ISTITUTO GEOGRAFICO MILITARE
FIRENZE
Bollettino di Geodesia e Scienze Affini

Deterministic versus stochastic modelling of an unknown bias in linear Gauss-Markov models

CRISTOPHER KOTSAKIS

ESTRATTO DAL «*BOLLETTINO DI GEODESIA E SCIENZE AFFINI*»
RIVISTA DELL'ISTITUTO GEOGRAFICO MILITARE
ANNO LXIV - N. 1 - GENNAIO-FEBBRAIO-MARZO 2005

Deterministic versus stochastic modelling of an unknown bias in linear Gauss-Markov models

CHRISTOPHER KOTSAKIS
Department of Geodesy and Surveying, Aristotle University of Thessaloniki, Greece

Summary. – The option for dual modelling in optimal estimation problems, adopting either a deterministic or a probabilistic description for (some or all of) the unknown quantities involved, holds a special place in the development of modern operational geodesy. In this paper one more paradigm of the fruitful interaction between deterministic and stochastic modelling is provided by looking into alternative schemes for the optimal analysis of biased data with a linear Gauss-Markov model. The parallel study of these different schemes will hopefully contribute to a stronger understanding on the ability to model unknown data biases within the least-squares estimation framework.

MODELLIZZAZIONE STOCASTICA E MODELLIZZAZIONE DETERMINISTICA DI UN «BIAS» INCOGNITO NEI MODELLI LINEARI DI GAUSS-MARKOV.

Sommario. – L'opzione relativa alla duplice modellizzazione nei problemi di stima ottimale, adottando una descrizione deterministica oppure probabilistica per tutte o parte delle quantità incognite di interesse, occupa un posto speciale nello sviluppo della geodesia operativa moderna. In questo lavoro viene presentato un ulteriore paradigma della fruttuosa interazione tra modellizzazione deterministica e stocastica, esaminando schemi alternativi per l'analisi ottimale di dati deviati condotta per mezzo di un modello lineare di Gauss-Markov. Lo studio parallelo di questi schemi diversi contribuirà - ci si augura - ad una più profonda comprensione della possibilità di modellizzare dati deviati incogniti nell'ambito della stima ai minimi quadrati.

Keywords: bias, Gauss-Markov, least-squares, stochastic regularization.
Parole Chiave: bias, Gauss-Markov, minimi quadrati, regolarizzazione stocastica.

INTRODUCTION

The linear *Gauss-Markov* (G-M) model is one of the most popular and widely applied tools for the optimal processing of discrete noisy data. Its usefulness stems from the fact that numerous applications in geodetic data analysis, as well as in other branches of geosciences, are essentially reduced to an inversion problem for a system of linear(-ized) equations

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{v} \tag{1}$$

where $\boldsymbol{y}$ is a known vector of observations, $\boldsymbol{x}$ is a vector of unknown parameters and $\boldsymbol{A}$ is a full-column rank matrix of known coefficients. The residual vector $\boldsymbol{v}$ contains unknown random errors (data noise) whose statistical characteristics are typically given in terms of their first and second order moments, i.e.

$$E\{\boldsymbol{v}\} = \boldsymbol{0}, \qquad E\{\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}\} = \boldsymbol{C}. \tag{2}$$

The error covariance (CV) matrix $\boldsymbol{C}$ is often considered only partially known. Its uncertainty is commonly controlled by one or more unknown scaling factors (variance components) which can be estimated in practice from the available data. The above classic model has been studied extensively in the context of linear statistical inference (Rao, 1983) and it still provides a powerful framework for optimal parameter estimation problems (Koch, 1987). The range of its geodetic applications is very rich and the associated topics vary from network adjustment and deformation analysis to satellite orbit determination and gravity field approximation, and from digital terrain modelling to the harmonic analysis of geodetic time series; for an in-depth discussion on the use of the G-M linear model in the geodetic estimation process, see Dermanis and Rummel (2000).

Frequently, the data sets $\boldsymbol{y}$ that enter into eq. (1) are affected not only by random noise, but also from other external disturbances that cause a systematic offset in the observed values. Such biases are usually not included in the initial parameterization ($\boldsymbol{A}\boldsymbol{x}$) which is dictated by the theoretical laws that describe the physical system under consideration. Given the importance and the wide applicability of the linear inverse problem that corresponds to eq. (1), the aim of this paper is to study the modifications that need to be applied to the standard G-M model in the presence of an unknown bias in the input data.

The frist plausible correction that we could intuitively apply to the standard G-M model, when it is *a priori* known that the observation vector is affected by some external bias, is to revise eq. (1) according to the form

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} + \boldsymbol{v} \tag{3}$$

where $\boldsymbol{b}$ denotes the unknown bias vector. However, this modelling choice does not allow us to obtain a well defined and unique estimate for the unknown quantities $\boldsymbol{x}$, $\boldsymbol{b}$ and $\boldsymbol{v}$. Indeed, if we attempt to solve eq. (3) using the least-squares optimal criterion

$$L\,(\hat{\boldsymbol{x}},\hat{\boldsymbol{b}}) = \min_{\hat{x},\hat{b}}\,(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{v}) = \min_{\hat{x},\hat{b}}\left[(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})^{\mathrm{T}}\boldsymbol{P}\,(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})\right] \tag{4}$$

then we obtain the following system of normal equations

$$\begin{bmatrix} \boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{A} & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{P} \\ \boldsymbol{P}\boldsymbol{A} & \boldsymbol{P} \end{bmatrix}\begin{bmatrix} \hat{\boldsymbol{x}} \\ \hat{\boldsymbol{b}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{y} \\ \boldsymbol{P}\boldsymbol{y} \end{bmatrix} \tag{5}$$

which is always singular for any symmetric positive-definite weight matrix $\boldsymbol{P}$ ([1]).

The singularity of eq. (5) signifies the need to model further the unknown bias vector, if a meaningful LS solution is to be obtained from the given data. In principle, the situation is similar to the rank-deficient inversion problems that arise in geodetic network adjustments. In such cases the reference system is not fully defined by the actual geometrical observations and some external constraints are generally required for the estimation of the coordinates at the network points. Similarly, the observation vector $\boldsymbol{y}$ in the linear model of eq. (3) is not able to distinguish between data noise ($\boldsymbol{v}$) and data bias ($\boldsymbol{b}$), unless some additional modelling step takes place before the LS adjustment is performed.

2. – TWO DIFFERENT SCHEMES FOR BIAS MODELLING

Two different approaches for representing the data bias effect in linear G-M models will be investigated in this paper. The first option, which will be identified as the *type I* model, assumes that the bias vector can be expressed in the product form $\boldsymbol{b} = \beta\boldsymbol{s}$, where β denotes a scalar parameter and $\boldsymbol{s}$ is a vector of ones. This scheme

([1]) The solution of Eq. (5) requires the inversion of the *projection matrix* $[\boldsymbol{I} - \boldsymbol{A}\,(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}\,]$ which is always singular for any full column-rank matrix $\boldsymbol{A}$ (Harville, 1997) implies that all observations are affected by a common systematic offset which is modelled deterministically in terms of a single unknown parameter.

The general form of the type I model is shown in Box 1.

Box 1 – *Type I* modelling scheme for data bias in G-M models

$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \beta \boldsymbol{s} + \boldsymbol{v}$ $E\{\boldsymbol{v}\} = \boldsymbol{0}$ $E\{\boldsymbol{v}\boldsymbol{v}^T\} = \boldsymbol{C}$	$\boldsymbol{y}$: vector of biased data $\boldsymbol{x}$: vector of unknown parameters $\boldsymbol{A}$: full-column rank design matrix $\boldsymbol{s} = [1 \ldots 1]^T$ β: unknown bias parameter (scalar) $\boldsymbol{v}$: data noise vector $\boldsymbol{C}$: known CV matrix for the data noise

The second alternative does not use any specific parameterization scheme for the bias vector $\boldsymbol{b}$. Instead, the effect of the data bias is now modelled stochastically through a modification of the initial error CV matrix that appears in the G-M model of eqs. (1)-(2). The general form of this approach, which will be referred as the *type II* model, is given in Box 2. Note that the observation vector $\boldsymbol{y}$, the vector of the unknown parameters $\boldsymbol{x}$, the design matrix $\boldsymbol{A}$, and the *noise* CV matrix $\boldsymbol{C}$ are the same in both modelling schemes (*type I* and *type II*).

Box 2 – *Type II* modelling scheme for data bias in G-M models

$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \tilde{\boldsymbol{v}}$ $E\{\tilde{\boldsymbol{v}}\} = \boldsymbol{0}$ $E\{\tilde{\boldsymbol{v}}\tilde{\boldsymbol{v}}^T\} = \boldsymbol{C} + \delta\boldsymbol{C}$	$\boldsymbol{y}$: vector of biased data $\boldsymbol{x}$: vector of unknown parameters $\boldsymbol{A}$: full-column rank design matrix $\tilde{\boldsymbol{v}}$ = residual vector (includes both noise and bias) $\boldsymbol{C}$: known CV matrix for the data noise $\delta\boldsymbol{C}$: modifying term to account for a common data bias

The common analysis of the models of *type I* and *type II* is the focal point of this paper. It is important to understand that although no bias parameter *per se* has been included in the formulation of the *type II* model, the effect of a systematic offset on the observation vector $\boldsymbol{y}$ is still «emulated» by modifying its associated CV matrix $E\{(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})^T\}$. It will actually be shown that, for a certain simple form of the additive term $\delta\boldsymbol{C}$, the LS solutions obtained from the two different models become identical.

3. – LEAST-SQUARES ADJUSTMENT OF THE TYPE I MODEL

A LS inversion analysis for the model of *type I* is presented in this section. Let us repeat here that the *type I* model is simply a modification of the standard G-M model using a single deterministic parameter β to describe any common biases in the input data. The following equations give the functional and the stochastic components, respectively, of this model

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \beta \boldsymbol{s} + \boldsymbol{v} \tag{6.a}$$

$$E\{\boldsymbol{v}\} = \boldsymbol{0}, \qquad E\{\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}\} = \boldsymbol{C} \tag{6.b}$$

where $\boldsymbol{s}$ is a vector of ones, and $\boldsymbol{C}$ is the CV matrix of the actual data noise.

In order to avoid any rank deficiency problems the design matrix $\boldsymbol{A}$ should not contain any columns that can be expressed as multiples of the constant vector $\boldsymbol{s}$. Thus, if the physical system under consideration contains a «useful» parameter that affects equally every observable in the data vector $\boldsymbol{y}$ (e.g., zero degree term in polynomial surface fitting), then the term $\beta\boldsymbol{s}$ absorbs both the effect of this parameter and any external common bias in the input data.

Using the LS optimal principle for the inversion of the *type I* model

$$L(\hat{\boldsymbol{x}},\hat{\beta}) = \min_{\hat{x},\hat{\beta}} (\boldsymbol{v}\boldsymbol{C}^{-1}\boldsymbol{v}) = \min_{\hat{x},\hat{\beta}} [(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} - \beta\boldsymbol{s})^{\mathrm{T}} \boldsymbol{C}^{-1} (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} - \beta\boldsymbol{s})] \tag{7}$$

the following system of normal equations is obtained

$$\begin{bmatrix} \boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{A} & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{s} \\ \boldsymbol{s}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{A} & \boldsymbol{s}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{s} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{x}} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{y} \\ \boldsymbol{s}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{y} \end{bmatrix}. \tag{8}$$

Assuming that the matrix $\boldsymbol{A}$ has full-column rank, the unique solution of the above system will be expressed with the help of the ancillary quantities

$$\boldsymbol{N}_0 = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{A} \tag{9.a}$$

$$\hat{\boldsymbol{x}}^{(\mathrm{b})} = (\boldsymbol{N}_0^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^{-1})\,\boldsymbol{y} \tag{9.b}$$

$$\xi = (\boldsymbol{N}_0^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^{-1})\,\boldsymbol{s}. \tag{9c}$$

Note that $\hat{\boldsymbol{x}}^{(\mathrm{b})}$ is the LS-based solution that we would obtain by ignoring the bias presence in the input data. It corresponds to a biased estimate of the true unknown parameters, since

$$E\{\hat{\boldsymbol{x}}^{(\mathrm{b})}\} = \boldsymbol{x} + \beta\xi \neq \boldsymbol{x}. \tag{10}$$

The vector ξ identifies a characteristic quantity for the *type I* model and it can be directly computed from the known matrices $\boldsymbol{A}$ and $\boldsymbol{C}$, without the knowledge of the data values. In fact, the vector ξ can be viewed as a measure of the relative distortion that a common data bias would cause on the LS solution, if β was left out of the *type I* model.

Taking into account eqs. (9.a), (9.b), (9.c), it is easily shown that the LS solution of the *type I* model can be expressed by the equations

$$\boxed{\begin{aligned} \hat{\boldsymbol{x}} &= \hat{\boldsymbol{x}}^{(\mathrm{b})} - k\boldsymbol{s}^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{y}-\boldsymbol{A}\hat{\boldsymbol{x}}^{(\mathrm{b})})\,\xi \qquad (11)\\ \hat{\beta} &= k\boldsymbol{s}^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{y}-\boldsymbol{A}\hat{\boldsymbol{x}}^{(\mathrm{b})}) \qquad (12)\end{aligned}}$$

where k is used to denote the scalar quantity

$$k = \frac{1}{\boldsymbol{s}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{s} - \xi^{\mathrm{T}}\boldsymbol{N}_0\xi}\,. \tag{13}$$

The estimated parameters $\hat{\boldsymbol{x}}$ can also be expressed as a function of the estimated data bias $\hat{\beta}$ using the equation

$$\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}^{(\mathrm{b})} - \hat{\beta}\xi \tag{14}$$

or equivalently

$$\hat{\boldsymbol{x}} = \boldsymbol{N}_0^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{y}-\hat{\beta}\boldsymbol{s}). \tag{15}$$

According to the last two equations, the LS estimate for the unknown parameters $\boldsymbol{x}$ can be computed by applying appropriate corrections either to the initial biased estimate $\hat{\boldsymbol{x}}^{(\mathrm{b})}$ obtained by eq. (9b), or to the original data vector $\boldsymbol{y}$ that has been affected by the unknown bias β.

Remark. The scalar quantity k given by eq. (13) *always* has a positive value. This is an important result that will be used in the following sections of this paper. In order to establish the positivity of k, it is sufficient to show that it corresponds to the value of a positive-definite quadratic form. This can easily be verified by analyzing the denominator of eq. (13) as follows:

$$\begin{aligned} \boldsymbol{s}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{s} - \xi^{\mathrm{T}}\boldsymbol{N}_0\xi &\overset{\text{eq.(9.c)}}{=} \boldsymbol{s}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{s} - (\boldsymbol{s}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{A}\boldsymbol{N}_0^{-1})\,\boldsymbol{N}_0\,(\boldsymbol{N}_0^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{s})\\ &= \boldsymbol{s}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{s} - \boldsymbol{s}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{A}\boldsymbol{N}_0^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{s}\\ &= \boldsymbol{s}^{\mathrm{T}}\boldsymbol{C}^{-1}[\boldsymbol{I} - \boldsymbol{A}\boldsymbol{N}_0^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^{-1}]\,\boldsymbol{s}\\ &= \boldsymbol{s}^{\mathrm{T}}\boldsymbol{C}^{-1}[\boldsymbol{C} - \boldsymbol{A}\boldsymbol{N}_0^{-1}\boldsymbol{A}^{\mathrm{T}}]\,\boldsymbol{C}^{-1}\boldsymbol{s}\\ &= \boldsymbol{s}^{\mathrm{T}}\boldsymbol{C}^{-1}[\boldsymbol{C} - \boldsymbol{A}\,(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}]\,\boldsymbol{C}^{-1}\boldsymbol{s}\\ &= \boldsymbol{a}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{a}\,. \end{aligned} \tag{16}$$

Since $\boldsymbol{C}$ is a symmetric positive-definite matrix, the term $\boldsymbol{R} = \boldsymbol{C} - \boldsymbol{A}\,(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{A})^{-1}\,\boldsymbol{A}^{\mathrm{T}}$ will also correspond to a positive-definite symmetric matrix for any full-column rank matrix $\boldsymbol{A}$; see, e.g., Harville (1997). As a result, the quadratic form $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{a}$ always attains positive values for any vector $\boldsymbol{a}$, including the case of $\boldsymbol{a} = \boldsymbol{C}^{-1}\boldsymbol{s}$.

4. – ACCURACY EVALUATION IN THE TYPE I MODEL

The LS solution given by eqs. (11) and (12) provides the best linear unbiased estimators for the unknown parameters of the *type I* model. Their optimality stems from the well known *Gauss-Markov* theorem and it relies on the fact that they yield the minimum mean squared estimation error among any other linear unbiased estimator within the same model. Furthermore, if the data noise follows a Gaussian probability distribution, the LS estimators will offer the best mean squared error performance among every other unbiased estimator within the same model (both linear and nonlinear).

Since the previous LS estimates are unbiased, the assessment of their accuracy can be solely based on their variances and co-variances. In particular, the CV matrix of the estimated parameters $\hat{\boldsymbol{x}}$ is given by the formula

$$\boldsymbol{C}_{\hat{x}} = E\,\{(\hat{\boldsymbol{x}} - \boldsymbol{x})\,(\hat{\boldsymbol{x}} - \boldsymbol{x})^{\mathrm{T}}\} = \boldsymbol{N}_0^{-1} + k\xi\xi^{\mathrm{T}} \tag{17}$$

whereas the variance of the bias estimate $\hat{\beta}$ is

$$\sigma_{\hat{\beta}}^2 = E\,\{(\hat{\beta} - \beta)^2\} = k. \tag{18}$$

Note that k is the same positive quantity that was previously defined in eq. (13). Finally, the cross-CV vector between $\hat{\boldsymbol{x}}$ and $\hat{\beta}$ will be given by the expression

$$\boldsymbol{C}_{\hat{x}\hat{\beta}} = E\,\{(\hat{\boldsymbol{x}} - \boldsymbol{x})\,(\hat{\beta} - \beta)^2\} = -k\xi. \tag{19}$$

The last three equations can be derived in a straightforward way by simply applying the variance/co-variance propagation law to eqs. (11) and (12).

The previous results show that the presence of the bias parameter β in the *type I* model will *always* degrade the estimation accuracy of the rest of the parameters. This is easy to realize from eq. (17), since k is a positive scalar and $\boldsymbol{N}_0^{-1}$ corresponds to the CV matrix of the LS estimate for $\boldsymbol{x}$ in the absence of β from the *type I* model. Hence, an interesting conclusion that can be drawn is the following: despite the inclusion of the data bias in the mathematical model of the adjustment (which prevents the LS-based solution for $\boldsymbol{x}$ from being biased) the estimation accuracy of $\hat{\boldsymbol{x}}$ will still become worse than in the case of unbiased data since the CV matrix $\boldsymbol{C}_{\hat{x}}$ «increases» in this case by the amount of $k\xi\xi^{\mathrm{T}}$.

5. – LEAST-SQUARES ADJUSTMENT OF THE TYPE II MODEL

A LS inversion analysis for the model of *type II* is performed in this section. Note that the *type II* model corresponds to the following stochastically modified version of the standard G-M model

$$y = Ax + \tilde{v} \tag{20.a}$$

$$E\{\tilde{v}\} = \mathbf{0}, \quad E\{\tilde{v}\tilde{v}^{\mathrm{T}}\} = C + \delta C \tag{20.b}$$

where C is the same CV matrix that was employed by the *type I* model, and δC denotes a corrective term that can generate a bias-like behaviour in the associated random error vector $\tilde{v}$.

In contrast to the *type I* model which used a single deterministic parameter (β) to describe «zero-order effects» in the input data, the previous model absorbs such effects into the stochastic description of the residual vector via the inclusion of an additive component δC in its CV matrix. For the purpose of this paper, we will adopt the following form for the term δC

$$\delta C = \gamma^2 s s^{\mathrm{T}} \tag{21}$$

where γ^2 denotes some positive constant whose actual value is left unspecified for now. Since s is a vector of ones, δC then becomes a square matrix whose elements are all equal to the same positive value.

Using the LS optimal principle for the inversion of the type II model

$$L(\hat{x}) = \min_{\hat{x}} (\tilde{v}^{\mathrm{T}} \tilde{C}^{-1} \tilde{v}) = \min_{\hat{x}} [(y - Ax)^{\mathrm{T}} \tilde{C}^{-1} (y - Ax)] \tag{22}$$

the following system of normal equations is now obtained

$$(A^{\mathrm{T}} \tilde{C}^{-1} A)\, \hat{x} = A^{\mathrm{T}} \tilde{C}^{-1} y. \tag{23}$$

Note that the error CV matrix $\tilde{C}$ is now given by the expression

$$\tilde{C} = C + \gamma^2 s s^{\mathrm{T}}. \tag{24}$$

Assuming again that the design matrix A has full-column rank and using the well known matrix identities (Harville 1997)

$$(C + \gamma^2 s s^{\mathrm{T}})^{-1} = C^{-1} - \frac{\gamma^2 C^{-1} s s^{\mathrm{T}} C^{-1}}{1 + \gamma^2 s^{\mathrm{T}} C^{-1} s} \tag{25}$$

$$(N_0 - \lambda A^{\mathrm{T}} C^{-1} s s^{\mathrm{T}} C^{-1} A)^{-1} = N_0^{-1} + \frac{\lambda N_0^{-1} A^{\mathrm{T}} C^{-1} s s^{\mathrm{T}} C^{-1} A N_0^{-1}}{1 - \lambda s^{\mathrm{T}} C^{-1} A N_0^{-1} A^{\mathrm{T}} C^{-1} s} \tag{26}$$

with λ being some arbitrary positive scalar, it is easily shown that the LS solution of the *type II* model can be expressed in the form

$$\boxed{\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}^{(b)} - \tilde{k}\boldsymbol{s}^{T}\boldsymbol{C}^{-1}\,(\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}^{(b)})\,\xi} \tag{27}$$

where $\hat{\boldsymbol{x}}^{(b)}$ and ξ are the same quantities that were previously introduced in eqs. (9.b) and (9.c).

The above result is almost identical with the corresponding estimate that was obtained from the LS solution of the *type I* model; see eq. (11). The only difference is that a new multiplicative factor $\tilde{k}$ is now used in place of k. In particular, the factor $\tilde{k}$ incorporates the variance factor γ^2 as follows:

$$\tilde{k} = \frac{1}{\dfrac{1}{\gamma^2} + \boldsymbol{s}^{T}\boldsymbol{C}^{-1}\boldsymbol{s} - \xi^{T}\boldsymbol{N}_0\xi}. \tag{28}$$

Taking into account eq. (13), we can obtain the following relationship

$$\tilde{k} = \frac{1}{\dfrac{1}{\gamma^2} + \dfrac{1}{k}} = k\,\frac{\gamma^2}{\gamma^2 + k}. \tag{29}$$

Since k is always a positive scalar (see the remark in section 3), it is evident from the last formula that $\tilde{k}$ will also be a positive quantity which is always smaller than k.

6. – ACCURACY EVALUATION IN THE TYPE II MODEL

As in the case of the *type I* model, the optimality of the LS estimator for the *type II* model follows from the *Gauss-Markov* theorem and it relies on the fact that it yields the minimum mean squared estimation error among any other linear unbiased estimator.

The CV matrix of the estimated parameters in the *type II* model can be easily derived by applying the variance/co-variance propagation law to eq. (27). The final result is given below

$$\boldsymbol{C}_{\hat{x}} = E\,\{(\hat{\boldsymbol{x}} - \boldsymbol{x})\,(\hat{\boldsymbol{x}} - \boldsymbol{x})^{T}\} = \boldsymbol{N}_0^{-1} + \tilde{k}\xi\xi^{T}. \tag{30}$$

The above matrix is similar to the corresponding CV matrix obtained from the LS solution of the *type I* model, with the only difference being the presence of the multiplicative factor $\tilde{k}$ instead of k. Note again the degradation in the accuracy of the estimated parameters that is caused by the additive positive component $\tilde{k}\xi\xi^{\mathrm{T}}$. Here, however, the reason for the accuracy degradation is not the inclusion of an additional bias parameter in the mathematical model of the adjustment, but the modification of the stochastic model for the input data by the amount $\gamma^2 \boldsymbol{s}\boldsymbol{s}^{\mathrm{T}}$.

An interesting point is that the degrading term $k\xi\xi^{\mathrm{T}}$ for the accuracy of the estimated parameters $\hat{\boldsymbol{x}}$ in the *type I* model does not depend at all on the bias parameter β, whereas the degrading term $\tilde{k}\xi\xi^{\mathrm{T}}$ for the accuracy of the estimated parameters $\hat{\boldsymbol{x}}$ in the *type II* model depends directly on the selected variance factor γ^2.

7. – EQUIVALENCE BETWEEN THE MODELS OF TYPE I AND TYPE II

From the results of the previous sections, we see that a strong link exists between the two modified versions (*type I* and *type II*) of the standard G-M model. The key point of their inherent connection lies on the relationship between the scalar factors k and $\tilde{k}$ that appear in the LS solutions for the two model types. Using eq. (29), we can verify that the value of $\tilde{k}$ will continuously approach k as γ^2 increases, i.e.

$$\lim_{\gamma^2 \to \infty} \tilde{k} = \lim_{\gamma^2 \to \infty} \left(k \frac{\gamma^2}{\gamma^2 + k} \right) = k. \tag{31}$$

Hence, the linear models of *type I* and *type II* can be considered equivalent when the variance factor γ^2 in the data CV matrix $\tilde{\boldsymbol{C}}$ becomes infinite! Indeed, in this case the corresponding LS estimates for the unknown parameters $\boldsymbol{x}$, as well as their associated CV matrices, become exactly equal.

Given the particular structure of the *type I* and *type II* models, such a result can be summarized in the following words: *a constant unknown bias that is modelled deterministically through a single scalar parameter has the same effect in a LS estimation problem as a fully-correlated random noise component whose CV matrix has the form $\gamma^2 \boldsymbol{s}\boldsymbol{s}^{\mathrm{T}}$ and its corresponding variance factor γ^2 approaches infinity.*

The equivalence between the models of *type I* and *type II* originates from the fact that the latter is essentially a stochastically transformed version of the former. In order to recognize this dualism, let us consider again the basic form of the *type I* model

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \beta\boldsymbol{s} + \boldsymbol{v} \tag{32.a}$$

$$E\{\boldsymbol{v}\} = \boldsymbol{0}, \qquad E\{\boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}\} = \boldsymbol{C} \tag{32.b}$$

and let us now *associate the scalar parameter* β *with a zero-mean random variable which is uncorrelated with the random observation errors*. In statistical terms we will have

$$E\{\beta\} = 0 \tag{33.a}$$

$$E\{\beta^2\} = \sigma_\beta^2 \tag{33.b}$$

$$E\{\beta \boldsymbol{v}\} = \mathbf{0}. \tag{33.c}$$

The preceding approach introduces a *Bayesian-like* viewpoint to our estimation problem by assigning a stochastic interpretation to the bias effect in the input data. Under this new setting, the linear model of eq. (32a) can take the equivalent form

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \tilde{\boldsymbol{v}} \tag{34}$$

where $\tilde{\boldsymbol{v}}$ contains the combined effect of the independent random components $\boldsymbol{v}$ and $\beta\boldsymbol{s}$. The first and second order moments of the random vector $\tilde{\boldsymbol{v}}$ can be computed as follows

$$E\{\tilde{\boldsymbol{v}}\} = E\{\beta\boldsymbol{s} + \boldsymbol{v}\} = E\{\beta\}\,\boldsymbol{s} + E\{\boldsymbol{v}\} = \mathbf{0} \tag{35.a}$$

$$E\{\tilde{\boldsymbol{v}}\tilde{\boldsymbol{v}}^{\mathrm{T}}\} = E\{(\beta\boldsymbol{s} + \boldsymbol{v})(\beta\boldsymbol{s} + \boldsymbol{v})^{\mathrm{T}}\} = \boldsymbol{C} + \sigma_\beta^2 \boldsymbol{s}\boldsymbol{s}^{\mathrm{T}} \tag{35.b}$$

where $\boldsymbol{C}$ is the CV matrix of the noise vector $\boldsymbol{v}$.

Therefore, if we identify the parameter β in the *type I* model with a zero-mean random variable whose variance is identical to the factor γ^2, then we arrive directly at the formulation of the *type II* model. In the absence of any prior information on β, its variance σ_β^2 should be set equal to infinity. This is not an absurd choice and it just reflects the maximum level of uncertainty that we have regarding the magnitude of the data bias effect. Such a choice has been also justified through the algorithmic equivalence that was established between the LS solutions of the *type I* and *type II* models, when $\gamma^2 = \infty$.

8. – THE ROLE OF THE VARIANCE FACTOR γ^2

In the previous sections it was established that, if the variance factor γ^2 in the *type II* model is set equal to an infinitely large value, then the results for $\hat{\boldsymbol{x}}$ and $\boldsymbol{C}_{\hat{x}}$ are the same as the ones obtained by the *type I* model for the same data set. It was also explained that such an outcome is totally reasonable and it complies with the lack of any *a priori* information for the unknown bias parameter β in the *type I* model. In this way, the factor γ^2 that is employed in the *type II* model has the role of expressing in statistical terms, the uncertainty level about any common bias in the input data.

The choice of $\gamma^2 = \infty$ is merely related to a maximum level of uncertainty regarding the magnitude of the bias parameter β.

One can reasonably claim that, in practice, the data bias β cannot exceed some maximum limit and the range of its possible values is certainly bounded. Therefore, it would seem logical to favor the use of the *type II* model (with a finite variance factor γ^2) for the LS analysis of a biased data set, rather than following the deterministic modelling choice of the *type I* model. The value of γ^2 in such cases should be tuned to some level which will represent a «worst-case» bias effect in the input data.

Note that, if we use the bias modelling choice induced by the *type II* model, the accuracy of the estimated parameters $\hat{\boldsymbol{x}}$ improves since their CV matrix has smaller diagonal elements than the corresponding CV matrix of the *type I* model. This is true since we always have that $\tilde{k} < k$; see eqs. (17) and (30). An open problem is of course the actual determination of a proper finite value for γ^2, which nonetheless goes beyond the scope of the present paper.

An alternative interpretation for the factor γ^2 is that of an arbitrary *regularization parameter* that leads to a family of *biased estimators* for the *type I* model. In contrast to the preceding viewpoint, the value of γ^2 does not reflect now any prior statistical knowledge on the data bias. In this case, we simply replace the coefficient k with the coefficient $\tilde{k}$ in the LS estimators of the *type I* model. The resulting formulae, which now depend directly on γ^2, generally provide biased estimates for the unknown quantities of the *type I* model. It can actually be shown that these new estimators offer a guaranteed improvement in terms of mean squared error (MSE) reduction over the standard LS estimators, with a minimal bias cost in the solution for $\boldsymbol{x}$ and β; for more details, see Kotsakis (2005). As in the previous viewpoint, a procedure for selecting a suitable value for γ^2 is still needed, which is not a trivial problem to solve.

A third alternative is finally to treat the factor γ^2 as an *unknown variance component* in the *type II* model and apply some of the well known variance component estimation (VCE) techniques for its optimal determination. In this case, if the CV matrix $\boldsymbol{C}$ in eq. (20b) describes realistically the statistical behaviour of the random errors in the observations, an exceptionally large value for the *a posteriori* estimate of γ^2 would indicate the existence of a possible common bias in the input data.

9. – CONCLUSIONS

The option of dual modelling in optimal estimation problems, adopting either a deterministic or a probabilistic description for (some or all of) the unknown quantities, holds a special place in the development of modern operational geodesy; see, e.g., Dermanis (1976), Dermanis and Rummel (2000), Sansò (1980), Moritz

(1980). The dilemma whether to use a purely algebraic or a purely stochastic representation for the geodetic parameters within the LS estimation framework has always stimulated interesting theoretical arguments among scientists (Moritz and Sansò, 1980; Dermanis and Sansò, 1993; Scales and Snieder, 1998), without however seriously hampering the geodetic practice over the years. In this paper, one more paradigm of the fruitful interaction between the two approaches has been rovided through the joint study of the models of *type I* and *type II*. It has been shown that these dual schemes for treating a common data bias in linear G-M models lead to algorithmically similar results, whose differences are controlled by the value of a single positive factor γ^2. The different interpretations of this factor have been emphasized, and some open problems that require additional research have been pointed out.

REFERENCES

A. Dermanis (1976), *Probabilistic and deterministic aspects of linear estimation in geodesy*. OSU Report Series, no. 244, Department of Geodetic Science, Ohio State University, Columbus.

A. Dermanis, F. Sansò (1993), *A dialogue on geodetic theory today*. IAG Section IV Bulletin, no. 1, pp. 6-9.

A. Dermanis, R. Rummel (2000), *Data analysis methods in geodesy*. In: A. Dermanis, A. Gruen and F. Sanso (eds.) Geomatic Methods for the Analysis of Data in Earth Sciences. Lecture Notes in Earth Sciences Series, vol. 95, pp. 17-92, Springer Verlag, Berlin Heidelberg.

D.A. Harville (1997), *Matrix algebra from a statistician's perspective*. Springer Verlag, Berlin Heidelberg.

K-R Koch (1987), *Parameter estimation and hypothesis testing in linear models*. Springer Verlag, Berlin Heidelberg.

C. Kotsakis (2003), *A new type of biased estimators for linear models*. Journal of Geodesy, submitted.

H. Moritz (1980), *Advanced physical geodesy*. Herbert Wichmann Verlag, Karlsruhe.

H. Moritz, F. Sansò (1980), *A dialogue on collocation*. Bolletino di Geodesia e Scienze Affini, Vol. 39, no. 1, pp. 49-51.

C.R. Rao (1973), *Linear statistical inference and its applications*. 2nd edition, Wiley, New York.

F. Sansò (1980), *The minimum mean square estimation error principle in physical geodesy (stochastic and non-stochastic interpretation)*. Bolletino di Geodesia e Scienze Affini, Vol. 39, no. 2, pp. 111-129.

J.A. Scales, R. Snieder (1998), *What is noise?* Geophysics, Vol. 63, no. 4, pp. 1122-1124.