A Monte-Carlo technique for weight estimation in satellite geodesy

J. Kusche

Physical, Geometrical, and Space Geodesy, DEOS, Delft University of Technology, Thijsseweg 11, 2629 JA Delft, The Netherlands
e-mail: j.kusche@citg.tudelft.nl; Tel.: +31-15-278-2562; Fax: +31-15-278-3711

Abstract. Precise orbit determination, and satellite-geodetic applications such as gravity field modelling or satellite altimetry, rely on different observation types and groups that have to be processed in a common parameter estimation scheme. Naturally, the choice of the relative weights for these data sets as well as for added prior information is of importance for obtaining reliable estimates of the unknown parameters and their associated covariance matrices. If the observations are predominantly affected by random errors and systematic errors play a minor role, variance component models can be applied. However, most of the methods proposed so far for variance component estimation involve repeated inversion of large matrices, resulting in intensive computations and large storage requirements if more than a few hundred unknowns are to be determined. In addition, these matrices are not necessarily provided as standard output from common geodetic least-squares estimation software. Therefore, a method is proposed which is based on Monte-Carlo estimation of the redundancy contributions of disjunctive observation groups. The method can handle unknown variance components without the need for repeated inversion of matrices. It is computationally simple, numerically stable and easy to implement. Its application is demonstrated in an experiment concerning low-medium-degree gravity field recovery from simulated orbit perturbations of the GOCE mission, and compared in performance with Lerch's method of subset solutions.

Keywords: Weight estimation – Variance component estimation – Satellite geodesy – Gravity field modelling – Monte-Carlo method

1 Introduction

Precise orbit determination, and satellite-geodetic applications and techniques such as global and regional gravity field modelling or sea surface mapping by satellite altimetry, rely on many different observation types or groups that have to be processed in a common parameter estimation scheme. Global gravity field models, for example, are computed nowadays from multi-arc, multi-satellite, multi-station laser tracking data, altimetry, terrestrial gravimetry, GPS-based satellite-to-satellite tracking, and specific techniques such as intersatellite ranging with the Gravity Recovery And Climate Experiment (GRACE) spacecraft or satellite gravity gradiometry with the Gravity and Steady-State Ocean Circulation Explorer Mission (GOCE). Needless to say, the choice of the relative weights for these different observation groups as well as for those models which introduce additional prior information, e.g. on the gravity field, is of importance for obtaining reliable estimates of the unknown parameters. The reliability of estimated and/or calibrated covariance matrices of the unknown parameters also depends on the assumptions about the observation weights.

Therefore, a weight optimization process normally forms a part of the activities. This involves assessing the formal instrumental errors as well as carrying out empirical considerations and intercomparison with independent models and data sets. Optimization with respect to the weights given by formal errors is necessary due to the limited knowledge of the instrument noise characteristics, and due to the imperfections in the physical models involved in the estimation process. Down- or overweighting in global gravity field modelling or orbit determination can easily reach a factor of 10 or more (Marsh et al. 1988; Lerch 1991; Schwintzer et al. 1997; Kizilsu and Sahin 2000). The question as to whether re-weightings are appropriate is likely to appear in the course of the European Space Agency's (ESA's) GOCE satellite mission, although a lot of effort is currently put into the investigation of error models for the gradiometer instrument and the precise orbit determination (POD) process.

A method proposed by Lerch (1991) establishes the weighting scheme based on the behaviour of the com-

mon parameters – the spherical harmonic coefficients of the Earth's gravity field in his paper – and their error estimates when we derive subset solutions where data groups have been excluded. This method is approximative in the sense that it assumes that each data subset in question has only small influence on the overall solution in comparison to the full data set. Consequently, in a variance component model, this method gives unbiased estimates to the same degree of approximation. Lerch's method requires repeated evaluation of the trace of the solution covariance matrix and the subset solution covariance matrices, which certainly will pose a problem for application to GOCE data analysis given the number of unknowns and the size of these covariance matrices. Lerch's method is, however, known to be quite robust with respect to unmodelled systematic errors.

It should be added that an assessment of the proper weight of prior information can be viewed as Tikhonov regularization with a posteriori parameter choice. Within this approach the regularization or weighting parameter is to be derived from minimization of some statistically or mathematically motivated target function, which usually involves the least-squares (LS) residuals of the data. Triggered by the new gravity field missions, quite recently different techniques have been studied in this field [see Kusche and Klees (2002) and the references therein]. If more than one regularization or weighting parameter is to be determined, however, these methods lead to multi-dimensional optimization problems for which numerical solutions lack efficiency in computation or may not be feasible at all.

The weighting method presented in this paper is based on the estimation of variance components (VCE), which goes back to Helmert (1924). Rao (1973) developed the MINQUE technique, and Grafarend and d'Hone (1978) investigated estimation methods for analysing geodetic data. For further references, see e.g. Schaffrin (1983), Rao and Kleffe (1988), Koch (1990), or Grafarend and Schaffrin (1993). A recent survey on exact and simplified VCE approaches when observation groups are disjunctive is provided by Crocetto et al. (2000). In global gravity modelling, VCE has been proposed by Schwintzer (1990) for assessing the proper weight of prior information on the Earth's low-degree gravity field. Sahin et al. (1992) and Kizilsu and Sahin (2000) apply Helmert's VCE for precision analysis in satellite laser ranging. Recently, VCE has been applied by Zhang et al. (2001) for improving orbit determination for the ERS-2 satellite. Lucas and Dillinger (1998) applied MINQUE to very-long-baseline interferometry (VLBI) data analysis and found good agreement with the conventional weighting methods in this field. Apart from in their problem, only relatively few unknown parameters were involved in the applications as reported. Looking at the GOCE data analysis concepts, the main limitation of the VCE techniques proposed so far seems the costly and repeated computation of the redundancy contributions of the observation groups, involving the normal matrix contribution of the particular data set as well as the inverse of the weighted combined normal matrix. For large systems, like those encountered when solving for a high-resolution gravity field model, this is prohibitive. The inverse normal matrix will not necessarily be computed if iterative solvers are employed. Moreover, the matrices involved are generally not provided as standard output by common geodetic parameter adjustment software. Even the technique worked out by Lucas and Dillinger (1998), which reduces the size of the problem to the number of common parameters present in the adjustment problem, cannot be expected to be of great improvement if the number of these global parameters dominates the problem. Therefore we apply a variant described by Koch and Kusche (2002), which makes use of a stochastic trace estimation technique invented by Girard (1989) and Hutchinson (1990).

The method is re-structured and developed further in a Monte-Carlo sense, that is, on input for an arbitrary LS estimation software we use cyclically randomized versions of the original data set where for each individual observation group an artificial noise sequence has to be added in turn. On output, from a comparison of the residuals obtained with the original data and the randomized data, the new weights are estimated in an iterative sense. The method will be called Monte-Carlo VCE (MCVCE) in the remainder of this paper. Three sub-variants will be described which differ slightly in the artificial noise characteristics and the necessary matrix–vector operations. A Monte-Carlo approach appears particularly attractive since it seeks to extract statistical information on the inversion scheme by passing random input through the given algorithm, without requiring modifications of the code. This means that an existing software package for solving the data inversion problem can be used as a black box. A similar method has been proposed by Purser and Parrish (2000) in the context of variational assimilation of meteorological data.

The material is organized as follows. First, we briefly review the model setup for the estimation problem in satellite geodesy. Then, VCE and the particular algorithm we intend to use will be introduced. We will show how this algorithm can indeed be recast in a Monte-Carlo fashion when using a randomized trace estimation technique. Geodetic parameter estimation software can be imbedded in an iterative MCVCE process without internal modifications or computing non-standard output. Finally, a simulation of low-/ medium-degree gravity field recovery from the GPS orbit determination for GOCE is considered. For comparison, we include results obtained with a randomized version of Lerch's method for weight estimation. A discussion closes the article.

2 Data combination in satellite geodesy

The linear observation model which will be adopted throughout this contribution reads

$$\mathbf{X}\boldsymbol{\beta} = \mathbf{y} + \mathbf{e} \qquad E(\mathbf{e}) = \mathbf{0} \qquad D(\mathbf{e}) = \boldsymbol{\Sigma} \tag{1}$$

with $n \times u$ design matrix $\mathbf{X}$, $u \times 1$ vector $\boldsymbol{\beta}$ of unknowns, $n \times 1$ vector $\mathbf{y}$ of observations, and $n \times 1$ vector $\mathbf{e}$ of

stochastic observation errors. We assume that the observation vector consists of p disjunctive groups of observations $\mathbf{y}_i$, which may be internally correlated but are uncorrelated with respect to each other. Moreover, we assume that the positive definite covariance matrices $\boldsymbol{\Sigma}_i = \sigma_i^2 \mathbf{P}_i^{-1}$ of these observation groups are only given up to unknown scaling factors, the p variance components σ_i^2. Let n_i be the size of $\mathbf{y}_i$. Then, Eq. (1) is completed by

$$\boldsymbol{\Sigma} = \sum_{i=1}^{p} \sigma_i^2 \mathbf{V}_i = \begin{pmatrix} \sigma_1^2 \mathbf{P}_1^{-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_2^2 \mathbf{P}_2^{-1} & \mathbf{0} & \mathbf{0} \\ \vdots & & & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_p^2 \mathbf{P}_p^{-1} \end{pmatrix} \quad (2)$$

where the $n_i \times n_i$ weight matrices $\mathbf{P}_i$ are given, and

$$\mathbf{V}_1 = \begin{pmatrix} \mathbf{P}_1^{-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & & & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$$

$$\mathbf{V}_2 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_2^{-1} & \mathbf{0} & \mathbf{0} \\ \vdots & & & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \cdots \mathbf{V}_p = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & & & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{P}_p^{-1} \end{pmatrix} \quad (3)$$

Disjunctive observation groups may, for example, be given by the tracking data collected from a single ground station, by orbit perturbations observed for a single satellite, or by all data produced from a single observation technique. The unknown parameters are usually separated into common or global parameters $\boldsymbol{\beta}_c$ that are present in all observation equations, and the arc or local parameters $\boldsymbol{\beta}_{i,a}$ which are unique to a single observation group $\mathbf{y}_i$. This means that all unknown parameters of the problem of Eqs. (1) and (2) may be collected as

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_c \\ \boldsymbol{\beta}_{1,a} \\ \vdots \\ \boldsymbol{\beta}_{p,a} \end{pmatrix} \qquad \boldsymbol{\sigma} = \begin{pmatrix} \sigma_1^2 \\ \vdots \\ \sigma_p^2 \end{pmatrix} \quad (4)$$

We may then write the original linear model as a sequence of smaller problems, which are linked to each other by the common parameters

$$\mathbf{X}_{i,c} \boldsymbol{\beta}_c + \mathbf{X}_{i,a} \boldsymbol{\beta}_{i,a} = \mathbf{y}_i + \mathbf{e}_i$$

$$D(\mathbf{e}_i) = \sigma_i^2 \mathbf{P}_i^{-1} \qquad i \in \{1, \ldots, p\} \quad (5)$$

Throughout what follows we will often make use of the abbreviated form

$$\tilde{\mathbf{X}}_i \boldsymbol{\beta} = \mathbf{y}_i + \mathbf{e}_i \quad i \in \{1, \ldots, p\} \quad (6)$$

with $\tilde{\mathbf{X}}_i = (\mathbf{X}_{i,c}\ \mathbf{0} \cdots \mathbf{0}\ \mathbf{X}_{i,a}\ \mathbf{0} \cdots)$. Obviously, in Eq. (1) it holds that

$$\mathbf{X} = \begin{pmatrix} \tilde{\mathbf{X}}_1 \\ \vdots \\ \tilde{\mathbf{X}}_p \end{pmatrix} = \begin{pmatrix} \mathbf{X}_{1,c} & \mathbf{X}_{1,a} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & & & & \vdots \\ \mathbf{X}_{p,c} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{X}_{p,a} \end{pmatrix}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_c \\ \boldsymbol{\beta}_{1,a} \\ \vdots \\ \boldsymbol{\beta}_{p,a} \end{pmatrix} \qquad \mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_p \end{pmatrix} \quad (7)$$

We assume here that no prior information has to be added, and that the method of partitioning will be applied when forming the normal equations, according to the separation into common and arc parameters (e.g. Reigber 1989). From the first assumption we do not loose any generality since adding prior information may be formally replaced by introducing additional pseudo-observations, with regularization parameters substituted by possibly unknown variance components. The LS estimate $\hat{\boldsymbol{\beta}}_c$ for the common parameters then follows from solving the accumulated normal equations

$$\left(\sum_{i=1}^{p} \frac{1}{\sigma_i^2} \bar{\mathbf{N}}_{i,cc} \right) \hat{\boldsymbol{\beta}}_c = \sum_{i=1}^{p} \frac{1}{\sigma_i^2} \bar{\mathbf{b}}_{i,c} \quad (8)$$

where the individual contributions, reduced by the influence of the arc parameters, are

$$\bar{\mathbf{N}}_{i,cc} = \mathbf{X}_{i,c}' \mathbf{P}_i \mathbf{X}_{i,c} - \mathbf{X}_{i,c}' \mathbf{P}_i \mathbf{X}_{i,a} (\mathbf{X}_{i,a}' \mathbf{P}_i \mathbf{X}_{i,a})^{-1} \mathbf{X}_{i,a}' \mathbf{P}_i \mathbf{X}_{i,c} \quad (9)$$

$$\bar{\mathbf{b}}_{i,c} = \mathbf{X}_{i,c}' \mathbf{P}_i \mathbf{y}_i - \mathbf{X}_{i,c}' \mathbf{P}_i \mathbf{X}_{i,a} (\mathbf{X}_{i,a}' \mathbf{P}_i \mathbf{X}_{i,a})^{-1} \mathbf{X}_{i,a}' \mathbf{P}_i \mathbf{y}_i \quad (10)$$

The covariance matrix of the common parameters is given by

$$D(\hat{\boldsymbol{\beta}}_c) = \left(\sum_{i=1}^{p} \frac{1}{\sigma_i^2} \bar{\mathbf{N}}_{i,cc} \right)^{-1} \quad (11)$$

and arc parameters $\hat{\boldsymbol{\beta}}_{i,a}$ then follow from

$$\frac{1}{\sigma_i^2} (\mathbf{X}_{i,a}' \mathbf{P}_i \mathbf{X}_{i,a}) \hat{\boldsymbol{\beta}}_{i,a} = \frac{1}{\sigma_i^2} (\mathbf{X}_{i,a}' \mathbf{P}_i \mathbf{y}_i - \mathbf{X}_{i,a}' \mathbf{P}_i \mathbf{X}_{i,c} \hat{\boldsymbol{\beta}}_c)$$

$$= \frac{1}{\sigma_i^2} \mathbf{X}_{i,a}' \mathbf{P}_i (\mathbf{y}_i - \mathbf{X}_{i,c} \hat{\boldsymbol{\beta}}_c) \quad (12)$$

The same can be written, if no partitioning and reduction is applied at all, as

$$\mathbf{N} \hat{\boldsymbol{\beta}} = \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{y} = \mathbf{b} \quad (13)$$

or

$$
\begin{pmatrix}
\left(\sum_i^p \frac{1}{\sigma_i^2}\mathbf{X}'_{i,c}\mathbf{P}_i\mathbf{X}_{i,c}\right) & \frac{1}{\sigma_1^2}\mathbf{X}'_{1,c}\mathbf{P}_1\mathbf{X}_{1,a} & \frac{1}{\sigma_2^2}\mathbf{X}'_{2,c}\mathbf{P}_2\mathbf{X}_{2,a} & \cdots & \frac{1}{\sigma_p^2}\mathbf{X}'_{p,c}\mathbf{P}_p\mathbf{X}_{p,a} \\
\frac{1}{\sigma_1^2}\mathbf{X}'_{1,a}\mathbf{P}_1\mathbf{X}_{1,c} & \frac{1}{\sigma_1^2}\mathbf{X}'_{1,a}\mathbf{P}_1\mathbf{X}_{1,a} & \mathbf{0} & \cdots & \mathbf{0} \\
\frac{1}{\sigma_2^2}\mathbf{X}'_{2,a}\mathbf{P}_2\mathbf{X}_{2,c} & \mathbf{0} & \frac{1}{\sigma_2^2}\mathbf{X}'_{2,a}\mathbf{P}_2\mathbf{X}_{2,a} & \cdots & \mathbf{0} \\
\vdots & & & & \vdots \\
\frac{1}{\sigma_p^2}\mathbf{X}'_{p,a}\mathbf{P}_p\mathbf{X}_{p,c} & \mathbf{0} & \mathbf{0} & \cdots & \frac{1}{\sigma_p^2}\mathbf{X}'_{p,a}\mathbf{P}_p\mathbf{X}_{p,a}
\end{pmatrix}
\hat{\boldsymbol{\beta}} =
\begin{pmatrix}
\sum_i^p \frac{1}{\sigma_i^2}\mathbf{X}'_{i,c}\mathbf{P}_i\mathbf{y}_i \\
\frac{1}{\sigma_1^2}\mathbf{X}'_{1,a}\mathbf{P}_1\mathbf{y}_1 \\
\frac{1}{\sigma_2^2}\mathbf{X}'_{2,a}\mathbf{P}_2\mathbf{y}_2 \\
\vdots \\
\frac{1}{\sigma_p^2}\mathbf{X}'_{p,a}\mathbf{P}_p\mathbf{y}_p
\end{pmatrix}
\tag{14}
$$

The normal matrix shown in Eq. (14) is said to be block-diagonal bordered due to its special structure.

3 Estimation of weights

3.1 Variance component estimation (VCE) in the Bayesian picture

The principle of VCE that we intend to use may be introduced briefly by way of the Bayesian point of view. We refer to Koch (1990) for more details and references. Let a vector $\boldsymbol{\lambda}$ of random parameters be realized with a prior probability density, $p(\boldsymbol{\lambda})$. For our application, we can think of collecting all unknown common and arc-dependent parameters, and all unknown variance components according to Eq. (4) in $\boldsymbol{\lambda}$, thus

$$
\boldsymbol{\lambda} = \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\sigma} \end{pmatrix}
\tag{15}
$$

Let the conditional density for the vector of observations $\mathbf{y}$ be $p(\mathbf{y}|\boldsymbol{\lambda})$. Then, according to elementary probability theory, the joint density for $\boldsymbol{\lambda}$ and $\mathbf{y}$ is $p(\mathbf{y}, \boldsymbol{\lambda}) = p(\mathbf{y}|\boldsymbol{\lambda})\, p(\boldsymbol{\lambda})$. Equally, we may express the joint density as the product of the conditional density $p(\boldsymbol{\lambda}|\mathbf{y})$ and the unconditional density $p(\mathbf{y})$. Combining these, we obtain Bayes' theorem

$$
p(\boldsymbol{\lambda}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\lambda})p(\boldsymbol{\lambda})}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\boldsymbol{\lambda})p(\boldsymbol{\lambda})}{\int p(\mathbf{y}|\boldsymbol{\lambda})p(\boldsymbol{\lambda})\mathrm{d}\boldsymbol{\lambda}} \propto p(\mathbf{y}|\boldsymbol{\lambda})p(\boldsymbol{\lambda})
\tag{16}
$$

Here $\propto$ means proportionality, and $\mathrm{d}\boldsymbol{\lambda}$ is the volume measure in the parameter space. We refer to $p(\boldsymbol{\lambda})$ as being the prior density, and $p(\boldsymbol{\lambda}|\mathbf{y})$ as being the posterior density, which has been modulated by the likelihood function $p(\mathbf{y}|\boldsymbol{\lambda})$. With the posterior density determined, the problem of statistical inference is solved: the estimate on $\boldsymbol{\lambda}$ can, for example, be obtained by maximization of the posterior density; it is then called the MAP estimator. Alternatively, we may derive the Bayes estimator $\int \boldsymbol{\lambda} p(\boldsymbol{\lambda}|\mathbf{y})\mathrm{d}\boldsymbol{\lambda}$, which can be shown to minimize the quadratic loss function $L = (\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})'D(\boldsymbol{\lambda})^{-1}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})$ (Koch 1990). Finally, the maximum-likelihood estimator, maximizes the likelihood function without considering the prior density. It can be viewed as an MAP estimator where implicitly the prior has been assumed as a constant for all parameters; for the $\boldsymbol{\beta}$ portion of the parameter vector this is called a non-informative prior.

Let us now assess the linear model given by Eqs. (1) and (2) with unknown common and arc parameters and unknown variance components. Assuming normal statistics for the observations, the likelihood function is then explicitly given by

$$
p(\mathbf{y}|\boldsymbol{\lambda}) = p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\sigma}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}}
$$
$$
\times \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right)
\tag{17}
$$

where $|\cdot|$ is the matrix determinant. Through $\boldsymbol{\Sigma}$ [Eq. (2)], the likelihood function depends on the variance components $\boldsymbol{\sigma}$. It is useful to refer to the negative logarithm of the probability densities, thereby converting multiplicative relationships into additive ones

$$
L_{\mathbf{y}|\boldsymbol{\beta},\boldsymbol{\sigma}} = -\log p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\sigma}) = \frac{1}{2}\log((2\pi)^n|\boldsymbol{\Sigma}|)
$$
$$
+ \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})
\tag{18}
$$

$$
L_{\boldsymbol{\beta},\boldsymbol{\sigma}} = -\log p(\boldsymbol{\beta}, \boldsymbol{\sigma}), \quad L_{\boldsymbol{\beta},\boldsymbol{\sigma}|\mathbf{y}} = -\log p(\boldsymbol{\beta}, \boldsymbol{\sigma}|\mathbf{y})
\tag{19}
$$

By integration over the $\boldsymbol{\beta}$ parameter space, the likelihood function can be expressed by a function of $\boldsymbol{\sigma}$ only

$$
p(\mathbf{y}|\boldsymbol{\sigma}) \propto \frac{1}{|\boldsymbol{\Sigma}|^{1/2}|\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}|^{1/2}}
$$
$$
\times \exp\left(-\frac{1}{2}\mathbf{y}'\boldsymbol{\Sigma}^{-1}(\mathbf{I} - \mathbf{X}\mathbf{N}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1})\mathbf{y}\right)
\tag{20}
$$

$$
L_{\mathbf{y}|\boldsymbol{\sigma}} = \frac{1}{2}\log(|\boldsymbol{\Sigma}|) + \frac{1}{2}\log(|\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}|)
$$
$$
+ \frac{1}{2}\mathbf{y}'\boldsymbol{\Sigma}^{-1}(\mathbf{I} - \mathbf{X}\mathbf{N}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1})\mathbf{y} + C
\tag{21}
$$

where C is just a constant (Koch 1990). Using the last result in a maximum-likelihood estimate (Koch 1986) leads to the algorithm used in the next section, with posterior densities for the variance components and their ratios as given in Koch and Kusche (2002). However, the main computational difficulty becomes obvious from the fact that derivatives of the likelihood function involve the derivatives of the log-determinant matrix function: let $\mathbf{A}$ be a regular symmetric matrix which is a function of σ_i^2, then $\partial \log(|\mathbf{A}|)/\partial\sigma_i^2 = \mathrm{trace}(\mathbf{A}^{-1}\partial\mathbf{A}/\partial\sigma_i^2)$.

3.2 The VCE algorithm

The algorithm described in this chapter goes back on Förstner (1979). It can be derived by the maximum-likelihood method from Eq. (20) if the observation groups are disjunctive, as shown in Koch (1986). If its iterations converge, this algorithm gives the same numerical results as the iterated MINQUE technique (see e.g. Rao and Kleffe 1988, Chap. 9.2). The non-negative estimates for the unknown variance components follow from

$$\hat{\sigma}_i^2 = \frac{\hat{\mathbf{e}}_i' \mathbf{P}_i \hat{\mathbf{e}}_i}{r_i}, \quad i \in \{1, \ldots, p\} \tag{22}$$

where the nominator is the weighted residual square sum for the ith observation group

$$\hat{\mathbf{e}}_i = \mathbf{X}_{i,c}\hat{\boldsymbol{\beta}}_c + \mathbf{X}_{i,a}\hat{\boldsymbol{\beta}}_{i,a} - \mathbf{y}_i = -\left(\mathbf{I} - \tilde{\mathbf{X}}_i \mathbf{N}^{-1} \tilde{\mathbf{X}}_i' \frac{1}{\sigma_i^2} \mathbf{P}_i\right)\mathbf{y}_i$$
$$+ \tilde{\mathbf{X}}_i \mathbf{N}^{-1} \left(\sum_{j \neq i}^{p} \tilde{\mathbf{X}}_j' \frac{1}{\sigma_j^2} \mathbf{P}_j \mathbf{y}_j\right) \tag{23}$$

A residual file is standard output of common LS estimation software. The denominator in Eq. (22) is given by the redundancy contribution of the particular observation group, also called the group redundancy number. This quantity involves the trace of the observation group influence matrix $\tilde{\mathbf{X}}_i' \mathbf{P}_i \tilde{\mathbf{X}}_i \mathbf{N}^{-1}$

$$r_i = -\text{trace}(\partial \hat{\mathbf{e}}_i / \partial \mathbf{y}_i) = n_i - \frac{1}{\sigma_i^2} \text{trace}(\tilde{\mathbf{X}}_i' \mathbf{P}_i \tilde{\mathbf{X}}_i \mathbf{N}^{-1}) \tag{24}$$

For numerical evaluation of the trace term it is more suitable to deal with symmetric matrices. Therefore, in the next section we will make use of the following formulation:

$$r_i = n_i - \frac{1}{\sigma_i^2} \text{trace}(\mathbf{G}_i' \tilde{\mathbf{X}}_i \mathbf{N}^{-1} \tilde{\mathbf{X}}_i' \mathbf{G}_i) \tag{25}$$

where $\mathbf{P}_i = \mathbf{G}_i \mathbf{G}_i'$. The partial redundancies sum up to the total redundancy of the LS problem

$$\sum_i^p r_i = r = n - u \tag{26}$$

The estimation of variance components generally leads to a coupled iterative process. The strategy is then as follows

1. Select start values for $\sigma_j^{2(0)}$, $j = 1 \ldots p$
   For $k = 0, 1, 2 \ldots$
2. Compute $\hat{\boldsymbol{\beta}}_c^{(k)}$, $\hat{\boldsymbol{\beta}}_{j,a}^{(k)}$, $\hat{\mathbf{e}}_j^{(k)}$, $j = 1 \ldots p$
3. Compute $r_j^{(k)}$, $j = 1 \ldots p$
4. Determine $\sigma_j^{2(k+1)}$, $j = 1 \ldots p$, and continue with step 2 until convergence is achieved.

Step 3, according to Eq. (24) or (25), poses a problem if the related matrices are not available from the estimation software, or if the dimension of the problem is so large that they cannot be computed in reasonable time.

Lucas and Dillinger (1998) in their treatment reduced the complexity of the problem by making explicit use of the block-diagonal bordered structure of the matrices involved in Eq. (24), and they were able to show that only matrices and vectors that appeared already in the reduced normal equations, Eqs. (9) and (10), showed up here. However, their method still requires the inverse of the normal matrix partition associated with the common parameters to be computed explicitly, which is one of the calculations we would like to avoid. Therefore, in the following we aim at obtaining a computationally cheap estimate of these quantities.

4 The MCVCE method for weight estimation

4.1. Randomized trace estimation for the observation group redundancies

Within the kth cycle of the iteration (i.e. given a set of approximate variance components) we have to compute the redundancy contributions $r_i^{(k)}$ for all observation groups $\mathbf{y}_i$, $i \in \{1 \ldots p\}$. This computation can be greatly simplified if we apply the idea of randomized trace estimation to Eq. (25), as proposed by Koch and Kusche (2002).

As is well known, for any symmetric positive definite matrix $\mathbf{T}$ and any random vector $\mathbf{z}$ with $E(\mathbf{z}) = \mathbf{0}$, $D(\mathbf{z}) = \mathbf{I}$, regardless the type of its probability density function, holds

$$E(\mathbf{z}'\mathbf{T}\mathbf{z}) = \text{trace}(\mathbf{T}) \tag{27}$$

Estimation of the trace of $\mathbf{T}$ holds in that q realizations $\bar{\mathbf{z}}^j$ of $\mathbf{z}$ are being obtained from a random number generator. The trace of $\mathbf{T}$ is then replaced by the unbiased estimate $\frac{1}{q}\sum_{j=1}^{q} \bar{\mathbf{z}}^{j'}\mathbf{T}\bar{\mathbf{z}}^j$. If, in addition, $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ follows from the multivariate standard normal distribution, the variance of the trace estimate can be bounded by (Girard 1989)

$$V(\mathbf{z}'\mathbf{T}\mathbf{z}) = 2\,\text{trace}(\mathbf{T}^2) < 2\,\text{trace}(\mathbf{T}) \tag{28}$$

The inequality holds if the matrix $\mathbf{T}$ is a smoother matrix, that is, with all its eigenvalues between 0 and 1. The observation group influence matrices $\tilde{\mathbf{X}}_i' \mathbf{P}_i \tilde{\mathbf{X}}_i \mathbf{N}^{-1}$ indeed share this property. However, it has been shown by Hutchinson (1990) that the estimator assumes minimum possible variance, if $\mathbf{z} \sim \mathcal{U}(\mathbf{0}, \mathbf{I})$, where $\mathcal{U}$ is the multivariate discrete probability function, so that each component of the sample vector takes the value -1 or 1 with equal probability $1/2$. The variance of this particular trace estimate can be bounded by (Hutchinson 1990).

$$V(\mathbf{z}'\mathbf{T}\mathbf{z}) = 2\sum_{k \neq l}^{n}(T_{kl})^2 < 2\,\text{trace}(\mathbf{T}) - \frac{2}{n}(\text{trace}(\mathbf{T}))^2 \tag{29}$$

where $n \times n$ is the dimension of $\mathbf{T}$. The variance of the trace estimate can be further reduced if $q > 1$ sample vectors are employed, clearly at the cost of increasing the computational load. There is no general prescription

for the choice of q, since $V(\mathbf{z}'\mathbf{Tz})$ and – what is more important for the variance component estimate of Eq. (22) – the relative variance $V(\mathbf{z}'\mathbf{Tz})/(\text{trace}(\mathbf{T}))^2$ depend on the matrix $\mathbf{T}$ according to Eqs. (28) and (29). Sharper bounds have been derived by Hutchinson (1990) but they are difficult to apply. In principle we could evaluate the bounds of Eq. (28) or (29) as soon as a first trace estimate is given, and conclude on the necessary number $q - 1$ of additional realizations to match a certain relative accuracy. But the bounds of Eqs. (28) and (29) turn out to be rather pessimistic. Instead, after carrying out some tests, we found it fully sufficient to work with only $q = 1$ realization of $\mathbf{z}$ for the GOCE–SST gravity field recovery experiment conducted in Sect. 5. We reported similar conclusions in Kusche and Klees (2002) in the context of regularization parameter choice for GOCE gradiometry.

In the following the application of trace estimation to VCE in the model of Eqs. (1) and (2) will be described. Three algorithms are given, with minor differences in the characteristics of the artificial random sequence and the necessary matrix–vector operations.

4.2 Monte-Carlo algorithm: first variant

A direct application of Eq. (27) on the computation of the redundancy contribution r_i according to Eq. (25) gives immediately

$$\hat{r}_i = n_i - \frac{1}{\sigma_i^2}\mathbf{z}_i'\mathbf{G}_i'\tilde{\mathbf{X}}_i\mathbf{N}^{-1}\tilde{\mathbf{X}}_i'\mathbf{G}_i\mathbf{z}_i \tag{30}$$

with random vector $\mathbf{z}_i$, $\dim(\mathbf{z}_i) = n_i$, $E(\mathbf{z}_i) = \mathbf{0}$, $D(\mathbf{z}_i) = \mathbf{I}$. Throughout the following we assume that applying a $q = 1$ sample of the random vector will meet our requirements in terms of accuracy of the trace estimate; otherwise, the mean of q estimates has to be taken. The evaluation of Eq. (30) for all p observation groups can then be formulated as a two-step procedure.

1. Solve

$$\mathbf{Nq}_i = \tilde{\mathbf{X}}_i'\mathbf{G}_i\mathbf{z}_i \tag{31}$$

for the 'random parameter solutions' $\mathbf{q}_i$, $i = \{1\ldots p\}$.
2. Compute

$$\hat{r}_i = n_i - \frac{1}{\sigma_i^2}\mathbf{z}_i'\mathbf{G}_i'\tilde{\mathbf{X}}_i\mathbf{q}_i \tag{32}$$

instead of the exact r_i.

Step 1 means nothing other than solving the original normal equations [Eq. 14] with the right-hand side replaced by

$$\tilde{\mathbf{X}}_i'\mathbf{G}_i\mathbf{z}_i = \begin{pmatrix} \mathbf{X}_{i,c}'\mathbf{G}_i\mathbf{z}_i \\ \mathbf{0} \\ \vdots \\ \mathbf{X}_{i,a}'\mathbf{G}_i\mathbf{z}_i \\ \mathbf{0} \\ \vdots \end{pmatrix} \tag{33}$$

i.e. with random observations for observation group i, scaled by $\mathbf{G}_i$, and zeros for all other observations. Re-assembling the right-hand side is therefore necessary only for the ith observation group. In other words, at step 1 we pass scaled standard-distributed random noise for the observation group in question and zeros for all other data through our LS parameter estimation algorithm, and at step 2 the redundancy estimate follows from the solved-for parameters and the (un-reduced) right-hand-side vectors which have been built up in the course of the first step. The numerical benefit of this technique depends on the way we solve the normal equations – it will be great for Cholesky solvers since we have to solve for just one additional right-hand-side vector and might be moderate for conjugate gradient solvers – but anyway it will clearly outperform the exact evaluation of the matrix trace and avoids the explicit formulation of the full normal matrix.

More explicitly, making use of the reduced normal equations [Eqs. (8) and (12)] we solve in addition

$$\left(\sum_{i=1}^{p}\frac{1}{\sigma_i^2}\bar{\mathbf{N}}_{i,cc}\right)\mathbf{q}_{i,c} = \mathbf{X}_{i,c}'\mathbf{G}_i\mathbf{z}_i - \mathbf{X}_{i,c}'\mathbf{P}_i\mathbf{X}_{i,a}$$

$$\times (\mathbf{X}_{i,a}'\mathbf{P}_i\mathbf{X}_{i,a})^{-1}\mathbf{X}_{i,a}'\mathbf{G}_i\mathbf{z}_i \tag{34}$$

$$\frac{1}{\sigma_i^2}(\mathbf{X}_{i,a}'\mathbf{P}_i\mathbf{X}_{i,a})\mathbf{q}_{i,a} = \mathbf{X}_{i,a}'\mathbf{G}_i\mathbf{z}_i - \mathbf{X}_{i,a}'\mathbf{P}_i\mathbf{X}_{i,c}\hat{\mathbf{q}}_{i,c}$$

$$= \mathbf{X}_{i,a}'\mathbf{G}_i(\mathbf{z}_i - \mathbf{G}_i'\mathbf{X}_{i,c}\hat{\mathbf{q}}_{i,c}) \tag{35}$$

After that we have the estimates

$$\hat{r}_i = n_i - \frac{1}{\sigma_i^2}\mathbf{z}_i'\mathbf{G}_i'(\mathbf{X}_{i,c}\hat{\mathbf{q}}_{i,c} + \mathbf{X}_{i,a}\hat{\mathbf{q}}_{i,a})$$

$$= n_i - \frac{1}{\sigma_i^2}(\hat{\mathbf{q}}_{i,c}'\mathbf{X}_{i,c}'\mathbf{G}_i\mathbf{z}_i + \hat{\mathbf{q}}_{i,a}'\mathbf{X}_{i,a}'\mathbf{G}_i\mathbf{z}_i) \tag{36}$$

The first expression in Eq. (36) convolves the input random sequence with a sequence of synthesized observations from the estimated random parameters, thus requiring the application of the design matrix on a parameter vector (synthesis step). The second expression relies on a product of the right-hand-side with the estimated parameter vector. The general principle is outlined in Fig. 1.

4.3 Monte-Carlo algorithm: second variant

This variant differs from the previous one in the scaling of the artificial random sequence. Instead of applying Eq. (30) we can form alternatively

$$\hat{r}_i = n_i - \frac{1}{\sigma_i^4}\mathbf{w}_i'\mathbf{P}_i\tilde{\mathbf{X}}_i\mathbf{N}^{-1}\tilde{\mathbf{X}}_i'\mathbf{P}_i\mathbf{w}_i \tag{37}$$

with random vector $\mathbf{w}_i$, $\dim(\mathbf{w}_i) = n_i$, $E(\mathbf{w}_i) = \mathbf{0}$, $D(\mathbf{w}_i) = \sigma_i^2\mathbf{P}_i^{-1}$. The difference is that here the random
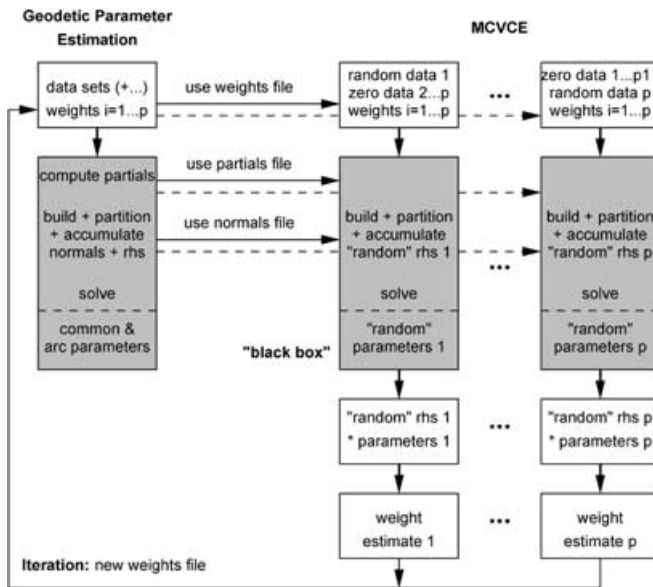
Fig. 1. Use of 'black-box' least-squares estimation software modules in MCVCE; first variant

sample vector $\mathbf{w}_i$ should possess exactly the same statistical first and second moments as are known or at least assumed for the error vector $\mathbf{e}_i$. If, for instance, an algorithm exists that simulates the observation errors given a certain power spectral density model, the same algorithm might be used for generating the random realization $\mathbf{w}_i$. In comparison with the previous variant, obviously $\mathbf{w}_i$ and $\mathbf{z}_i$ are related by $\mathbf{G}'_i\mathbf{w}_i = \sigma_i\mathbf{z}_i$. The first step of Eq. (37) then involves solving the normal equations of Eq. (14) for the unknowns $\mathbf{p}_i$ with the right-hand side

$$\tilde{\mathbf{X}}'_i\mathbf{P}_i\mathbf{w}_i = \begin{pmatrix} \mathbf{X}'_{i,c}\mathbf{P}_i\mathbf{w}_i \\ \mathbf{0} \\ \vdots \\ \mathbf{X}'_{i,a}\mathbf{P}_i\mathbf{w}_i \\ \mathbf{0} \\ \vdots \end{pmatrix} \tag{38}$$

i.e. with unscaled random observations for observation group i, and zeros for all other observations. Making use of the reduced normal equations, this is

$$\left(\sum_{i=1}^{p} \frac{1}{\sigma_i^2} \bar{\mathbf{N}}_{i,cc}\right)\mathbf{p}_{i,c}$$

$$= \mathbf{X}'_{i,c}\mathbf{P}_i\mathbf{w}_i - \mathbf{X}'_{i,c}\mathbf{P}_i\mathbf{X}_{i,a}(\mathbf{X}'_{i,a}\mathbf{P}_i\mathbf{X}_{i,a})^{-1}\mathbf{X}'_{i,a}\mathbf{P}_i\mathbf{w}_i$$

$$= \mathbf{X}'_{i,c}\mathbf{P}_i\mathbf{Q}_{i,a}\mathbf{w}_i \tag{39}$$

$$\frac{1}{\sigma_i^2}(\mathbf{X}'_{i,a}\mathbf{P}_i\mathbf{X}_{i,a})\mathbf{p}_{i,a} = \mathbf{X}'_{i,a}\mathbf{P}_i\mathbf{w}_i - \mathbf{X}'_{i,a}\mathbf{P}_i\mathbf{X}_{i,c}\hat{\mathbf{p}}_{i,c}$$

$$= \mathbf{X}'_{i,a}\mathbf{P}_i(\mathbf{w}_i - \mathbf{X}_{i,c}\hat{\mathbf{p}}_{i,c}) \tag{40}$$

and finally

$$\hat{r}_i = n_i - \frac{1}{\sigma_i^4}\mathbf{w}'_i\mathbf{P}_i(\mathbf{X}_{i,c}\hat{\mathbf{p}}_{i,c} + \mathbf{X}_{i,a}\hat{\mathbf{p}}_{i,a})$$

$$= n_i - \frac{1}{\sigma_i^4}(\hat{\mathbf{p}}'_{i,c}\mathbf{X}'_{i,c}\mathbf{P}_i\mathbf{w}_i + \hat{\mathbf{p}}'_{i,a}\mathbf{X}'_{i,a}\mathbf{P}_i\mathbf{w}_i) \tag{41}$$

Concerning Eq. (41), the same algorithmic structure as mentioned for Eq. (36) applies; thus Fig. 1 remains valid for this variant as well.

4.4. Monte-Carlo algorithm: third variant

Since the LS estimator for the common parameters and the arc parameters is a linear estimator, we might equally well work simultaneously with the complete original data set $\mathbf{y}$, supplemented by p artificially perturbed data vectors $\mathbf{y}_i^*$

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_i \\ \vdots \\ \mathbf{y}_p \end{pmatrix} \qquad \mathbf{y}_i^* = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_i + \mathbf{w}_i \\ \vdots \\ \mathbf{y}_p \end{pmatrix} \tag{42}$$

with $\mathbf{w}_i$, $\dim(\mathbf{w}_i) = n_i$, $E(\mathbf{w}_i) = \mathbf{0}$, $D(\mathbf{w}_i) = \sigma_i^2\mathbf{P}_i^{-1}$. In this case we obtain by accumulation of the normal equations, partitioning and solving the parameter sets $\hat{\boldsymbol{\beta}}_c$, $\hat{\boldsymbol{\beta}}_{i,a}$, $\hat{\boldsymbol{\beta}}_{i,c}^*$, $\hat{\boldsymbol{\beta}}_{i,a}^*$. These are related to $\hat{\mathbf{p}}_{i,c}$ and $\hat{\mathbf{p}}_{i,a}$ as follows:

$$\hat{\mathbf{p}}_{i,c} = \hat{\boldsymbol{\beta}}_{i,c}^* - \hat{\boldsymbol{\beta}}_c \qquad \hat{\mathbf{p}}_{i,a} = \hat{\boldsymbol{\beta}}_{i,a}^* - \hat{\boldsymbol{\beta}}_{i,a} \tag{43}$$

The estimate for the redundancy contributions is then

$$\hat{r}_i = n_i - \frac{1}{\sigma_i^4}\mathbf{w}'_i\mathbf{P}_i\left(\mathbf{X}_{i,c}(\hat{\boldsymbol{\beta}}_{i,c}^* - \hat{\boldsymbol{\beta}}_c) + \mathbf{X}_{i,a}(\hat{\boldsymbol{\beta}}_{i,a}^* - \hat{\boldsymbol{\beta}}_{i,a})\right)$$

$$= n_i - \frac{1}{\sigma_i^4}\left((\hat{\boldsymbol{\beta}}_{i,c}^* - \hat{\boldsymbol{\beta}}_c)'\mathbf{X}'_{i,c}\mathbf{P}_i\mathbf{w}_i + (\hat{\boldsymbol{\beta}}_{i,a}^* - \hat{\boldsymbol{\beta}}_{i,a})'\mathbf{X}'_{i,a}\mathbf{P}_i\mathbf{w}_i\right) \tag{44}$$

With the LS residuals $\hat{\mathbf{e}}_i = \mathbf{X}_{i,c}\hat{\boldsymbol{\beta}}_c + \mathbf{X}_{i,a}\hat{\boldsymbol{\beta}}_{i,a} - \mathbf{y}_i$ and the corresponding residuals $\hat{\mathbf{e}}_i^* = \mathbf{X}_{i,c}\hat{\boldsymbol{\beta}}_c^* + \mathbf{X}_{i,a}\hat{\boldsymbol{\beta}}_{i,a}^* - \mathbf{y}_i$ this can be written as

$$\hat{r}_i = n_i - \frac{1}{\sigma_i^4}\mathbf{w}'_i\mathbf{P}_i(\hat{\mathbf{e}}_i^* - \hat{\mathbf{e}}_i) \tag{45}$$

This last expression for the redundancy contributions needs only the original random sequences $\mathbf{w}_i$ which were used when passing perturbed data through the estimation algorithm, and the observation residuals obtained with the original data as well as with the perturbed data. Even more so than with the previously derived expressions, Eqs. (36) and (41), the right-hand-side and the solved-for parameter vectors may remain hidden in the LS algorithm, although of course implicitly needed in the residuals synthesis. It can be said that this variant is 'end-to-end' directed in the sense that the user does not

even need to know exactly what kind of parameters are represented in the functional model. An outline of the procedure is shown in Fig. 2.

5 Numerical experiments: gravity field determination from simulated orbit perturbations of the GOCE mission

5.1 Description

The objective of the ESA's GOCE satellite mission is the determination of a model of the Earth's static gravity field with an accuracy of better than 1–2 mGal in terms of gravity anomalies and 1 cm in geoid heights, at a resolution of 100 km (ESA 1999). The spacecraft will be launched in 2006 as the first satellite in the framework of the ESA's Earth Explorer Mission programme. GOCE will be equipped with a gravity gradiometer and a geodetic GPS receiver, providing observations of the gravity gradient tensor at satellite altitude (about 250 km) and satellite-to-satellite observations between GOCE and the GPS satellites.

The GPS receiver plays a multiple role; it enables a high-precision orbit determination, allowing the gradiometer data to be processed without estimation of orbit errors, and it serves for recovery of the long-to-medium part of the gravity field. The POD is expected to be at the centimetre accuracy level (Visser and van den IJssel 2000).

In this study a simulated 10-day GOCE orbit solution has been used as pseudo observation, i.e. Cartesian x, y and z coordinates in an Earth-centered quasi-inertial reference frame, split up into 10 data sets (orbital arcs) of 1 day each. In an adjustment for gravity parameters $\boldsymbol{\beta}_c = (\delta c_{02}, \delta c_{12}, \delta s_{12}, \ldots)^T$ (i.e. the difference of the harmonic coefficients between the 'true' model OSU91a and the adopted initial model JGM-3) as well as for 10 sets

of state–vector epoch parameters $\boldsymbol{\beta}_{i,a} = \mathbf{x}_i = (x, y, z, v_x, v_y, v_z)^T$, $i = 1 \ldots 10$, the partial derivatives of the satellite positions with respect to the unknowns have been obtained by numerical integration of the variational equations. The OSU91a model has been used complete up to degree 50 for the orbit generation. It should be noted that the parameter estimation process actually has to be iterated due to linearization errors; however, these errors are systematic but small for degree 50 and no re-computation of the partials has been applied. Also, no regularization has been imposed on the estimation process. Of course, these considerations should be revised when higher resolutions are taken into account. Simulated noise-free data and partials files, computed from the GEODYN II orbit determination software (McCarthy et al. 1993), have kindly been provided by P. Visser. A detailed description is given in Visser et al. (2001), where this data set has been used in a comparative study on the quality of different recovery methods. At the time of writing, it is indeed expected that the variational approach will be followed in the derivation of the official ESA level-2 gravity field product from the GPS POD.

5.2 Results

The pseudo-observed coordinates x, y and z have been assumed as uncorrelated in time as well as with respect to each other. They have been analysed in the model of Eqs. (1) and (2) with $\mathbf{P}_i = \mathbf{I}$, by assigning a variance component σ_i^2 or a weight $\omega_i = 1/\sigma_i^2$ to each orbital arc data set. For all experiments we have assumed equal weights $\omega_i = 1$ for the first iteration. The normal system has been solved by the method of partitioning [Eqs. (8) and (12)]. Results are given for the estimated common gravity parameters $\hat{\boldsymbol{\beta}}_c$, expressed in terms of geoid height errors, and for the estimated variance components. The first variant of the MCVCE method has been applied as described in previous sections. The random number generator IGNUIN of the statistics package RANLIB has been used for generation of the artificial noise sequences. The overall procedure, of course, forms a rather simplified picture of the long-to-medium-degree gravity field recovery for GOCE, but it serves perfectly as a test bed for the Monte-Carlo approach.

For comparison, we will give results obtained with a randomized version of Lerch's method (Lerch 1991). This method aims at estimating weights that balance the difference of the common parameter solution and certain subset solutions, i.e. where an observation group has been excluded, with the associated covariance matrix. It requires, within each iteration, for each observation group in question, the trace of the difference of two covariance matrices for the common parameters to be evaluated. Again, we have used randomized trace estimation according to Eq. (27) for this purpose. For both methods, all trace evaluations have been estimated using one single realization of a random vector. To validate this assumption, an additional experiment has been performed: Fig. 3 shows a histogram for the trace
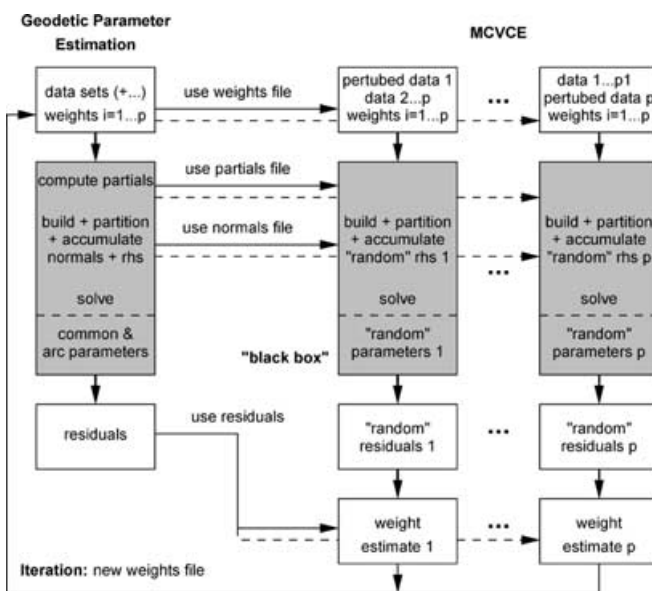

Fig. 2. Use of 'black-box' least-squares estimation software modules in MCVCE; third variant
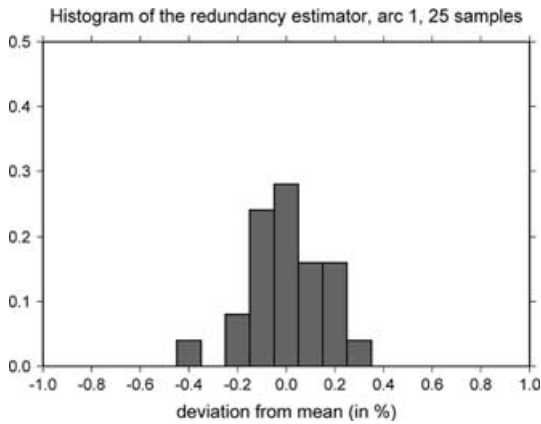
Histogram of the redundancy estimator, arc 1, 25 samples

Fig. 3. Histogram of the estimation of the first arc's redundancy contribution $\hat{r}_1$. 25 realizations have been used

estimates of $\hat{r}_1$, from 25 realizations of a random vector. We have found that all samples for $\hat{r}_1$ are within 0.5% deviation, which, according to Eq. (22), also holds for the variance component estimate samples. We conclude that for our problem the accuracy obtained from a single sample will be sufficient. In addition, the total redundancy, Eq. (26), has been used as a condition for 'calibrating' the estimated group redundancy numbers, simply by adjusting for the error $\sum_i \hat{r}_i - (n - u) \neq 0$.

In what follows, six test cases have been investigated. For a fair comparison, we have included one test with error-free data, two experiments where purely stochastic noise is involved, and three tests where significant systematic effects are modulated on top of the stochastic noise.

First, error-free pseudo-observations x, y and z have been used. For this case, which is also shown in Visser et al. (2001), geoid errors are generally negligible with a maximum of 0.4 cm and an RMS of 0.4 mm if no weight optimization takes place at all. (See Table 1, which gives maximum, RMS and average geoid errors as well as the range of the estimated variance components.)

These small errors are due to numerical round-off and linearization effects, which are non-stochastic in nature and depend on the quality of the underlying gravity model used for integration of the variational equations. Nevertheless, even these results improve

significantly when we apply MCVCE or Lerch's method for weight optimization. Iterations of Lerch's randomized method converge faster in terms of geoid errors, but MCVCE gives a more reasonable estimate of the size of linearization effects. This behaviour can generally be expected since Lerch's method focuses exclusively on the common parameters, i.e. the gravity unknowns. When we tried to force convergence by running many iterations, unexpectedly Lerch's method started to diverge (although this was the only test case where we observed this behaviour). The reason might be that after some iterations the estimated weights become strongly unbalanced, and, as a consequence, the underlying assumption that no single data set has more than a moderate influence on the solution becomes violated. However, given the magnitude of geoid errors, results from this test case should not be over-estimated.

Second, all observations have been corrupted artificially by an uncorrelated random noise sequence of $\sigma_{x,y,z} = 1$ cm standard deviation. This is an ideal case since the assumed stochastic model is then correct from the very beginning. Consequently, MCVCE performs excellently and convergence is reached after only one iteration. Lerch's method, nevertheless, provides us with comparable results, and even the weight estimates are 90% correct. A summary is shown in Table 2.

In the third test case for two of the 10 orbital arcs, the noise level has been increased from $\sigma_{x,y,z} = 1$ cm to $\sigma_{x,y,z} = 4$ cm. Results are summarized in Table 3. The MCVCE does exactly what we want it to do: it converges quickly (as does Lerch's method) and provides us with very accurate estimates of the variance components. Two iterations of MCVCE appear to be sufficient. Although this test clearly oversimplifies the expected real situation, stochastic noise levels varying over time may very well be encountered when observation configurations, solar activity or environmental conditions change.

For the fourth experiment, we have again added to all observations the uncorrelated random noise sequence of $\sigma_{x,y,z} = 1$ cm. In addition, for two particular orbital arcs the x, y and z pseudo-observations have been perturbed by a frequency-dependent radial error of 5 cm at the orbit frequency, 1 cycle/revolution (CPR). The intention is to mimic in a very simplistic way an error introduced by the GPS POD for these observation groups. As in the first

Table 1. Results for the case of noise-free observations. Geoid errors are caused purely by linearization errors and numerical round-off

Case 1	Geoid errors (cm)			Variance components (cm) $\hat{\sigma}_i = \dfrac{1}{\hat{w}_i^{1/2}}$
	Max	RMS	Average	
Equal weights[a]	0.39	0.04	0.06	1
MCVCE, 1st iteration	0.41	0.04	0.06	$1.4 \cdot 10^{-3} \ldots 2.1 \cdot 10^{-3}$
MCVCE, 5th iteration	0.34	0.03	0.05	$0.4 \cdot 10^{-3} \ldots 3.7 \cdot 10^{-3}$
MCVCE, 9th iteration	0.26	0.03	0.04	$0.4 \cdot 10^{-3} \ldots 5.0 \cdot 10^{-3}$
MCVCE at convergence	0.21	0.02	0.03	$0.4 \cdot 10^{-3} \ldots 4.8 \cdot 10^{-3}$
Lerch's method, 1st iteration	0.26	0.03	0.04	$0.8 \cdot 10^{-3} \ldots 4.1 \cdot 10^{-3}$
Lerch's method, 5th iteration	0.19	0.02	0.03	$0.7 \cdot 10^{-3} \ldots 0.02$
Lerch's method, 9th iteration	0.20	0.03	0.04	$0.7 \cdot 10^{-3} \ldots 0.16$
Lerch's method at convergence[b]	–	–	–	–

[a] This is the case shown in Table 4 of Visser et al. (2001).
[b] No convergence achieved!

Table 2. Results for the case of noisy observations. Uncorrelated random errors of $\sigma_{x,y,z} = 1$ cm standard deviation have been added to the original orbit pseudo-observations

Case 2	Geoid errors (cm)			Variance components (cm) $\hat{\sigma}_i = \dfrac{1}{\hat{w}_i^{1/2}}$
	Max	RMS	Average	
Equal weights	90.45	13.68	18.91	1
MCVCE, 1st iteration	90.45	13.68	18.91	0.99 ... 1.01
MCVCE at convergence	90.45	13.69	18.91	0.99 ... 1.01
Lerch's method, 1st iteration	92.87	13.67	18.92	0.91 ... 1.10
Lerch's method at convergence	93.25	13.68	18.93	0.89 ... 1.12

Table 3. Results for the case of noisy observations. Uncorrelated random errors of $\sigma_{x,y,z} = 1$ cm (arcs 1 ... 8) and $\sigma_{x,y,z} = 4$ cm (arcs 9 and 10) standard deviation have been added to the original orbit pseudo-observations

Case 3	Geoid errors (cm)			Variance components (cm) $\hat{\sigma}_i = \dfrac{1}{\hat{w}_i^{1/2}}$
	Max	RMS	Average	
Equal weights	238.54	27.42	38.00	1
MCVCE, 1st iteration	99.67	15.22	21.15	1.01 ... 1.05, 3.90, 3.98
MCVCE, 3rd iteration	100.02	15.20	21.13	0.99 ... 1.01, 3.93, 4.01
MCVCE at convergence	100.02	15.20	21.13	0.99 ... 1.01, 3.93, 4.01
Lerch's method, 1st iteration	102.81	15.40	21.36	1.12 ... 1.38, 3.75, 3.74
Lerch's method, 3rd iteration	100.84	15.16	21.14	0.86 ... 1.13, 3.97, 4.02
Lerch's method, 5th iteration	100.71	15.16	21.15	0.85 ... 1.13, 3.98, 4.02
Lerch's method at convergence	100.69	15.16	21.15	0.85 ... 1.13, 3.98, 4.02

Table 4. Results for the case of noisy observations, plus systematic errors. Uncorrelated random errors of $\sigma_{x,y,z} = 1$ cm standard deviation have been added to the original orbit pseudo-observations for all arcs. In addition, a bias of 5 cm at the 1-CPR orbit frequency of GOCE has been added to the radial component of x, y, z coordinates of arcs 9 and 10

Case 4	Geoid errors (cm)			Variance components (cm) $\hat{\sigma}_i = \dfrac{1}{\hat{w}_i^{1/2}}$
	Max	RMS	Average	
Equal weights	120.79	19.25	25.84	1
MCVCE, 1st iteration	99.91	17.44	23.09	1.00 ... 1.01, 2.10, 2.10
MCVCE, 3rd iteration	100.06	17.41	23.06	0.99 ... 1.01, 2.11, 2.11
MCVCE at convergence	100.06	17.41	23.06	0.99 ... 1.01, 2.11, 2.11
Lerch's method, 1st iteration	94.07	18.04	23.99	0.93 ... 1.10, 1.29, 1.83
Lerch's method, 3rd iteration	98.10	15.11	20.51	0.86 ... 1.12, 1.14, 10.68
Lerch's method, 5th iteration	97.81	15.48	20.98	0.89 ... 1.12, 1.19, 22.41
Lerch's method at convergence	97.79	15.49	20.99	0.89 ... 1.12, 1.21, 22.39

case, here again we assess a problem where systematic effects dominate against the stochastic noise. If no weight optimization takes place at all – i.e. no down-weighting of the 'bad' data sets – the RMS geoid errors are roughly 50% increased when compared to the second case (compare Table 4 to Table 2). MCVCE estimates a sigma value of 2.1 cm for the questionable data sets, and this down-weighting reduces the RMS geoid errors by about 2 cm. It is here that Lerch's randomized method outperforms MCVCE, since by its drastic down-weighting RMS errors are reduced by about 4 cm. On the other hand, the estimated weights for the systematically perturbed arcs are strongly unbalanced and by no means representative of the magnitude of the true simulated errors, whereas MCVCE gives quite accurate weight estimates.

In a fifth test case, all observation arcs have been corrupted by uncorrelated random noise of $\sigma_{x,y,z} = 1$ cm and the systematic 5-cm radial orbit error at 1 CPR. The unexpected result occurs that the best gravity parameter solution can be obtained from the original equal-weight

assumption, whereas MCVCE iterations finally increase the geoid error RMS by about 0.5 cm (see Table 5). The variance component estimates are still quite representative at about 2 cm. Lerch's method performs even more strangely; after an early improvement it converges slowly towards a solution where geoid errors are increased by about 4.7 cm.

In the final, sixth test case, for which results are shown in Table 6, uncorrelated random errors of $\sigma_{x,y,z} = 1$ cm standard deviation (arcs 1 ... 8) and $\sigma_{x,y,z} = 4$ cm standard deviation (arcs 9 and 10) have been applied. In addition, the 5-cm radial bias at 1 CPR has been added to the orbit pseudo-observations for all arcs. Again, MCVCE performs well in terms of geoid errors, weight estimates and speed of convergence. Lerch's randomized method, on the other hand, gives a superior gravity field solution after one iteration but then things start getting worse, and at convergence the geoid RMS errors are about 2.6 cm larger than those obtained with MCVCE.

Table 5. Results for the case of noisy observations, plus systematic errors. Uncorrelated random errors of $\sigma_{x,y,z} = 1$ cm standard deviation as well as a radial bias of 5 cm at the 1-CPR orbit frequency of GOCE have been added to the original orbit pseudo-observations for all arcs

Case 5	Geoid errors (cm)			Variance components (cm) $\hat{\sigma}_i = \dfrac{1}{\hat{w}_i^{1/2}}$
	Max	RMS	Average	
Equal weights	94.78	14.78	19.69	1
MCVCE, 1st iteration	94.84	15.12	19.98	2.06 ... 2.15
MCVCE, 3rd iteration	94.83	15.25	20.10	2.04 ... 2.15
MCVCE, 5th iteration	94.83	15.26	20.11	2.04 ... 2.16
MCVCE at convergence	94.83	15.27	20.11	2.04 ... 2.16
Lerch's method, 1st iteration	95.77	14.03	19.20	1.01 ... 1.41
Lerch's method, 3rd iteration	99.58	15.15	20.29	0.91 ... 2.01
Lerch's method, 5th iteration	101.50	17.45	22.68	0.87 ... 3.42
Lerch's method at convergence	101.62	19.95	25.84	0.86 ... 15.21

Table 6. Results for the case of noisy observations, plus systematic errors. Uncorrelated random errors of $\sigma_{x,y,z} = 1$ cm standard deviation (arcs 1 ... 8) and $\sigma_{x,y,z} = 4$ cm standard deviation (arcs 9 and 10) have been applied. Moreover, a radial bias of 5 cm at the 1-CPR orbit frequency of GOCE has been added to the original orbit pseudo-observations for all arcs

Case 6	Geoid errors (cm)			Variance components (cm) $\hat{\sigma}_i = \dfrac{1}{\hat{w}_i^{1/2}}$
	Max	RMS	Average	
Equal weights	235.78	28.41	38.52	1
MCVCE, 1st iteration	116.66	17.90	23.93	2.07 ... 2.16, 4.32, 4.42
MCVCE, 3rd iteration	115.23	17.89	23.90	2.05 ... 2.16, 4.36, 4.46
MCVCE, 5th iteration	115.21	17.89	23.90	2.05 ... 2.16, 4.36, 4.46
MCVCE at convergence	115.21	17.90	23.90	2.05 ... 2.16, 4.36, 4.46
Lerch's method, 1st iteration	103.27	17.14	22.84	1.21 ... 1.80, 3.79, 3.86
Lerch's method, 3rd iteration	104.88	17.89	23.72	0.87 ... 1.57, 4.11, 4.44
Lerch's method, 5th iteration	106.00	18.15	24.05	0.84 ... 1.62, 4.11, 4.51
Lerch's method at convergence	114.63	20.51	26.56	0.79 ... 8.81, 4.14, 9.40

6. Discussion

An algorithm has been presented that facilitates VCE for disjunctive observation groups, without the need for inversion of the observation group influence matrices or even for explicit computation of the normal matrix. It has been shown that the method, then denoted as MCVCE, can be re-formulated as a Monte-Carlo method. Thus, an existing LS estimation software can be used in a black-box fashion to derive weight estimates. In an iterative procedure random data or randomly perturbed data have to be passed through the given algorithm.

The computational costs of MCVCE, when compared with VCE, reduce considerably. Moreover, the method applies for problems where the matrices involved in a VCE are not available due to restrictions imposed by the LS estimation software in use. Application in satellite geodesy might range from gravity field determination with the new gravity missions in combination with altimetry, airborne or terrestrial data, to orbit and ground station determination. Clearly, we had the GOCE data analysis in our mind, where the amount of gravity parameters makes conventional VCE techniques unattractive. Here the real challenge will be to combine all available data from gradiometry, GPS orbit tracking and other sources in an optimal way.

In a simplified numerical experiment we can show that the proposed algorithm works well for simulated satellite orbit perturbations contaminated with pure stochastic noise, and can reasonably deal with orbit data where simulated systematic errors are present. In comparison, Lerch's method equipped with randomized trace estimation proves to perform generally less stably in terms of convergence, although in some cases with comparable success. If we aim to estimate the variance components of the orbit perturbation data (or at least a representative absolute weight if systematic noise is present), MCVCE clearly outperforms Lerch's method. It should be acknowledged, however, that the major concern in the design of Lerch's method was to obtain useful relative weights, not absolute ones.

A number of important issues have not yet been touched within this paper, remaining for future investigation, among others: (1) how can the proposed method be integrated optimally in nonlinear estimation procedures where many iterations of a linearization and estimation process have to be performed?; (2) how can automatic observation editing algorithms be dealt with?; and (3) the formulation and use of prior information about the weights.

References

Crocetto N, Gatti M, Russo P (2000) Simplified formulae for the BIQUE estimation of variance components in disjunctive observation groups. J Geod 74: 447–457

European Space Agency (1999) Gravity Field and Steady-State Ocean Circulation Mission. Reports for mission selection, ESA SP-1233(1). ESTEC, Noordwijk

Förstner W (1979) Ein Verfahren zur Schätzung von Varianz- und Kovarianzkomponenten. Allg Vermess-Nachr 86: 446–453

Girard DA (1989) A fast 'Monte-Carlo Cross-Validation' procedure for large least squares problems with noisy data. Numer Math 56: 1–23

Golub, GH, van Loan CF (1996) Matrix computations. John Hopkins University Press, Baltimore, MD

Grafarend EW, d'Hone A (1978) Gewichtsschätzung in geodätischen Netzen. Series A, no. 88, German Geodetic Commission, Munich

Grafarend EW, Schaffrin B (1993) Ausgleichungsrechnung in linearen Modellen. BI Wissenschaftsverlag, Mannheim

Helmert FR (1924) Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate, 3rd edn. Teubner, Leipzig

Hutchinson MF (1990) A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. Commun Statist Simulat Comput 19: 433–450

Kizilsu G, Sahin M (2000) SLR precision analysis for LAGEOS I and II. Earth Plan Space 52: 789–794

Koch K-R (1986) Maximum likelihood estimate of variance components. Bull Geod 60: 329–338

Koch K-R (1990) Bayesian inference with geodetic applications. Lecture Notes in Earth Sciences. Springer, Berlin Heidelberg New York

Koch K-R, Kusche J (2002) Regularization of geopotential determination from satellite data by variance components. J Geod 76: 259–268

Kusche J, Klees R (2002). Regularization of gravity field estimation from satellite gravity gradients. J Geod 76: 359–368

Lerch FJ (1991) Optimum data weighting and error calibration for estimation of gravitational parameters. Bull Geod 65: 44–52

Lucas JR, Dillinger WH (1998) MINQUE for block diagonal bordered systems such as those encountered in VLBI data analysis. J Geod 72: 343–349

Marsh JG, Lerch FJ, Putney BH, Christodoulis DC, Smith DE, Felsentreger TL, Sanchez BV, Klosko, SM, Pavlis EC, Martin TV, Robbins JW, Williamson RG, Colombo OL, Rowlands DD, Eddy WF, Chandler NL, Rachlin KE, Patel GB, Bhati S, Chinn, DS (1988) A new gravitational model for the Earth from satellite tracking data: GEM-T1. J Geophys Res 93(B6): 6169–6215

McCarthy JJ, Rowton S, Moore D, Pavlis DE, Luthcke SB, Tsaoussi LS (1993) GEODYN II systems description. NASA Goddard Space Flight Center, Greenbelt, MD

Purser RJ, Parrish DF (2000) A Bayesian technique for estimating continuously varying statistical parameters of a variational assimilation. Office note 429, NOAA National Weather Service/National Centers for Environmental Prediction

Rao CR (1973) Linear statistical inference and its applications. Wiley, New York

Rao CR, Kleffe J (1988) Estimation of variance components and applications. Elsevier, Amsterdam

Reigber C (1989) Gravity field recovery from satellite tracking data. In: Sanso F, Rummel R (eds) Theory of satellite geodesy and gravity field determination. Lecture Notes in Earth Sciences, 25. Springer, Berlin Heidelberg New York

Sahin M, Cross PA, Sellers PC (1992) Variance component estimation applied to satellite laser ranging. Bull Geod 66: 284–295

Schaffrin B (1983) Varianz–Kovarianz-Komponenten-Schätzung bei der Ausgleichung heterogener Wiederholungsmessungen. Series C, no. 282, German Geodetic Commission, Munich

Schwintzer P (1990) Sensitivity analysis in least squares gravity field modelling – redundancy decomposition of stochastic a priori information. Internal rep Deutsches Geodätisches Forschungsinstitut, Munich

Schwintzer P, Reigber C, Bode A, Kang Z, Zhu SY, Massmann F-H, Raimondo JC, Biancale R, Balmino G, Lemoine JM, Moynot B, Marty JC, Barlier F, Boudon Y (1997) Long-wavelength global gravity field models: GRIM4-S4, GRIM4-C4. J Geod 71: 189–208

Visser P, van den IJssel J (2000) GPS-based precise orbit determination of the very low Earth-orbiting gravity mission GOCE. J Geod 74: 590–602

Visser P, van den IJssel J, Koop R, Klees R (2001) Exploring gravity field determination from orbit perturbations of the European gravity mission GOCE. J Geod 75: 89–98

Zhang F, Huang C, Feng C, Dong X, Liao X (2001). Application of variance component estimation to precise orbit determination for ERS-2. Chin Astron Astrophys 25: 478–489