## C. Kotsakis

# On the trade-off between model expansion, model shrinking, and parameter estimation accuracy in least-squares data analysis 

Received: 9 December 2004 / Accepted: 20 June 2005 / Published online: 22 August 2005
© Springer-Verlag 2005


#### Abstract

A systematic investigation of over-parameterized and under-parameterized formulations in the least-squares adjustment of linear models is performed in this paper. Overparameterization and under-parameterization are modeling effects that can often occur in the adjustment of geodetic data. The former refers to situations where new unknown parameters are added to an existing model in order to provide a more precise deterministic description for a given data set. Such an expansion may either correspond to a physically meaningful and necessary model improvement (e.g. due to the presence of unknown systematic errors in the input data) or to a fabricated data over-fitting through the inclusion of fictitious parametric terms in the mathematical model for the data adjustment. On the other hand, under-parameterization schemes emerge when the effects of existing systematic disturbances are omitted from the mathematical model that is employed for the data adjustment, thus causing a bias in the estimates for the remaining model parameters. The main focus of this study is the statistical accuracy of the estimated model parameters and the conditions under which it can be improved, either through an over-parameterized model formulation or through an under-parameterized model formulation.


Keywords Least-squares • Over-parameterization • Under-parameterization • Systematic effects • Accuracy trade-off

## 1 Introduction

The model of a linear(-ized) system of observation equations with additive random noise is a fundamental component in geodetic data analysis (Dermanis and Rummel 2000).

[^0]Despite its simplistic linear character and its inherent restriction for additive measurement errors, this model (commonly termed by statisticians as the Gauss-Markov linear model) is overwhelmingly used in all fields of geodetic research; for a systematic overview of various types of modeling alternatives in geodetic estimation problems, see Dermanis and Rossikopoulos (1988).

Two basic modeling "effects" can often take place in the adjustment of geodetic data with the linear(-ized) GaussMarkov model, namely over-parameterization and underparameterization. The first of these effects refers to situations where new unknown parameters are added to an existing model in order to provide a more detailed deterministic description of a given data set. Such an expansion may either correspond to a physically meaningful and necessary model improvement, or to a fabricated data over-fitting through the inclusion of fictitious parametric terms in the mathematical model for the data adjustment. Examples where the use of an over-parameterized model constitutes a justifiable modeling choice for the adjustment of experimental data occur in cases where the input data are affected by unknown systematic errors, or in cases where the measurement noise level has declined to such a level where new higher-order systematic effects can be detected from the adjustment of the given data. Under-parameterization schemes, on the other hand, emerge when the effects of existing systematic disturbances in the observables are omitted from the mathematical model that is employed for the data adjustment, thus causing a bias in the final estimates for the remaining model parameters.

In this paper, a theoretical analysis of over-parameterized and under-parameterized formulations for the least-squares (LS) adjustment of linear models is presented. The consequences from over-parameterized LS adjustments have been investigated in the geodetic literature mainly in the framework of partitioned model representations (Teunissen 2000, pp 92-93; Koch 1999, pp 178-179) and the study of various parametric schemes for the elimination of systematic effects from the input data (Harvey 1985, p 145; Gaspar et al. 1994; Ineichen et al. 2001; Vermeer 1997; Rutkowska 1999). The main focus of our analysis is placed on the statistical
accuracy of the estimated parameters and the conditions under which it can be improved either through an over-parameterized or through an under-parameterized model formulation. In particular, a systematic investigation is presented that will address and answer the following questions: (1) does lower data noise lead to improved parameter estimates when the underlying model is augmented by a new parametric term to account for hidden systematic effects in the input data? (2) can we obtain more accurate parameter estimates from the LS adjustment of a linear model that is simplified by ignoring certain parametric terms from its initial rigorous formulation? Both questions are studied within the standard framework of LS adjustment for full-rank linear models with heteroscedastic (and generally correlated) random errors.

## 2 Methodology

### 2.1 The basic model

The standard form of the linear model that is commonly used in geodetic data analysis is given below
$\mathbf{y}_{1}=\mathbf{A x}+\mathbf{v}_{1}$
$E\left\{\mathbf{v}_{1}\right\}=0, \quad E\left\{\mathbf{v}_{1} \mathbf{v}_{1}^{\mathrm{T}}\right\}=\mathbf{C}_{1}$
where $\mathbf{y}_{1}$ is a known observation vector, $\mathbf{x}$ is an unknown (non-random) parameter vector, $\mathbf{A}$ is a design matrix of known (non-random) coefficients, and $\mathbf{v}_{1}$ is a vector of zero-mean random errors with a prescribed covariance (CV) matrix $\mathbf{C}_{1}$. For the purpose of this paper, we assume that the design matrix A has full column rank, in which case the LS optimal inversion of Eq. (1) yields the result
$\hat{\mathbf{x}}_{(1)}=\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{y}_{1}$
Regardless of the probability distribution function for the random errors, the above solution gives the best (minimum variance) unbiased linear estimate for $\mathbf{x}$, provided that there are no other systematic (non-random) disturbances in the input data (i.e. $E\left\{\mathbf{y}_{1}\right\}=\mathbf{A x}$ ). In addition, for the special case of normal (Gaussian) random errors, the solution in Eq. (3) provides also the maximum likelihood estimate for the unknown model parameters.

The quality of the previous unbiased estimate is typically described by its CV matrix
$\mathbf{C}_{\hat{\mathbf{x}}_{(1)}}=\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\right)^{-1}$
and its diagonal elements give a quantitative measure of the statistical accuracy associated with the elements of $\hat{\mathbf{x}}_{(1)}$. Further details for the LS adjustment of linear models can be found in the textbooks by Koch (1999), Rao and Toutenburg (1999) and Teunissen (2000), and in the monograph by Grafarend and Schaffrin (1993).

### 2.2 The expanded model

The linear model of the previous section (Eqs. (1), (2)) is now revised according to the general form
$\mathbf{y}_{2}=\mathbf{A x}+\mathbf{B z}+\mathbf{v}_{2}$
$E\left\{\mathbf{v}_{2}\right\}=0, \quad E\left\{\mathbf{v}_{2} \mathbf{v}_{2}^{\mathrm{T}}\right\}=\mathbf{C}_{2}$
where $\mathbf{z}$ denotes a vector of additional unknown (non-random) parameters, and $\mathbf{B}$ is an associated known design matrix such that rank $[\mathbf{A l B}]=\operatorname{rank} \mathbf{A}+\operatorname{rank} \mathbf{B}$. The input data are affected by zero-mean random errors $\mathbf{v}_{2}$ whose statistical behavior is described by a new CV matrix $\mathrm{C}_{2}$.

The importance of the augmented model in Eq. (5) is twofold. First, it provides a straightforward option for LS adjustment problems in the presence of unknown systematic errors (biases) in the input data. In such cases, the term $\mathbf{B z}$ corresponds to the linear(-ized) effect of the external systematic errors, while the noise level remains the same as in the case with bias-free data $\left(\mathbf{C}_{1}=\mathbf{C}_{2}\right)$. Note that the observation vector is now denoted by $\mathbf{y}_{2}$ in order to distinguish it from the vector data bias-free $\mathbf{y}_{1}$ that is used in Eq. (1).

Parametric modeling is a standard tool for dealing with non-random error sources in LS adjustments and it has been extensively used in various geodetic applications, including the optimal analysis of GPS code and carrier phase measurements (Vanicek et al. 1985; Vermeer 1997; Satirapod et al. 2003), the compensation of systematic errors in the observational and kinematic models for real-time Kalman filtering navigation (Yang and Zhang 2005), the processing of satellite altimetry measurements (Tscherning and Knudsen 1986; Gaspar et al. 1994; van Gysen and Coleman 1994), the detectability of systematic effects in satellite gradiometry observations from dedicated gravity field missions (Preimesberger and Pail 2004), the integration of heterogeneous height data sets (Jiang and Duquenne 1996; Kotsakis and Sideris 1999), the consistent combination of GPS and Galileo signals for geodetic positioning (Moudrak et al. 2005), the estimation of station-dependent range biases in the adjustment of satellite laser ranging networks (Rutkowska 1999), etc.

Alternatively, Eq. (5) can be viewed as a more "detailed" image of the physical system under consideration (e.g. satellite orbit, geodetic network, gravity field, etc.), when a significant increase in the measurement precision of its observable quantities occurs. The term $\mathbf{B z}$ now represents the higherorder features of the underlying physical system that can be detected due to the use of more precise data than in the case of the simpler model in Sect. 2.1 (i.e. trace $\mathbf{C}_{2}<\operatorname{trace} \mathbf{C}_{1}$ ). As an example, we can refer to the processing of gravity measurements obtained from absolute free-fall or rise-and-fall gravimeters using an expanded form of the Newtonian equation of motion that takes into account the variations of the gravity gradient along the trajectory of the test mass (Hipkin 1999; Nagornyi 1995).

The implementation of a LS adjustment using an expanded model is not always endorsed by the existence of systematic errors in the input measurements and/or the considerable reduction in the measurement noise level. Often, the parameterization for our data analysis models is not uniquely specified a priori, and various choices need to be examined for the optimal analysis of geodetic measurements. Some examples include the study of different orbit parameterizations for the combined analysis of GPS and GLONASS data (Ineichen et al. 2001), the experimentation with different types
of corrector surfaces for the assimilation of GPS, levelling and geoid height data (Fotopoulos 2003), the use of different functional models for the parameterization of the tropospheric delay in GPS observables (Kleijer 2004), and the combination of different functional models for the optimal description of crustal deformation patterns (Liu and Chen 1998). In such cases, several parameterization choices can be integrated and tested with the help of the general model in Eq. (5), where the basic parametric form $\mathbf{A x}$ is enhanced by an additional term $\mathbf{B z}$ that may significantly improve the LS fit with the given data.

The LS inversion of the augmented linear model in Eq. (5) leads to the following solution
$\hat{\mathbf{x}}_{(2)}=\mathbf{x}_{\mathrm{o}}-\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{2}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{C}_{2}^{-1} \mathbf{B} \mathbf{M}^{-1} \mathbf{B}^{\mathrm{T}} \mathbf{C}_{2}^{-1}\left(\mathbf{y}_{2}-\mathbf{A} \mathbf{x}_{\mathrm{o}}\right)$
$\hat{\mathbf{z}}=\mathbf{M}^{-1} \mathbf{B}^{\mathrm{T}} \mathbf{C}_{2}^{-1}\left(\mathbf{y}_{2}-\mathbf{A} \mathbf{x}_{0}\right)$
where the quantities $\mathbf{x}_{0}$ and $\mathbf{M}$ are given by the equations

$$
\begin{equation*}
\mathbf{x}_{\mathrm{o}}=\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{2}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{C}_{2}^{-1} \mathbf{y}_{2} \tag{9}
\end{equation*}
$$

$\mathbf{M}=\mathbf{B}^{\mathrm{T}}\left(\mathbf{C}_{2}^{-1}-\mathbf{C}_{2}^{-1} \mathbf{A}\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{2}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{C}_{2}^{-1}\right) \mathbf{B}$
The auxiliary term $\mathbf{x}_{0}$ corresponds to the (biased) LS solution that is obtained if we ignore the dependence of the observable vector on the "nuisance" term $\mathbf{B z}$, whereas the matrix $\mathbf{M}$ is commonly known as the Schur complement for the system of normal equations that is formed from the linear model of Eqs. (5) and (6).

The application of covariance propagation to Eqs. (7) and (8) yields the CV matrix for the estimates of the original model parameters $\mathbf{x}$

$$
\begin{align*}
\mathbf{C}_{\hat{\mathbf{x}}_{(2)}}= & \left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{2}^{-1} \mathbf{A}\right)^{-1} \\
& +\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{2}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{C}_{2}^{-1} \mathbf{B} \mathbf{M}^{-1} \mathbf{B}^{\mathrm{T}} \mathbf{C}_{2}^{-1} \mathbf{A}\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{2}^{-1} \mathbf{A}\right)^{-1} \tag{11}
\end{align*}
$$

and the CV matrix for the estimates of the additional model parameters $\mathbf{z}$
$\mathbf{C}_{\hat{\mathbf{z}}}=\mathbf{M}^{-1}$
More details on the LS adjustment of extended linear models and the theory of nuisance parameter elimination can be found in Welsch (1975), Schaffrin and Grafarend (1986), Schaffrin (2004), and Teunissen (2000, Chap. 6).

### 2.3 Neglecting data biases: does the noise level matter?

A common practice in geodetic data analysis is to neglect unmodeled systematic effects, if their magnitude is known to lie below the noise level of the available measurements. However, we should point out that the difference between the LS solutions obtained from (1) bias-free data, and (2) biased data without concurrent modeling of their systematic measurement errors, is not always affected by the bias-tonoise ratio. In particular, when the data CV matrix has the homoscedastic form $\mathbf{C}=\sigma^{2} \mathbf{I}$, then the LS estimate of the
parameter vector $\mathbf{x}$ which is determined without taking into account the presence of systematic errors in the input data, has the form

$$
\begin{align*}
\underbrace{\mathbf{x}_{\mathrm{o}}}_{\begin{array}{c}
\text { LS solution from biased data } \\
\text { (ignoring the bias presence) }
\end{array}}= & \left(\mathbf{A}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{C}^{-1} \overbrace{(\mathbf{A} \mathbf{x}+\mathbf{B} \mathbf{Z}+\mathbf{v})}^{\text {biased data }} \\
= & \underbrace{\left(\mathbf{A}^{\mathrm{T}} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}} \overbrace{(\mathbf{A} \mathbf{x}+\mathbf{V})}^{\text {bias-free data }}} \\
& +\underbrace{\left(\mathbf{A}^{\mathrm{T}} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{B} \mathbf{Z}}_{\mathbf{L S}^{\text {solution from bias-free data }}}
\end{align*}
$$

where the term $\mathbf{B z}$ describes the linear(-ized) effect of the systematic errors in the observations. Equation (13) shows that the closeness of $\mathbf{x}_{0}$ to the optimal LS solution which is obtained from bias-free data is independent of the data noise level, since the bias term $\left(\mathbf{A}^{\mathrm{T}} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{B z}$ is not affected by $\sigma^{2}$.

On the other hand, for cases with correlated and/or nonhomogeneous noise in the input data, the difference between the non-rigorous (biased) LS solution $\mathbf{x}_{0}$ obtained from biased measurements without modeling their systematic errors, and the optimal (unbiased) LS solution $\hat{\mathbf{x}}$ obtained from biasfree measurements, will depend on the magnitude of the systematic errors and the data noise structure. In such cases, we have that

$$
\underbrace{\mathbf{x}_{0}} \quad=\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{C}^{-1} \overbrace{(\mathbf{A x}+\mathbf{B} \mathbf{z}+\mathbf{v})}^{\text {biased data }}
$$

LS solution from biased data
(ignoring the bias presence)

$$
\begin{align*}
= & \underbrace{\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{C}^{-1} \overbrace{(\mathbf{A x}+\mathbf{v})}^{\text {bias-free data }}}_{\text {LS solution from bias-free data }} \\
& +\underbrace{\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{B z}}_{\text {depends on the data noise structure }} \\
= & \hat{\mathbf{x}}+\text { bias } \tag{14}
\end{align*}
$$

In general, for a given systematic error vector $\mathbf{B z}$, the size of the bias term $\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{B z}$ varies for different types of CV matrices, except of course when $\mathbf{C}=\sigma^{2} \mathbf{I}$.

## 3 Unbiasedness, over-parameterization, and under-parameterization

Geodetic data sets are often used for fitting or testing a number of different parametric models with varying degrees of complexity. Examples include the assessment of different kinematic models for the description of local and regional crustal deformation fields, the use of several types of corrector surfaces for the optimal fitting of GPS, levelling and
gravimetric geoid data, and the investigation of various coordinate transformation models for the optimal analysis of distortions in geodetic networks. An important problem in such cases is to determine which parameter(s) can be uniquely estimated under a given model design. Note that, within a linear modeling framework, a parameter is called estimable if it admits a linear unbiased estimator (Koch 1999, p 183). The estimability structure of partitioned linear models and submodels has been a classic problem in statistical estimation theory (Seely and Birkes 1980; Stewart and Wynn 1981), while in the geodetic literature a systematic investigation of estimability issues in partitioned linear models has been given by Schaffrin and Grafarend (1986), Schaffrin and Baki-Iz (2001), and Schaffrin (2004).

In the context of the present paper, a key point is the effect of over-parameterization schemes on the estimability of linear model parameters. The LS estimator $\hat{\mathbf{x}}_{(2)}$ in Eq. (7) provides the best (minimum variance) linear unbiased (BLU) estimate for the parameter vector $\mathbf{x}$, assuming that $E\left\{\mathbf{y}_{2}\right\}=$ $\mathbf{A x}+\mathbf{B z}$. In addition, the estimator $\hat{\mathbf{x}}_{(2)}$ retains its unbiasedness property even if the underlying true model does not contain the "nuisance" parametric term $\mathbf{B z}$. This means that $E\left\{\hat{\mathbf{x}}_{(2)}\right\}=\mathbf{x}$ not only when $E\left\{\mathbf{y}_{2}\right\}=\mathbf{A x}+\mathbf{B z}$ but also for $E\left\{\mathbf{y}_{2}\right\}=\mathbf{A x}$, as it can easily be verified from Eq. (7). In the latter case, $\hat{\mathbf{x}}_{(2)}$ corresponds to the result of an over-parameterized LS adjustment and it lacks the minimum-variance optimality property since the trace of its CV matrix is not minimum anymore within the class of all linear unbiased estimators for the simple linear model $E\left\{\mathbf{y}_{2}\right\}=\mathbf{A x}$.

On the other hand, the LS estimator $\hat{\mathbf{x}}_{(1)}$ of Eq. (3) provides the BLU estimate for the parameter vector $\mathbf{x}$ when $E\left\{\mathbf{y}_{1}\right\}=\mathbf{A x}$, and it does not remain unbiased if the true underlying model contains an additional parametric term (in which case we have an under-parameterized data adjustment). In fact, it is easily verified that $E\left\{\hat{\mathbf{x}}_{(1)}\right\}=\mathbf{x}+\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}}$ $\mathbf{C}^{-1} \mathbf{B z} \neq \mathbf{x}$ when $E\left\{\mathbf{y}_{1}\right\}=\mathbf{A x}+\mathbf{B z}$.

We can thus conclude that LS-type estimators remain unbiased when applied to over-parameterized models, but they become biased when used with under-parameterized data sets; see also Koch (1999, pp 178-180).

## 4 A condition for improved parameter estimates from over-parameterized linear models

The scope of this section is to derive a simple condition that can guarantee the improvement in the LS estimation accuracy of $\mathbf{x}$ when the linear model of Sect. 2.1 is replaced by the extended model of Sect. 2.2. For this purpose, we need to compare the performance of the parameter estimates $\hat{\mathbf{x}}_{(1)}$ and $\hat{\mathbf{x}}_{(2)}$, which are obtained from the LS adjustment of the corresponding models. Note that both estimators are considered unbiased and their quality assessment can be solely based on the comparison of their CV matrices. Their unbiasedness is justified from the following assumptions
$E\left\{\mathbf{y}_{1}\right\}=\mathbf{A x}$
and
$E\left\{\mathbf{y}_{2}\right\}=\mathbf{A x}+\mathbf{B z} \quad$ or $\quad E\left\{\mathbf{y}_{2}\right\}=\mathbf{A x}$
Under the above setting, the estimator $\hat{\mathbf{x}}_{(2)}$ can admit two alternative interpretations:
(i) if $E\left\{\mathbf{y}_{2}\right\}=\mathbf{A x}$ then $\hat{\mathbf{x}}_{(2)}$ corresponds to a LS adjustment where the extra (non-random) parametric term $\mathbf{B z}$ is related to "fictitious" systematic effects which do not really exist in the input data $\mathbf{y}_{2}$.
(ii) if $E\left\{\mathbf{y}_{2}\right\}=\mathbf{A x}+\mathbf{B z}$ then $\hat{\mathbf{x}}_{(2)}$ corresponds to a LS adjustment where the extra (non-random) parametric term $\mathbf{B z}$ is related to real systematic effects that exist in the input data $\mathbf{y}_{2}$.
Using Eqs. (4) and (11), we have
$\mathbf{C}_{\hat{\mathbf{x}}_{(1)}}-\mathbf{C}_{\hat{\mathbf{x}}_{(2)}}=\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\right)^{-1}-\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{2}^{-1} \mathbf{A}\right)^{-1}-\mathbf{Q}$
where the auxiliary matrix $\mathbf{Q}$ has the form
$\mathbf{Q}=\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{2}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{C}_{2}^{-1} \mathbf{B} \mathbf{M}^{-1} \mathbf{B}^{\mathrm{T}} \mathbf{C}_{2}^{-1} \mathbf{A}\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{2}^{-1} \mathbf{A}\right)^{-1}$
In order for the LS estimator $\hat{\mathbf{x}}_{(2)}$ to provide better statistical accuracy than the LS estimator $\hat{\mathbf{x}}_{(1)}$, the difference between their CV matrices in Eq. (17) must correspond to a positivedefinite matrix.

Based on this criterion, an optimal "tuning" for the data accuracy can be determined that will guarantee the improvement in the estimation accuracy of $\mathbf{x}$, for any augmenting term $\mathbf{B z}$. For this purpose, we express the data CV matrix $\mathbf{C}_{2}$ as a scaled version of the initial data CV matrix $\mathbf{C}_{1}$
$\mathbf{C}_{2}=a \mathbf{C}_{1}$
where $a$ denotes a positive scaling factor. Taking into account Eq. (19), the expression in Eq. (17) takes the form
$\mathbf{C}_{\hat{\mathbf{x}}_{(1)}}-\mathbf{C}_{\hat{\mathbf{x}}_{(2)}}=(1-a)\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\right)^{-1}-a \tilde{\mathbf{Q}}$
where the auxiliary matrix $\tilde{\mathbf{Q}}$ is
$\tilde{\mathbf{Q}}=\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{B} \tilde{\mathbf{M}}^{-1} \mathbf{B}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\right)^{-1}$
and
$\tilde{\mathbf{M}}=\mathbf{B}^{\mathrm{T}}\left(\mathbf{C}_{1}^{-1}-\mathbf{C}_{1}^{-1} \mathbf{A}\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1}\right) \mathbf{B}$
The result in Eq. (20) gives a positive-definite matrix if and only if the quadratic form $\mathbf{q}^{\mathrm{T}}\left(\mathbf{C}_{\hat{\mathbf{x}}_{(1)}}-\mathbf{C}_{\hat{\mathbf{x}}_{(2)}}\right) \mathbf{q}$ takes positive values for any nonzero vector $\mathbf{q}$. In this way, the following condition is obtained that ensures the positive definiteness of $\mathbf{C}_{\hat{\mathbf{x}}_{(1)}}-\mathbf{C}_{\hat{\mathbf{x}}_{(2)}}$
$\frac{1-a}{a}>\frac{\mathbf{q}^{\mathrm{T}} \tilde{\mathbf{Q}} \mathbf{q}}{\mathbf{q}^{\mathrm{T}}\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\right)^{-1} \mathbf{q}}$
Since $\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\right)^{-1}$ is a symmetric and positive-definite matrix, it can be factorized via a Cholesky decomposition as follows
$\mathbf{C}_{\hat{\mathbf{x}}_{(1)}}=\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\right)^{-1}=\mathbf{L L}^{\mathrm{T}}$
where $\mathbf{L}$ is a unique lower-triangular invertible matrix. By introducing a new auxiliary vector $\mathbf{p}=\mathbf{L}^{\mathrm{T}} \mathbf{q}$, the inequality condition in Eq. (23) is replaced by the relationship
$\frac{1-a}{a}>\frac{\mathbf{p}^{\mathrm{T}} \mathbf{K} \mathbf{p}}{\mathbf{p}^{\mathrm{T}} \mathbf{p}}$
which should be satisfied for any real non-zero vector $\mathbf{p}$. The matrix $\mathbf{K}$ is given by
$\mathbf{K}=\mathbf{L}^{-1} \tilde{\mathbf{Q}}\left(\mathbf{L}^{\mathrm{T}}\right)^{-1}$
The ratio of the quadratic forms on the right-hand side of Eq. (25) is commonly known as Rayleigh's quotient and its maximum value, for any real vector $\mathbf{p}$, is equal to the largest eigenvalue of the kernel $\mathbf{K}$ (Strang 1988, pp 348-349).

In this way, a sufficient condition that ensures the positive definiteness of the difference $\mathbf{C}_{\hat{\mathbf{x}}_{(1)}}-\mathbf{C}_{\hat{\mathbf{x}}_{(2)}}$ is
$\frac{1-a}{a}>\lambda_{\max }$
or, equivalently
$a<\frac{1}{1+\lambda_{\max }}$
where $\lambda_{\text {max }}$ denotes the maximum eigenvalue of $\mathbf{K}$. The last inequality sets the required improvement of the data precision (i.e. in the sense of scaling the initial data CV matrix by a positive factor $a$ satisfying Eq. (28)) that guarantees more accurate parameter estimates when a linear model is augmented by an extra parametric term $\mathbf{B z}$.

Note that the computation of $\mathbf{K}$ does not require the availability of the data values but only the knowledge of the design matrices $\mathbf{A}$ and $\mathbf{B}$, and the knowledge of the initial noise CV matrix $\mathbf{C}_{1}$. Thus, the condition in Eq. (28) can be particularly useful, for example, in the optimization of geodetic networks with respect to the presence of systematic errors in the measurements, or with respect to the selection of a suitable deformation/kinematic model for the variations in the station coordinates (in the case of time-dependent geodetic networks).

Remark 1 In the case of homoscedastic uncorrelated errors, the data CV matrices take the diagonal forms $\mathbf{C}_{1}=\sigma_{1}^{2} \mathbf{I}$ and $\mathbf{C}_{2}=\sigma_{2}^{2} \mathbf{I}$. Such oversimplified CV matrices are often used in practice due to the lack of reliable information for the data noise characteristics. Since $\mathbf{C}_{2}=\left(\sigma_{2}^{2} / \sigma_{1}^{2}\right) \sigma_{1}^{2} \mathbf{I}=$ $\left(\sigma_{2}^{2} / \sigma_{1}^{2}\right) \mathbf{C}_{1}$, the condition for the tuning factor in Eq. (28) takes the form
$\sigma_{2}<\frac{\sigma_{1}}{\sqrt{1+\lambda_{\max }}}$
The above inequality gives the necessary reduction of the noise level in order to obtain more accurate parameter estimates from the LS adjustment of an over-parameterized linear model with homogeneous uncorrelated noise in its input data.

Remark 2 If the data noise level remains the same then the accuracy of the LS estimates for the original parameters $\mathbf{x}$ always gets poorer when the extended model is used. This can easily be deduced from Eq. (17), which yields the following result for $\mathbf{C}_{1}=\mathbf{C}_{2}$
$\mathbf{C}_{\hat{\mathbf{x}}_{(2)}}=\mathbf{C}_{\hat{\mathbf{x}}_{(1)}}+\mathbf{Q}$
Since the matrix $\mathbf{Q}$ is non-negative definite, the accuracy of the LS estimator $\hat{\mathbf{x}}_{(2)}$ will be lower than the accuracy of
the LS estimator $\hat{\mathbf{x}}_{(1)}$. A general conclusion, therefore, is that estimating external biases through the LS adjustment process weakens the solution for the original model parameters in a statistical sense, but it still provides a correct mechanism to optimally eliminate systematic data errors from them. Such a result manifests a well-known fact to geodesists: overparameterization (for a constant data noise level) degrades the statistical accuracy of the LS adjustment results (Teunissen 2000, pp 92-93; Harvey 1985, p 145, Rutkowska 1999, p 123; Vermeer 1997).

Remark 3 It is interesting to point out that the extent of the accuracy improvement or the accuracy degradation for the estimates of the model parameters $\mathbf{x}$ in an over-parameterized LS adjustment (i.e. the difference between the CV matrices $\mathbf{C}_{\hat{\mathbf{x}}_{(1)}}$ and $\mathbf{C}_{\hat{\mathbf{x}}_{(2)}}$ ) is independent of the magnitude of the new parameters $\mathbf{z}$ that enter in the expanded model formulation; see Eq. (17). A constant unknown bias, for example, which affects all (or a part of) the input measurements and it is taken into account by adding a single extra parameter into the LS adjustment, will cause the same reduction in the estimation accuracy of the original model parameters $\mathbf{x}$, regardless of its true magnitude.

## 5 A condition for improved parameter estimates from under-parameterized linear models

In this section, we investigate the effects of under-parameterization in the LS adjustment of linear models. In particular, we seek a condition that can guarantee the improvement in the LS estimation accuracy of $\mathbf{x}$ when the extended model of Sect. 2.2 (Eqs. (5), (6)) is replaced by the simpler model of Sect 2.1 (Eqs. (1), (2)). For this purpose, we need again to compare the performance of the parameter estimates $\hat{\mathbf{x}}_{(1)}$ and $\hat{\mathbf{x}}_{(2)}$ that are obtained from the LS adjustment of the corresponding models.

The important difference, compared to the analysis in Sect. 4 , is that the estimator $\hat{\mathbf{x}}_{(1)}$ should be presumed biased since we now consider the case
$E\left\{\mathbf{y}_{1}\right\}=\mathbf{A x}+\mathbf{B z}$
and thus

$$
\begin{align*}
E\left\{\hat{\mathbf{x}}_{(1)}\right\} & =\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{C}^{-1} E\left\{\mathbf{y}_{1}\right\} \\
& =\mathbf{x}+\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{B z} \\
& =\mathbf{x}+\mathbf{b} \tag{32}
\end{align*}
$$

where $\mathbf{b}$ denotes the bias vector. In reality (which is now to be understood in terms of Eq. (31)), the solution $\hat{\mathbf{x}}_{(1)}$ corresponds to the result of an under-parameterized adjustment of the data set $\mathbf{y}_{1}$.

Due to the fact that $\hat{\mathbf{x}}_{(1)}$ is a biased estimate when Eq. (31) is valid, whereas $\hat{\mathbf{x}}_{(2)}$ remains unbiased regardless of the inclusion of the parametric term $\mathbf{B z}$ in the model formulation, the assessment of the under-parameterization effect must be based on the comparison of the mean squared error matrices for these two estimators. In this way, both the internal
precision (variance) and the external accuracy (bias) of the solutions are properly accounted for, since the CV matrix of an arbitrary estimator does not reflect the magnitude of its possible associated bias.

The mean squared error matrix (MSEM) of $\hat{\mathbf{x}}_{(1)}$ is
$\operatorname{MSEM}_{\hat{\mathbf{x}}_{(1)}}=E\left\{\left(\hat{\mathbf{x}}_{(1)}-\mathbf{x}\right)\left(\hat{\mathbf{x}}_{(1)}-\mathbf{x}\right)^{\mathrm{T}}\right\}$
and taking into account Eqs. (3) and (31), we get
$\operatorname{MSEM}_{\hat{\mathbf{x}}_{(1)}}=\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\right)^{-1}+\mathbf{b b}^{\mathrm{T}}$
or more analytically

$$
\begin{gather*}
\operatorname{MSEM}_{\hat{\mathbf{x}}_{(1)}}=\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\right)^{-1}+\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{B z z}^{\mathrm{T}} \\
\times \mathbf{B}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\right)^{-1} \tag{35}
\end{gather*}
$$

The MSEM of $\hat{\mathbf{x}}_{(2)}$ is identical to its CV matrix (since $\hat{\mathbf{x}}_{(2)}$ is unbiased) and it is given by (see Eq. (11))

$$
\begin{align*}
\operatorname{MSEM}_{\hat{\mathbf{x}}_{(2)}}= & E\left\{\left(\hat{\mathbf{x}}_{(2)}-\mathbf{x}\right)\left(\hat{\mathbf{x}}_{(2)}-\mathbf{x}\right)^{\mathrm{T}}\right\} \\
= & \left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{2}^{-1} \mathbf{A}\right)^{-1}+\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{2}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{C}_{2}^{-1} \mathbf{B} \mathbf{M}^{-1} \\
& \quad \times \mathbf{B}^{\mathrm{T}} \mathbf{C}_{2}^{-1} \mathbf{A}\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{2}^{-1} \mathbf{A}\right)^{-1} \tag{36}
\end{align*}
$$

The difference between $\hat{\mathbf{x}}_{(1)}$ and $\hat{\mathbf{x}}_{(2)}$, when both are used with the same set of data, lies in the fact that the former estimator neglects the influence of certain systematic disturbances in the input data. In this way, the solution $\hat{\mathbf{x}}_{(1)}$ offers greater simplicity than $\hat{\mathbf{x}}_{(2)}$, at the cost of a nonzero estimation bias $\mathbf{b}$. However, the total statistical accuracy of the biased estimate $\hat{\mathbf{x}}_{(1)}$ will be better than the statistical accuracy of the unbiased estimate $\hat{\mathbf{x}}_{(2)}$, if the difference $\operatorname{MSEM}_{\hat{\mathbf{x}}_{(2)}}$ $\operatorname{MSEM}_{\hat{\mathbf{x}}_{(1)}}$ corresponds to a positive-definite matrix. Equivalently, this means that the quadratic form $\mathbf{q}^{\mathrm{T}}$ $\left(\right.$ MSEM $_{\hat{\mathbf{x}}_{(2)}}-$ MSEM $\left._{\hat{\mathbf{x}}_{(1)}}\right) \mathbf{q}$ should take positive values for every nonzero vector $\mathbf{q}$. Taking into account Eqs. (35) and (36), and also setting $\mathbf{C}_{1}=\mathbf{C}_{2}$ since the same set of data is supposed to be used by both estimators, we have

$$
\begin{align*}
\mathbf{q}^{\mathrm{T}} & \left(\mathrm{MSEM}_{\hat{\mathbf{x}}_{(2)}}-\mathrm{MSEM}_{\hat{\mathbf{x}}_{(1)}}\right) \mathbf{q}=\mathbf{q}^{\mathrm{T}}\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \\
& \left.\times \mathbf{B}^{-1} \mathbf{M}^{-1}-\mathbf{z z}^{\mathrm{T}}\right) \mathbf{B}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\right)^{-1} \mathbf{q} \\
= & \mathbf{p}^{\mathrm{T}}\left(\mathbf{M}^{-1}-\mathbf{z z}^{\mathrm{T}}\right) \mathbf{p} \tag{37}
\end{align*}
$$

where the auxiliary vector $\mathbf{p}$ denotes the quantity $\mathbf{B}^{\mathrm{T}} \mathbf{C}_{1}^{-1}$ $\mathbf{A}\left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{1}^{-1} \mathbf{A}\right)^{-1} \mathbf{q}$.

From Eq. (37), it can be concluded that in order for the matrix difference $\operatorname{MSEM}_{\hat{\mathbf{x}}_{(2)}}-\mathrm{MSEM}_{\hat{\mathbf{x}}_{(1)}}$ to be positive-definite, the following matrix should also be positive-definite $\mathbf{G}=\mathbf{M}^{-1}-\mathbf{z z}^{\mathbf{T}}$
The positive-definiteness of $\mathbf{G}$ represents a sufficient condition that will ensure the improvement in the estimation accuracy of the model parameters $\mathbf{x}$, when the under-parameterized (biased) LS estimator $\hat{\mathbf{x}}_{(1)}$ is used instead of the rigorous (unbiased) LS estimator $\hat{\mathbf{x}}_{(2)}$. Note that the matrix $\mathbf{M}$ corresponds to the Schur complement that was already defined in Eq. (10). Its determination requires the two design matrices $\mathbf{A}$ and $\mathbf{B}$, and the knowledge of the data CV matrix. On the other hand, the matrix $\mathbf{z z}^{\mathrm{T}}$ cannot be rigorously computed in practice, since it depends on the magnitude of the true unknown parameters that are ignored by the under-parameterized estimator $\hat{\mathbf{x}}_{(1)}$. Nevertheless, the positive-definiteness of $\mathbf{G}$ can be tested in practice by using some prior information on the parameters $\mathbf{z}$ (e.g. their approximate values that enter into the evaluation of the design matrix $\mathbf{B}$ or some other type of numerical information in terms of maximum and minimum bounds for the elements of $\mathbf{z}$ ).

## 6 Conclusions

A set of conditions that guarantee the accuracy improvement for the estimated model parameters, which are obtained from either an over-parameterized or an under-parameterized LS data adjustment, has been presented in this paper.

For the case of an over-parameterized model formulation, the associated condition requires the tuning of the data CV matrix by an appropriate scaling factor that satisfies the inequality restriction given in Eq. (28). Note that improved parameter estimates in such cases can only be obtained if the data accuracy gets better by a certain required amount. The required improvement of the data accuracy has been solely modeled in this paper through a CV scaling process $\left(\mathbf{C}_{2}=\right.$ $a \mathbf{C}_{1}$ ), leaving out more general schemes (e.g. $\mathbf{C}_{2}=\mathbf{C}_{1}+\delta \mathbf{C}$ ) that can be investigated in future work.

For the case of an under-parameterized model formulation, the associated condition requires the positive-definiteness of a "criterion matrix" $\mathbf{G}$ that is defined in Eq. (38). The accuracy improvement for the parameter estimates in such cases can be obtained without enhancing the accuracy of the input data, but the determination of $\mathbf{G}$ requires in principle the knowledge of the true values of the parameters that have not been included in the model formulation. Nonetheless, its positive-definiteness can be examined in practice by using approximate values (or some other type of prior numerical information) for the elements of the omitted model parameters.

Acknowledgements The author would like to acknowledge the comments and suggestions given by Burkhard Schaffrin, which led to a major improvement of this paper.

On the trade-off between model expansion, model shrinking, and parameter estimation accuracy in least-squares data analysis

## References

Dermanis A, Rossikopoulos D (1988) Modeling alternatives in fourdimensional geodesy. Proceedings of the international symposium "Instrumentation, theory and analysis for integrated geodesy", vol 2, pp 115-145, Sopron, Hungary, May 16-20, 1988
Dermanis A, Rummel R (2000) Data analysis methods in geodesy. In Dermanis A, Grun A, Sanso F (eds) Geomatic methods for the analysis of data in earth sciences. Lecture notes in earth sciences, vol 95, Springer, Berlin Heidelberg New York, pp 17-92
Fotopoulos G (2003) An analysis on the optimal combination of geoid, orthometric and ellipsoidal height data. UCGE Report No. 20185, Dept. of Geomatics Engineering, University of Calgary, Canada
Gaspar P, Ogor F, Le Traon P-Y, Zanife Q-Z (1994) Estimating the sea state bias of the TOPEX and POSEIDON altimeters from crossover differences. J Geoph Res 99(C12):24,981-24,994
Grafarend EW, Schaffrin B (1993) Adjustment computations in linear models (in German). Bibliograph Inst, Manheim, Germany
Harvey BR (1985) The combination of VLBI and ground data for geodesy and geophysics. UNISURV Report S-27, School of Surveying, University of New South Wales, Kensington NSW, Australia
Hipkin RG (1999) Absolute determination of the vertical gradient of gravity. Metrol 36:47-52
Ineichen D, Springer T, Beutler G (2001) Combined processing of the IGS and the IGEX network. J Geod 75:575-586
Jiang Z, Duquenne H (1996) On the combined adjustment of gravimetrically determined geoid and GPS levelling stations. J Geod 70:505514
Kleijer F (2004) Troposphere modeling and filtering for precise GPS leveling. Report of the Netherlands Geodetic Commission, vol 56, Delft, The Netherlands
Koch K-R (1999) Parameter estimation and hypothesis testing in linear models, 2nd ed. Springer, Berlin Heidelberg New York
Kotsakis C, Sideris MG (1999) On the adjustment of combined GPS/levelling/geoid networks. J Geod 73:412-421
Liu QW, Chen YQ (1998) Combining the geodetic models of vertical crustal deformation. J Geod 72:673-683
Moudrak A, Konovaltsev A, Furthner J, Hammesfahr J, Bauch A, Defraigne P, Bedrich S, Schroth A (2005) Interoperability on Time: GPS-Galileo offset will bias position. GPS World 16(3):24-32
Nagornyi VD (1995) A new approach to absolute gravimeter analysis. Metrol, 32: 201-208.
Preimesberger T, Pail R (2004) Detectability of systematic effects in GOCE gradiometer measurements. Presented at the IAG international symposium on gravity, geoid and satellite missions, Porto, Portugal, August 30 - September 3

Rao CR, Toutenburg H (1999) Linear Models: least squares and alternatives ( $2^{\text {nd }}$ edition). Springer-Verlag, Berlin Heidelberg New York.
Rutkowska M (1999) Investigation on stability of network solutions estimated from satellite laser measurements for 1993-1995. Art Sat 34 (2):77-135
Satirapod C, Wang J, Rizos C (2003) Comparing different global positioning system data processing techniques for modeling residual systematic errors. J Surv Eng 129(4):129-135
Schaffrin B (2004) Some generalized equivalence theorems for leastsquares adjustment. IAG Symposia, vol 127, Springer, Berlin Heidelberg New York, pp 107-112
Schaffrin B, Grafarend E (1986) Generating classes of equivalent linear models by nuisance parameter elimination. Manuscr Geod 11:262271
Schaffrin B, Baki-Iz H (2001) Integrating heterogeneous data sets with partial inconsistencies. IAG Symposia, vol 123, Springer, Berlin Heidelberg New York, pp 49-54
Seely J, Birkes D (1980) Estimability in partitioned linear models. Ann Stat 8(2):399-406
Stewart I, Wynn HP (1981) The estimability structure of linear models and submodels. J R Stat Soc Ser B 43(2):197-207
Strang G (1988) Linear algebra and its applications 3rd ed. Harcourt Brace \& Company, Orlando, Florida
Teunissen PJG (2000) Adjustment theory: an introduction. Series on mathematical geodesy and positioning, Delft University Press, Delft
Tscherning CC, Knudsen P (1986) Determination of bias parameters for satellite altimeter by least-squares collocation. Proceedings of the 1 st Hotine-Marussi symposium on mathematical geodesy, Rome, June 3-6, 1985, pp 833-851
van Gysen H, Coleman R (1995) A uniform approach to the estimation problems of satellite altimetry. IAG symposia, vol 113, Springer, Berlin Heidelberg New York, pp 307-316
Vanicek P, Kleusberg A, Langley RB, Santerre R, Wells DE (1985) On the elimination of biases in processing differential GPS observations. Proceedings of the first international symposium on precise positioning with the global positioning system, Rockville, Maryland, April 15-19, vol I, pp 315-324, US Department of Commerce, NOOA
Vermeer M (1997) The precision of geodetic GPS and one way of improving it. J Geod 71:240-245
Welsch W (1975) Uber eine allgemeine reduzierende Gewichtsmatrix zur Elimination von Orientierungsunbekannten. ZfV 100(2):83-86
Yang Y, Zhang S (2005) Adaptive fitting of systematic errors in navigation. J Geod 79:43-49


[^0]:    C. Kotsakis

    Department of Geodesy and Surveying, Aristotle University of Thessaloniki,
    University Box 440,
    Thessaloniki 54124, Greece
    Tel.: +30 2310-996150
    Fax: +30 2310-995948
    E-mail: kotsaki@topo.auth.gr

