectively, of Figure 1, and their orthometric heights are the first three elements in the parameter vector $\boldsymbol{\xi}$ of (13).

Figure 2: GPS baselines processed in OPUS-Projects

assigned from the NAVD 88 Surface Gravity model. Then, the file in REDUC6 format was loaded in ASTA. ASTA was then used to output both the average geopotential number difference and the average orthometric height difference for the two runs of each double-run leveling section. The average orthometric height difference values for each of the 12 sections in the leveling network became the 12×1 observation vector $\boldsymbol{y}$ in the model (cf. (13)). The individual, single runs for each section could also have been used, but this was deemed unnecessary for the purposes of this study.

The Federal Geodetic Control Subcommittee (FGCS, previously Federal Geodetic Control Committee, FGCC) first-order class I section mis-closure standards (FGCC, 1984) state

$$\textit{Section misclosure not to exceed } (\text{mm}) = \\ [3\,\text{mm}] \times \sqrt{\text{shortest one-way length of section in km}}.$$

It is noted that section misclosure standards are *maximum allowable* misclosures and thus deemed worst-case scenarios. Obviously, the typical misclosures for each section must be smaller than these. Moreover, there are other methods for assigning variances to leveling observations, including those that are based on instrument specifications, and such methods could have been applied in this experiment. However, our primary purpose is to evaluate the feasibility of the VCM with Stochastic Constraints, one advantage of which is that it provides an opportunity to estimate variance components that are expected to accurately account for mis-scaled cofactor matrices. As such, the following (conservative) formula was used for specifying the diagonal elements of the 12×12 cofactor submatrix P^{-1}:

$$\sigma_i^2 = \Big([3\,\text{mm}] \times \sqrt{\text{shortest one-way length of section } i \text{ in km}}\Big)^2.$$

In summary, Tables 3 and 4 list the inputs to the Variance Component Model with Stochastic Constraints.

Table 3: GPS-derived orthometric heights and associated dispersion (covariance) matrix used in model (13)–(15)

$\boldsymbol{z}_0$ [m]	$P_0^{-1}\,[10^{-6}\,\text{m}^2]$		
68.8569	2.84068	0.53399	0.53574
66.9471	0.53399	2.14133	0.53153
68.1559	0.53574	0.53153	2.19380

Various independent computer routines were built in MATLAB, C++, and FORTRAN to estimate the unknown quantities of the VCM with Stochastic Constraints. More specifically, they were designed to estimate orthometric heights as parameters and to estimate independent variance components for the leveling observations and stochastic constraints. The estimates from all the

[9] The "x" in the geoid model name stands for experimental. In preparation for NAPGD2022, NGS has annually released experimental gravimetric geoid models since 2014. The models are based on the gravity data from the latest satellite gravity models as well as from terrestrial, airborne and shipborne gravity surveys. The experimental geoid models are excellent estimates of the final gravimetric geoid model that will be developed for NAPGD2022, named GEOID2022.

Table 4: Leveling observations and their variances, with intra-station distances

From point	To point	$\boldsymbol{y}$ [m]	Diagonal elements of P^{-1} $[10^{-6}\,\text{m}^2]$	Distance [m]
6	1	0.333557	2.214	246
3	6	0.365859	1.440	160
4	1	2.850824	1.476	164
2	4	−0.948661	1.215	135
5	2	−1.040570	1.116	124
6	7	−0.824317	0.720	80
5	4	−1.989007	1.728	192
6	5	−0.528043	1.170	130
4	6	2.517497	1.413	157
7	4	−1.692892	0.864	96
5	7	−0.296337	0.765	85
3	5	−0.162582	0.999	111

routines converged to identical values at approximately the same rate (i.e., four to five iterations, where $\epsilon = 10^{-6}$ was set for the convergence criterion in (30)).

The estimated variance components are listed in Table 5. The variance component for the leveling observations is significantly less than 1 ($\hat{\sigma}_1^2 = 0.017381$), confirming our speculation that FGCS section misclosure standards are significantly larger than the actual random errors in leveling observations. The variance component for the GNSS-derived orthometric heights is significantly greater than 1 ($\hat{\sigma}_2^2 = 8.709800$), indicative of a well-known result that dispersion matrices from GNSS baseline processors are often too optimistic. The estimation of these two variance components helped with the pessimism and optimism of the a priori cofactor matrices for the leveling observations and GNSS-derived orthometric heights, respectively.

Incidentally, we note that for a numerical check on the accuracy of the coded algorithms, one could substitute into the GMM with Stochastic Constraints (equations (1)–(3)) the original cofactor submatrices P^{-1} and P_0^{-1} scaled by the estimated variance components $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$, respectively, and then solve equations (4)–(7). In this case, the estimated variance component associated with that model should turn out to be 1 (giving due consideration to the convergence criterion used and numerical computing precision).

Table 5: Estimated variance components

Quantity	Estimate
$\hat{\sigma}_1^2$	$(0.131838)^2 = 0.017381$
$\hat{\sigma}_2^2$	$(2.951237)^2 = 8.709801$

The estimated parameters (orthometric heights at the seven stations) and their estimated dispersion matrix are shown in Table 6. They are shown in further detail in Table 8. Finally, the adjusted leveling observations $\tilde{\boldsymbol{y}}$ (recall,

Table 6: Estimated orthometric heights $\hat{\boldsymbol{\xi}}$ and their estimated dispersion matrix

$\hat{\boldsymbol{\xi}}$ [m]	$\hat{D}\{\hat{\boldsymbol{\xi}}\}$ [10^{-6} m^2]						
68.8534	9.91288	9.89661	9.89624	9.90177	9.89915	9.90113	9.90065
66.9512	9.89661	9.90742	9.89675	9.90124	9.90100	9.89970	9.90060
68.1542	9.89624	9.89675	9.90805	9.89943	9.90156	9.90148	9.90089
66.0026	9.90177	9.90100	9.90156	9.90293	9.90652	9.90351	9.90436
67.9917	9.89915	9.90100	9.90156	9.90293	9.90652	9.90351	9.90436
68.5199	9.90113	9.89970	9.90148	9.90348	9.90351	9.90764	9.90499
67.6955	9.90065	9.90060	9.90089	9.90445	9.90436	9.90499	9.90913

they are differential orthometric heights) and their estimated dispersion matrix $\hat{D}\{\tilde{\boldsymbol{y}}\}$ are listed in Table 7. They are shown in further detail in Table 8.

Table 8 below shows a summary of the estimated orthometric heights, adjusted leveling observations, and the empirical standard deviations of each (being the square roots of the diagonal elements of the respective matrices in Tables 6 and 7). We note that the empirical standard deviations are relatively small. This could be explained in part by the rigor of the field work and in part by the relatively high redundancy in the observations (owing to the strong connectivity between stations 5–7). The small geographical size of the network most likely helped to limit the accumulation of any unaccounted-for biases, too. In any case, it is always recommended to pay attention to empirical standard deviations (and residuals), asking whether they seem reasonable, as part of a post-adjustment review.

Table 9 shows the residuals of the GPS-derived orthometric heights and the leveling observations. The residuals were divided by their corresponding empirical standard deviations (i.e., square roots of the diagonals of the estimated dispersion matrix from (31)) to produce *studentized residuals*.

Aside from the fact that the Variance Component Model with Stochastic Constraints proved to be suitable and that the least-squares solution within the model converged quickly, what is especially interesting about the numerical results is that the estimated heights are influenced by the random error of the GPS data at the 2–4-mm level, but the differential heights are changed only at the sub-mm level; apparently their residuals are governed almost entirely by the leveling data. In summary, given the relative precision of the observations (GPS and spirit leveling), we conclude that the GPS-derived orthometric heights primarily provided height datum information, while the leveling data determined the relative accuracy of the estimated heights. This is more-or-less as we expected.

Numerical results for Partial-MINOLESS

The Partial-MINOLESS was computed in accordance with the formulas presented in Section 7. Somewhat surprisingly, the estimated parameters (ortho-

Table 7: Adjusted leveling observation and their dispersion matrix

$\tilde{\boldsymbol{y}} = A\hat{\boldsymbol{\xi}}$ [m]	Columns 1–6 of estimated dispersion matrix $\hat{D}\{\tilde{\boldsymbol{y}}\}$ $[10^{-6}\,\mathrm{m}^2]$					
0.333523	0.0183	-0.0013	0.0135	0.0014	0.0013	0.0022
0.365687	-0.0013	0.0127	0.0008	0.0011	0.0010	-0.0021
2.850851	0.0135	0.0008	0.0167	-0.0009	-0.0008	-0.0014
-0.948630	0.0014	0.0011	-0.0009	0.0123	-0.0081	0.0001
-1.040546	0.0013	0.0010	-0.0008	-0.0081	0.0119	0.0001
-0.824411	0.0022	-0.0021	-0.0014	0.0001	0.0001	0.0068
-1.989176	0.0027	0.0021	-0.0018	0.0042	0.0038	0.0001
-0.528152	0.0021	-0.0042	-0.0014	-0.0019	-0.0017	0.0035
2.517328	-0.0048	0.0021	0.0032	-0.0023	-0.0021	-0.0036
-1.692917	0.0026	-0.0000	-0.0018	0.0022	0.0021	-0.0032
-0.296259	0.0000	0.0022	-0.0000	0.0019	0.0018	0.0033
-0.162464	0.0009	0.0085	-0.0006	-0.0008	-0.0007	0.0014
	Columns 7–12 of estimated dispersion matrix $\hat{D}\{\tilde{\boldsymbol{y}}\}$ $[10^{-6}\,\mathrm{m}^2]$					
	0.0027	0.0021	-0.0048	0.0026	0.0000	0.0009
	0.0021	-0.0042	0.0021	-0.0000	0.0022	0.0085
	-0.0018	-0.0014	0.0032	-0.0018	-0.0000	-0.0006
	0.0042	-0.0019	-0.0023	0.0022	0.0019	-0.0008
	0.0038	-0.0017	-0.0021	0.0021	0.0018	-0.0007
	0.0001	0.0035	-0.0036	-0.0032	0.0033	0.0014
	0.0080	-0.0036	-0.0044	0.0043	0.0037	-0.0015
	-0.0036	0.0071	-0.0036	0.0001	-0.0036	0.0029
	-0.0044	-0.0036	0.0080	-0.0044	-0.0000	-0.0015
	0.0043	0.0001	-0.0044	0.0076	-0.0032	0.0000
	0.0037	-0.0036	-0.0000	-0.0032	0.0069	-0.0015
	-0.0015	0.0029	-0.0015	0.0000	-0.0015	0.0114

metric heights) and the adjusted observations turned out to be identical to the values shown in Table 8. Thus, we may consider the respective adjustments equivalent for this example.

However, as expected, the estimated dispersion matrices between the two approaches were quite different, with that for Partial-MINOLESS computed by (52) and that within the VCM with Stochastic Constraints by (42). To give an idea of their difference in magnitudes, we give the square roots of the trace of their respective estimated dispersion matrices divided by their dimension m, that is $[\mathrm{tr}(\hat{D}\{\hat{\boldsymbol{\xi}}\})/m]^{1/2}$. For the Partial-MINOLESS we obtained ± 0.17 mm, and for the LESS within the VCM with Stochastic Constraints we obtained ± 4.2 mm, the latter perhaps seeming more realistic and the former appearing somewhat overly optimistic.

Table 8: Estimated parameters and adjusted observations in the Corbin test network. Empirical standard deviations are the square roots of the diagonal elements of the respective matrices in Tables 6 and 7.

Quantity	Estimate [m]	Empirical s.d. [m]	Quantity	Estimate [m]	Empirical s.d. [m]
$H_1\ (\hat{\xi}_1)$	68.8534	± 0.0031	$H_1 - H_6\ (\tilde{y}_1)$	0.33352	± 0.00014
$H_2\ (\hat{\xi}_2)$	66.9512	± 0.0031	$H_6 - H_3\ (\tilde{y}_2)$	0.36569	± 0.00011
$H_3\ (\hat{\xi}_3)$	68.1542	± 0.0031	$H_1 - H_4\ (\tilde{y}_3)$	2.85085	± 0.00013
$H_4\ (\hat{\xi}_4)$	66.0026	± 0.0031	$H_4 - H_2\ (\tilde{y}_4)$	-0.94863	± 0.00011
$H_5\ (\hat{\xi}_5)$	67.9917	± 0.0031	$H_2 - H_5\ (\tilde{y}_5)$	-1.04054	± 0.00011
$H_6\ (\hat{\xi}_6)$	68.5199	± 0.0031	$H_7 - H_6\ (\tilde{y}_6)$	-0.82441	± 0.00008
$H_7\ (\hat{\xi}_7)$	67.6955	± 0.0031	$H_4 - H_5\ (\tilde{y}_7)$	-1.98918	± 0.00009
			$H_5 - H_6\ (\tilde{y}_8)$	-0.52815	± 0.00008
			$H_6 - H_4\ (\tilde{y}_9)$	2.51733	± 0.00009
			$H_4 - H_7\ (\tilde{y}_{10})$	-1.69292	± 0.00009
			$H_7 - H_5\ (\tilde{y}_{11})$	-0.29626	± 0.00008
			$H_5 - H_3\ (\tilde{y}_{12})$	-0.16246	± 0.00011

Table 9: Residuals and studentized residuals of GPS-derived orthometric heights and leveling observations

Quantity	Residual [mm]	Student. residual	Quantity	Residual [mm]	Student. residual
$\tilde{e}_{0_1} = z_{0_1} - \tilde{z}_{0_1}$	3.5	0.703	$\tilde{e}_1 = y_6 - \tilde{y}_1$	0.034	0.174
$\tilde{e}_{0_2} = z_{0_2} - \tilde{z}_{0_2}$	-4.1	-0.945	$\tilde{e}_2 = y_3 - \tilde{y}_2$	0.172	1.087
$\tilde{e}_{0_3} = z_{0_3} - \tilde{z}_{0_3}$	1.7	0.391	$\tilde{e}_3 = y_4 - \tilde{y}_3$	-0.027	-0.170
			$\tilde{e}_4 = y_2 - \tilde{y}_4$	-0.032	-0.217
			$\tilde{e}_5 = y_5 - \tilde{y}_5$	-0.023	-0.168
			$\tilde{e}_6 = y_6 - \tilde{y}_6$	0.094	0.839
			$\tilde{e}_7 = y_5 - \tilde{y}_7$	0.169	0.976
			$\tilde{e}_8 = y_6 - \tilde{y}_8$	0.109	0.763
			$\tilde{e}_9 = y_4 - \tilde{y}_9$	0.169	1.077
			$\tilde{e}_{10} = y_7 - \tilde{y}_{10}$	0.025	0.205
			$\tilde{e}_{11} = y_5 - \tilde{y}_{11}$	-0.077	-0.672
			$\tilde{e}_{12} = y_3 - \tilde{y}_{12}$	-0.117	-0.890

9 Summary and recommendations

1. We have presented a new model for incorporating stochastic constraints on prior information and multiple unknown variance components, which we have called the Variance Component Model with Stochastic Constraints.

2. The least-squares solution within the VCM with Stochastic Constraints appears to converge relatively quickly and provide reasonable estimates of unknown variance components based on an example leveling network we presented and on a couple others we have investigated, too.

3. In addition, the Partial-MINOLESS was presented as an alternative least-squares solution for estimating heights from leveling data and GPS-derived heights.

4. The precise agreement of the estimated heights and the adjusted leveling observations between the two different least-squares solutions in the test network came as a bit of a surprise to us, and we have interpreted it to mean that the GPS-derived heights primarily provided datum information for the VCM with Stochastic Constraints.

5. However, the agreement between the two solutions should not be taken as conclusive evidence that the GPS-derived heights will not necessarily affect the adjusted leveling observations within the new model, as the least-squares adjustment within that model is not, in theory, of type minimal constraint.

6. In case the two different adjustments produce different solutions in a particular case, the question might be asked which one should be adopted. In such a case, the objectives of the adjustment should be given due consideration. If the variance component estimates look reasonable and the objective is to adjust all the data simultaneously, then one may prefer the adjustment within the VCM with stochastic constraints. On the other hand, if the GPS-derived heights are considered inferior to the leveling observations, or if no weights are available for the GPS-derived heights, then one may prefer the Partial-MINOLESS. Another case where Partial-MINOLESS may be favored is that of deformation monitoring, where the GPS-derived heights may have been provided initially to define the height datum and precise leveling is repeated periodically to monitor small motions of the deforming body.

7. In any case, we do not recommend the least-squares solution within the Gauss-Markov Model with Stochastic Constraints described in Section 2 since it involves two weight matrices with a common variance component, which implies that the weight matrices must be accurately known with respect to one another. As discussed in the introductory section, this condition might be hard to satisfy or confirm.

8. The small test network in Corbin, Virginia was suitable for this work, since our primary objective in using it was to show how the new model and associated least-squares adjustment algorithm could be used. Important work is still required for much larger, functioning networks. We have already begun this work and intend to publish the results of it soon in a geodetic journal.

References

Armstrong, M., Schenewerk, M., Mader, G., Martin, D., Stone, B., Foote, R., 2015. OPUS-Projects: Online Positioning User Service Baseline Processing and Adjustment Software User Instructions and Technical Guide, v. 2.5. National Oceanic and Atmospheric Administration, National Geodetic Survey.

FGCC, 1984. Standards and specifications for geodetic control networks. Federal Geodetic Control Committee.

Gillins, D., Eddy, M., 2017. Comparison of GPS Height Modernization Surveys Using OPUS-Projects and Following NGS-58 Guidelines. J. of Surv. Engrg. 143 (1), 05016007.
URL `https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29SU.1943-5428.0000196`

Koch, K.-R., 1997. Parameter Estimation and Hypothesis Testing in Linear Models. Springer-Verlag.

NGS, 2017. Blueprint for 2022, Part 2: Geopotential Coordinates. NOAA Technical Report NOS NGS 64.

Schaffrin, B., Snow, K., September 2017a. Adjustment Computations. Class notes, School of Earth Sciences, Geodetic Science Division, The Ohio State University, Columbus, Ohio, based on Geodetic Science Courses GS650 and GS651 taught by Prof. B. Schaffrin. Compiled by K. Snow.
URL `https://earthsciences.osu.edu/sites/earthsciences.osu.edu/files/AdjustmentNotes_2017-09-06_0.pdf`

Schaffrin, B., Snow, K., September 2017b. Advanced Adjustment Computations. Class notes, School of Earth Sciences, Geodetic Science Division, The Ohio State University, Columbus, Ohio, based on Geodetic Science Course GS762 taught by Prof. B. Schaffrin. Compiled by K. Snow.
URL `https://earthsciences.osu.edu/sites/earthsciences.osu.edu/files/AdvancedAdjustmentNotes_2017-09-06_0.pdf`

Schaffrin, B., Snow, K., Fang, X., 2018. Alternative approaches for the use of uncertain prior information to overcome the rank-deficiency of a linear model. In: Tez, M., von Rosen, D. (Eds.), Trends and Perspectives in Linear Statistical Inference. Springer, Heidelberg, New York, etc., Ch. 1, pp. 181–201.

Snow, K., Schaffrin, B., 2007. GPS network analysis with BLIMPBE: An alternative to least-squares adjustment for better bias control. J. of Surv. Engrg. 133 (3), 114–122.

10 Appendix A

The final page shows a comparative list of equations derived within the Gauss-Markov Model with Stochastic Constraints and the Variance Component Model with Stochastic Constraints.

Table 10: Comparative list of equations derived within the Gauss-Markov Model with Stochastic Constraints and the Variance Component Model with Stochastic Constraints

Quantity	Based on the Gauss-Markov Model with Stochastic Constraint	Based on the Variance Component Model with Stochastic Constraints
Method of LESS	Non-iterative	Iterative
Estimated variance component(s)	$\hat{\sigma}_0^2 = \frac{\tilde{\boldsymbol{e}}^T P \tilde{\boldsymbol{e}} + \tilde{\boldsymbol{e}}_0^T P_0 \tilde{\boldsymbol{e}}_0}{n - m + l}$	$\hat{\sigma}_1^2,\ \hat{\sigma}_2^2$ (computed iteratively)
Estimated parameters	$\hat{\boldsymbol{\xi}} = (A^T PA + K^T P_0 K)^{-1}(A^T P\boldsymbol{y} + K^T P_0 \boldsymbol{z}_0)$	$\hat{\boldsymbol{\xi}} = \left[\hat{\sigma}_1^{-2} \cdot A^T PA + \hat{\sigma}_2^{-2} \cdot K^T P_0 K\right]^{-1} \left[\hat{\sigma}_1^{-2} \cdot A^T P\boldsymbol{y} + \hat{\sigma}_2^{-2} \cdot K^T P_0 \boldsymbol{z}_0\right]$
Adjusted observations	$\begin{bmatrix} \tilde{\boldsymbol{y}} \\ \tilde{\boldsymbol{z}}_0 \end{bmatrix} = \begin{bmatrix} A \\ K \end{bmatrix} \hat{\boldsymbol{\xi}}$	$\begin{bmatrix} \tilde{\boldsymbol{y}} \\ \tilde{\boldsymbol{z}}_0 \end{bmatrix} = \begin{bmatrix} A \\ K \end{bmatrix} \hat{\boldsymbol{\xi}}$
Residuals	$\begin{bmatrix} \tilde{\boldsymbol{e}} \\ \tilde{\boldsymbol{e}}_0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{z}_0 \end{bmatrix} - \begin{bmatrix} A \\ K \end{bmatrix} \hat{\boldsymbol{\xi}}$	$\begin{bmatrix} \tilde{\boldsymbol{e}} \\ \tilde{\boldsymbol{e}}_0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{z}_0 \end{bmatrix} - \begin{bmatrix} A \\ K \end{bmatrix} \hat{\boldsymbol{\xi}}$
Estimated dispersion matrix for parameters	$\hat{D}\{\hat{\boldsymbol{\xi}}\} = \hat{\sigma}_0^2 (A^T PA + K^T P_0 K)^{-1}$	$\hat{D}\{\hat{\boldsymbol{\xi}}\} = \left(\hat{\sigma}_1^{-2} \cdot A^T PA + \hat{\sigma}_2^{-2} \cdot K^T P_0 K\right)^{-1}$
Estimated dispersion matrix for adjusted observations	$\hat{D}\{\begin{bmatrix} \tilde{\boldsymbol{y}} \\ \tilde{\boldsymbol{z}}_0 \end{bmatrix}\} = \hat{\sigma}_0^2 \begin{bmatrix} A(A^T PA + K^T P_0 K)^{-1} A^T & A(A^T PA + K^T P_0 K)^{-1} K^T \\ K(A^T PA + K^T P_0 K)^{-1} A^T & K(A^T PA + K^T P_0 K)^{-1} K^T \end{bmatrix}$	$\hat{D}\{\begin{bmatrix} \tilde{\boldsymbol{y}} \\ \tilde{\boldsymbol{z}}_0 \end{bmatrix}\} = \begin{bmatrix} A\left(\hat{\sigma}_1^{-2} \cdot A^T PA + \hat{\sigma}_2^{-2} \cdot K^T P_0 K\right)^{-1} A^T & A\left(\hat{\sigma}_1^{-2} \cdot A^T PA + \hat{\sigma}_2^{-2} \cdot K^T P_0 K\right)^{-1} K^T \\ K\left(\hat{\sigma}_1^{-2} \cdot A^T PA + \hat{\sigma}_2^{-2} \cdot K^T P_0 K\right)^{-1} A^T & K\left(\hat{\sigma}_1^{-2} \cdot A^T PA + \hat{\sigma}_2^{-2} \cdot K^T P_0 K\right)^{-1} K^T \end{bmatrix}$
Estimated dispersion matrix for residuals	$\hat{D}\{\begin{bmatrix} \tilde{\boldsymbol{e}} \\ \tilde{\boldsymbol{e}}_0 \end{bmatrix}\} = \hat{\sigma}_0^2 \begin{bmatrix} P^{-1} - A(A^T PA + K^T P_0 K)^{-1} A^T & -A(A^T PA + K^T P_0 K)^{-1} K^T \\ -K(A^T PA + K^T P_0 K)^{-1} A^T & P_0^{-1} - K(A^T PA + K^T P_0 K)^{-1} K^T \end{bmatrix}$	$\hat{D}\{\begin{bmatrix} \tilde{\boldsymbol{e}} \\ \tilde{\boldsymbol{e}}_0 \end{bmatrix}\} = \begin{bmatrix} \hat{\sigma}_1^2 P^{-1} - A\left(\hat{\sigma}_1^{-2} \cdot A^T PA + \hat{\sigma}_2^{-2} \cdot K^T P_0 K\right)^{-1} A^T & -A\left(\hat{\sigma}_1^{-2} \cdot A^T PA + \hat{\sigma}_2^{-2} \cdot K^T P_0 K\right)^{-1} K^T \\ -K\left(\hat{\sigma}_1^{-2} \cdot A^T PA + \hat{\sigma}_2^{-2} \cdot K^T P_0 K\right)^{-1} A^T & \hat{\sigma}_2^2 P_0^{-1} - K\left(\hat{\sigma}_1^{-2} \cdot A^T PA + \hat{\sigma}_2^{-2} \cdot K^T P_0 K\right)^{-1} K^T \end{bmatrix}$