# GJI Geodesy, potential field and applied geophysics

# A covariance-adaptive approach for regularized inversion in linear models

# Christopher Kotsakis

Department of Geodesy and Surveying, The Aristotle University of Thessaloniki, University Box 440, Thessaloniki GR-54124, Greece. E-mail: kotsaki@topo.auth.gr

Accepted 2007 June 19. Received 2006 November 21; in original form 2006 October 4

# SUMMARY

The optimal inversion of a linear model under the presence of additive random noise in the input data is a typical problem in many geodetic and geophysical applications. Various methods have been developed and applied for the solution of this problem, ranging from the classic principle of least-squares (LS) estimation to other more complex inversion techniques such as the Tikhonov–Philips regularization, truncated singular value decomposition, generalized ridge regression, numerical iterative methods (Landweber, conjugate gradient) and others. In this paper, a new type of optimal parameter estimator for the inversion of a linear model is presented. The proposed methodology is based on a linear transformation of the classic LS estimator and it satisfies two basic criteria. First, it provides a solution for the model parameters that is optimally fitted (in an average quadratic sense) to the classic LS parameter solution. Second, it complies with an external user-dependent constraint that specifies a priori the error covariance (CV) matrix of the estimated model parameters. The formulation of this constrained estimator offers a unified framework for the description of many regularization techniques that are systematically used in geodetic inverse problems, particularly for those methods that correspond to an eigenvalue filtering of the ill-conditioned normal matrix in the underlying linear model. Our study lies on the fact that it adds an alternative perspective on the statistical properties and the regularization mechanism of many inversion techniques commonly used in geodesy and geophysics, by interpreting them as a family of 'CV-adaptive' parameter estimators that obey a common optimal criterion and differ only on the pre-selected form of their error CV matrix under a fixed model design.

**Key words:** biased estimation, covariance constraint, decorrelation, discrete inverse problem, regularization.

# **1 INTRODUCTION**

A generic estimation problem in many geodetic and geophysical applications is that of determining a set of unknown deterministic parameters which are observed through a known transformation model and corrupted by additive random noise. This type of problem is commonly formulated in terms of a linear system of observation equations  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$ , which needs to be uniquely solved for the unknown parameter vector x given a set of noisy data y (Tarantola 1987; Menke 1989; Dermanis & Rummel 2000). The residual term v describes the effect of random disturbances in the input data, whereas the elements of the matrix A correspond to known coefficients that depend on the conditions of the physical system under study. Note that such problems are inherently ill-posed as they can never have a unique solution due to the existence of the unknown random observation errors. In algebraic terms, we deal with an underdetermined system that has an infinite number of solutions, since for every candidate solution  $\hat{\mathbf{x}}$  for the parameter vector there always exists a corresponding estimate of the unknown data errors,  $\mathbf{\hat{v}}=\mathbf{y}-\mathbf{A}\mathbf{\hat{x}},$  which satisfies the original system of observation equations.

Typical examples of such discrete inverse problems arise in several fields of geosciences, including gravity field modelling and satellite geodesy (Schwarz 1979; Rummel 1985; Sanso & Rummel 1989; Xu 1992; Xu & Rummel 1994; Ilk et al. 2002), satellite orbit determination (Engelis & Knudsen 1989; Beutler et al. 1994; Tapley et al. 2004), tectonic plate motions (Chase & Stuart 1972; Minster et al. 1974), satellite altimetry and ocean circulation studies (Wunsch & Minster 1982; Knudsen 1992; Fu & Cazenave 2001), magnetic field modelling (Parker et al. 1987; Backus 1988), atmospheric tomography (Kunitake 1996; Kleijer 2004), seismology and acoustic tomography (Crosson 1976; Aki & Richards 1980; Pavlis & Booker 1980), wave propagation and scattering theory (Bee & Jacobson 1984; Koch 1985) and core and mantle dynamics (Parker 1970; Gire et al. 1986; Nataf et al. 1986).

A fairly common approach for solving parameter estimation problems in linear models is the method of least-squares (LS) which provides an optimal solution  $\mathbf{\hat{x}}^{LS}$  that reproduces as close as possible the given data by minimizing the weighted differences between the original measurements **y** and the model-induced observables  $\mathbf{A}\hat{\mathbf{x}}^{LS}$ . In statistical terms, the LS solution gives an unbiased estimate for the unknown parameters provided that there are no modelling errors or other hidden systematic effects in the data. In addition, it has the minimum mean squared estimation error (MSEE) among all linear unbiased estimators if the chosen weight matrix for the input data is equal to the inverse covariance (CV) matrix of their random observation errors; for more details, see Menke (1989), Koch (1999) and Rao & Toutenburg (1999).

Within the context of geodetic and geophysical applications, the LS methodology has been used not only as an optimal algorithm for estimating a set of unknown parameters from noisy measurements in a given linear model, but also as a trend analysis tool for identifying the main features of partially known physical systems through the fitting of simple analytic models to complex data records. In any case, the accuracy of LS inversion results may not be satisfactory in practice if the underlying linear model that relates the available data and the unknown parameters is ill-conditioned. In such cases, one or more eigenvalues of the corresponding normal matrix are close to zero, thus causing a large statistical uncertainty in the LS solution even for very precise data sets (Hoerl & Kennard 1970; Sen & Srivastava 1990). In this category belong many geodetic estimation problems that can be expressed in the linear form y = Ax + v, including the downward continuation of gravity field measurements from satellite or aerial altitude to the surface of the earth, the inversion of altimetric data for marine gravity determination, and the computation of gravity gradients from gravity anomaly data. For a general overview of ill-posed inverse problems in geodesy, see Neyman (1985) and Ilk (1991).

Various alternative techniques to the classic LS method have been developed in order to improve the parameter estimation accuracy in linear model inversion. Despite their differences and (sometimes) conflicting interpretations, most of these regularization techniques seem to follow a common rationale which points to the fact that a successful solution to a discrete inverse problem should not be solely determined by the quality of the fit between the data and the corresponding observables obtained from the estimated model parameters (i.e. the typical 'LS logic'), but it requires also some quality control for the estimated parameters themselves. As an example, we can mention the method of Tikhonov-Philips regularization (Tikhonov & Arsenin 1977) which has been used extensively for the stable inversion of ill-conditioned linear models based on a hybrid optimal criterion that takes into account both the data-model fitting aspect and the size or smoothness of the actual model parameters (Schwarz 1979; Schaffrin 1980; Ilk et al. 2002; Kusche & Klees 2002; Ditmar et al. 2003). Other perspectives of the Tikhonov-Philips inversion scheme have also been considered by many researchers, including its implementation from a biased estimation or ridge regression viewpoint (Xu 1992; Xu & Rummel 1994; Bouman & Koop 1997), and its Bayesian interpretation which involves a priori parameter weighting (Xu & Rummel 1994) and variance component estimation procedures (Koch & Kusche 2002). Alternative regularization approaches that have been followed for the optimal inversion of linear models include the truncated singular value decomposition (TSVD) or bandpass filtering (Xu 1998), stochastic regularization techniques (Schaffrin 1985, 1989), LS collocation (Rummel et al. 1976, 1979) and numerical iterative methods (Landweber algorithm, conjugate gradient algorithm). A detailed comparison of regularization techniques for geodetic inverse problems has been given in Bouman (1998), where a unified framework for the spectral behaviour of several regularized estimators was presented; see also Gerstl & Rummel (1981).

The objective of this paper is to present a new parameter estimator  $\hat{\mathbf{x}}$  for linear models  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$ , which unifies under a common statistical framework all regularization methods that are systematically used in geodetic inverse problems. Our study lies on the fact that it adds an alternative perspective on the optimal statistical properties and the regularization mechanism of many inversion techniques commonly used in geodesy and geophysics, including Tikhonov-Philips regularization, generalized ridge regression (GRR) and TSVD estimation. The structure of the paper is organized as follows. In Section 2, the formulation and the properties of the new parameter estimator are presented along with the derivation of its analytic form. Additional topics of study in this section include the MSEE performance and the inherent bias characteristics of the new estimator, as well as the admissible range for the signal-to-noise ratio (SNR) of an arbitrary linear model that guarantees better MSEE performance when using the new estimator over the classic LS solution. Some particular regularization aspects associated with the new parameter estimator are discussed in Section 3, where it is shown that all regularized inversion techniques corresponding to an eigenvalue filtering of the normal matrix of the underlying linear model  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$  are essentially special cases of the general estimation framework presented herein. An example for the numerical solution of a typical ill-conditioned problem in physical geodesy, namely the downward continuation of satellite gradiometry data for the determination of a gravity anomaly grid on the surface of the Earth, is given in Section 4 to demonstrate the practical implementation and the usefulness of the new estimator in geodetic inverse problems. Finally, some concluding remarks are given at the end of the paper.

# 2 CV-ADAPTIVE PARAMETER ESTIMATION IN LINEAR MODELS

### 2.1 Problem formulation

Let us consider the standard linear model that is commonly used for tackling discrete inverse problems in geodesy (Xu & Rummel 1994; Xu 1998; Dermanis & Rummel 2000)

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

where y is a known observation vector, A is a matrix of known coefficients whose values depend on the geometrical and physical conditions of the data modelling process and  $\mathbf{x}$  is a vector of unknown deterministic parameters. The residual term v corresponds to a vector of zero-mean random variables with a CV matrix  $C_v$ , and it describes the effect of random errors and other stochastic disturbances in the available data. Although it is not crucial for our problem formulation to specify the exact nature of the observations and the unknowns, it is nevertheless worth mentioning that eq. (1)often corresponds to a discretized Fredholm integral equation of the first kind with noisy data. This type of integral equations provides the fundamental mathematical basis for describing and studying several ill-posed inverse problems in geodesy and geophysics (Neyman 1985; Ilk 1991; Bouman 1998). In gravity field modelling, for example, the system in eq. (1) can describe the problem of marine gravity determination from satellite altimetry data using Hotine's or Stokes' integral, or the downward continuation of satellite and/or aerial gravity field data based on Poisson's integral (Heiskanen & Moritz 1967).

The optimal inversion of eq. (1) will be based on a *tailoring methodology*, according to which a known reference estimator of **x** is modified in such a way that it produces an improved solution for the unknown model parameters. This kind of constructive reasoning has actually shaped the development of several parameter estimators in statistics theory (Bibby & Toutenburg 1977). Following Rao (1973), an improved estimator may be heuristically described as one which is derived in the following manner: starting with some prescribed solution, say  $\mathbf{\hat{x}}_1$ , having known statistical properties of bias, variance, etc., we try to adapt it in some way so that a newly generated estimator  $\mathbf{\hat{x}}_2$  (based on  $\mathbf{\hat{x}}_1$ ) is better in terms of some chosen criterion function. For instance, if the criterion function is the MSEE, then we should say that  $\mathbf{\hat{x}}_2$  is better than  $\mathbf{\hat{x}}_1$  if  $E\{(\mathbf{\hat{x}}_2 - \mathbf{x})^T(\mathbf{\hat{x}}_2 - \mathbf{x})\}$  is smaller than  $E\{(\mathbf{\hat{x}}_1 - \mathbf{x})^T(\mathbf{\hat{x}}_1 - \mathbf{x})\}$  (for more details, see Bibby & Toutenburg 1977).

In this paper, we form an optimal estimator  $\hat{\mathbf{x}}$  for the parameter vector  $\mathbf{x}$  in a linear model  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$  based on a linear transformation of the classic LS solution (or equivalently, the minimum-variance linear unbiased solution). Thus, we have the following general scheme

$$\hat{\mathbf{x}} = \mathbf{R}\hat{\mathbf{x}}^{LS} \tag{2}$$

with  $\mathbf{\hat{x}}^{LS}$  denoting the usual LS estimator

$$\hat{\mathbf{x}}^{LS} = \left(\mathbf{A}^{\mathrm{T}} \mathbf{C}_{\mathbf{v}}^{-1} \mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{C}_{\mathbf{v}}^{-1} \mathbf{y}$$
(3)

and **R** being a transformation matrix that needs to be determined according to some optimal criteria for the performance of  $\hat{\mathbf{x}}$ .

Despite its optimal statistical properties, the LS solution in eq. (3) gives poor results when the original linear model is ill-conditioned. The effect of the matrix  $\mathbf{R}$  in eq. (2) should thus be understood in a 'regularization' context as an attempt to correct the LS solution by adjusting it closer to the true vector x. In fact, many well-known regularization techniques are known to produce parameter estimators that can take the form of eq. (2) for specific expressions of the matrix R, including Tikhonov-Philips regularization, GRR, shrunken LS estimation, TSVD estimation and principal component estimation (Marquardt 1970; Mayer & Willke 1973). Note that the numerical implementation of the modified estimator  $\hat{\mathbf{x}}$  does not necessarily require the prior computation of the entire LS solution (which is actually problematic in the case of ill-conditioned problems), provided that the matrix product  $\mathbf{R} (\mathbf{A}^T \mathbf{C}_{\mathbf{v}}^{-1} \mathbf{A})^{-1}$  can be separately evaluated in a numerically stable and efficient manner (more details given in Section 3).

Our aim is to determine a general form for the regularization matrix **R** by combining (i) a 'global' optimal criterion that minimizes the weighted difference between the original LS solution and the transformed LS solution obtained from eq. (2) and (ii) a constraint that sets the error CV matrix of the modified estimator  $\hat{\mathbf{x}}$  equal to an *a priori*, user-specified form.

As it will be shown in later sections, all existing regularization methods for discrete inverse problems that correspond to an *eigenvalue filtering* of the ill-conditioned normal matrix  $\mathbf{N} = \mathbf{A}^T \mathbf{C}_v^{-1} \mathbf{A}$  can be formulated as special cases of the aforementioned framework. This kind of unification adds an interesting perspective on the statistical properties and the regularization mechanism of many inversion techniques that are currently used in geodesy and geophysics, by interpreting them as a collection of 'CV-adaptive' parameter estimators that obey the same optimal criterion and differ only on the pre-selected form of their error CV matrix under a fixed model design.

# 2.2 CV-adaptive inversion methodology and optimal regularization matrix

A fundamental characteristic of the LS solution is its unbiasedness,  $E{\hat{\mathbf{x}}^{LS}} = \mathbf{x}$  under the absence of any modelling errors. From the Gaussian viewpoint of LS estimation theory, the property of unbiasedness is essentially a *constraint* that is *a priori* imposed on the first moment of a linear estimator in conjunction with the minimum MSEE principle.

Here we seek an optimal parameter estimator  $\hat{\mathbf{x}}$  by imposing an *a priori* constraint on its second (centred) moment  $\mathbf{C}_{\hat{\mathbf{x}}} = E\{(\hat{\mathbf{x}} - E\{\hat{\mathbf{x}}\})^{\mathsf{T}}\}$ . Since the general form of  $\hat{\mathbf{x}}$  is taken as a linear transform of the LS solution according to eq. (2), a constraint on its second centred moment can be formulated in terms of the matrix equation

$$\mathbf{R}\mathbf{N}^{-1}\mathbf{R}^{\mathrm{T}} = \mathbf{C}_{\hat{\mathbf{x}}},\tag{4}$$

where  $C_{\hat{x}}$  is the user-specified CV matrix of the estimated parameters,  $N = A^T C_v^{-1} A$  is the normal matrix of the underlying linear model, and **R** is the required transformation matrix.

Note that the CV matrix of the estimated model parameters is identical to the CV matrix of the corresponding estimation errors  $\mathbf{e} = \hat{\mathbf{x}} - \mathbf{x}$ , since

$$\mathbf{C}_{\hat{\mathbf{x}}} = E\{(\hat{\mathbf{x}} - E\{\hat{\mathbf{x}}\})(\hat{\mathbf{x}} - E\{\hat{\mathbf{x}}\})^{\mathrm{T}}\}$$
  
=  $E\{(\hat{\mathbf{x}} - E\{\mathbf{e}\} - \mathbf{x})(\hat{\mathbf{x}} - E\{\mathbf{e}\} - \mathbf{x})^{\mathrm{T}}\}$   
=  $E\{(\mathbf{e} - E\{\mathbf{e}\})(\mathbf{e} - E\{\mathbf{e}\})^{\mathrm{T}}\}$   
=  $\mathbf{C}_{\mathbf{e}}.$ 

The required regularization matrix **R** cannot be uniquely determined through the matrix constraint in eq. (4). Hence, apart from its CV-adaptivity, an additional optimal property will be assigned to the estimator  $\hat{\mathbf{x}}$  in terms of the principle

$$\varphi(\mathbf{\hat{x}}) = E\{\|(\mathbf{\hat{x}} - \mathbf{\hat{x}}^{LS})\|^2\} = E\{(\mathbf{\hat{x}} - \mathbf{\hat{x}}^{LS})^{\mathsf{T}}\mathbf{P}(\mathbf{\hat{x}} - \mathbf{\hat{x}}^{LS})\} = \min, \quad (5)$$

where the weight matrix **P** is taken as the inverse CV matrix of the LS solution;  $\mathbf{P} = \mathbf{C}_{\mathfrak{g}^{LS}}^{-1} = \mathbf{N}$ .

The parameter estimator  $\hat{\mathbf{x}}$  is thus characterized by two basic statistical properties.

1. It adheres to a minimization criterion that provides an optimal weighted fit with the LS solution according to eq. (5). The particular choice of the weight matrix **P** as the inverse CV matrix of  $\hat{\mathbf{x}}^{LS}$  ensures that the information about the parameter estimation accuracy in the original LS solution is assimilated into the new estimator by adjusting the elements of  $\hat{\mathbf{x}}$  according to the dispersion of the corresponding elements of  $\hat{\mathbf{x}}^{LS}$ .

2. It complies with a user-dependent constraint that specifies the variance and the correlation structure of the parameter estimation errors through an *a priori* selected CV matrix  $C_{\hat{x}}$ . Particular strategies and further discussion pertaining to the choice of  $C_{\hat{x}}$  will be given in Section 3.

From the combination of the above properties we can determine the required regularization matrix  $\mathbf{R}$  according to an analytic procedure which is described in detail in Appendix A. The final result is given by the following formula:

$$\mathbf{R} = \mathbf{N}^{-1/2} \left( \mathbf{N}^{1/2} \mathbf{C}_{\hat{\mathbf{x}}} \mathbf{N}^{1/2} \right)^{1/2} \mathbf{N}^{1/2}.$$
 (6)

In Appendix B it is proven that the above non-symmetric matrix is also equal to the alternative expression

$$\mathbf{R} = \mathbf{C}_{\hat{\mathbf{x}}}^{1/2} \left( \mathbf{C}_{\hat{\mathbf{x}}}^{1/2} \mathbf{N} \mathbf{C}_{\hat{\mathbf{x}}}^{1/2} \right)^{1/2} \mathbf{C}_{\hat{\mathbf{x}}}^{-1/2}.$$
 (7)

# Remark 1:

The square root  $C^{1/2}$  of a symmetric non-negative definite matrix **C** is a unique matrix which is determined through the square roots of its eigenvalues. In particular, if  $C = U\Lambda U^T$  is the orthonormalized eigenvalue decomposition of **C**, then we have that  $C^{1/2} = U\Lambda^{1/2} U^T$ , where the diagonal matrix  $\Lambda^{1/2}$  contains the square roots of the eigenvalues of **C**. Also, the inverse square root of a symmetric positive definite matrix is defined as  $C^{-1/2} = (C^{1/2})^{-1} = (C^{-1})^{1/2}$  (for more details see Rao & Toutenburg 1999, pp. 362–363).

Remark 2:

Using the matrix identity  $\mathbf{ST}^{1/2} \mathbf{S}^{-1} = (\mathbf{STS}^{-1})^{1/2}$ , it is easily verified that the optimal regularization matrix can also be expressed through the formula

$$\mathbf{R} = (\mathbf{C}_{\hat{\mathbf{x}}} \mathbf{N})^{1/2}.$$
(8)

In the trivial case that we choose  $C_{\hat{x}} = N^{-1}$ , the regularization matrix becomes identity ( $\mathbf{R} = \mathbf{I}$ ) and our inversion scheme simply reproduces the LS solution, as it should be expected.

# 2.3 Bias of the CV-adaptive estimator

The CV-adaptive estimator that was introduced in the previous section gives, in general, a biased solution for the unknown model parameters. Its associated bias vector is

$$BIAS(\hat{\mathbf{x}}) = E\{\hat{\mathbf{x}}\} - \mathbf{x} = (\mathbf{R} - \mathbf{I})\mathbf{x},$$
(9)

where  $\mathbf{R}$  is the regularization matrix obtained from eq. (6) or (7). By making use of the unbiasedness property of the LS solution, the above bias vector can be also expressed as

$$BIAS(\hat{\mathbf{x}}) = E\{\hat{\mathbf{x}} - \hat{\mathbf{x}}^{LS}\} = E\{\mathbf{d}\},$$
(10)

where **d** denotes the difference between our CV-adaptive estimator and the LS estimator.

Let us now rewrite the minimization principle which was previously introduced for the construction of the regularization matrix  $\mathbf{R}$  (see eq. 5) in the equivalent form

$$\varphi(\mathbf{\hat{x}}) = E\{\|\mathbf{d}\|^2\} = E\{\mathbf{d}^{\mathrm{T}}\mathbf{N}\mathbf{d}\} = \min.$$
(11)

The optimality of the CV-adaptive estimator, in terms of satisfying the above criterion, is closely related to the reduction of its inherent bias. This relationship is more clearly seen through a spectral representation approach, and for this reason let us consider the eigenvalue decomposition of the normal matrix **N** 

$$\mathbf{N} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^{\mathrm{T}},\tag{12}$$

where  $\mathbf{U} = [\mathbf{u}_1 \dots \mathbf{u}_n]$  is a unitary matrix formed by the orthonormal eigenvectors and  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$  is a diagonal matrix containing the eigenvalues of **N**. Using the above factorization, the criterion function  $\varphi(\hat{\mathbf{x}})$  can be expressed as

.

$$\varphi(\mathbf{\hat{x}}) = E\left\{\mathbf{d}^{\mathrm{T}}\sum_{i=1}^{n}\lambda_{i}\mathbf{u}_{i}\mathbf{u}_{i}^{\mathrm{T}}\mathbf{d}\right\}$$
$$= \sum_{i=1}^{n}\lambda_{i}E\{(\mathbf{u}_{i}^{\mathrm{T}}\mathbf{d})^{2}\},$$
(13)

where *n* is equal to the number of the unknown parameters  $(n = \dim \mathbf{x})$ .

Also, using eq. (10), the bias vector can be represented in the following spectral form

$$BIAS(\hat{\mathbf{x}}) = E\{\mathbf{d}\} = E\{\mathbf{U}\mathbf{U}^{\mathrm{T}}\mathbf{d}\}$$
$$= E\left\{\sum_{i=1}^{n} \mathbf{u}_{i}\mathbf{u}_{i}^{\mathrm{T}}\mathbf{d}\right\}$$
$$= \sum_{i=1}^{n} E\{(\mathbf{u}_{i}^{\mathrm{T}}\mathbf{d})\}\mathbf{u}_{i}.$$
(14)

The squared magnitude (Euclidean length) of the above bias vector takes the generalized Pythagorean expression

$$\|\mathrm{BIAS}(\hat{\mathbf{x}})\|^{2} = \left[\sum_{i=1}^{n} E\left\{\left(\mathbf{u}_{i}^{\mathrm{T}}\mathbf{d}\right)\right\}\mathbf{u}_{i}\right]^{1}\left[\sum_{j=1}^{n} E\left\{\left(\mathbf{u}_{j}^{\mathrm{T}}\mathbf{d}\right)\right\}\mathbf{u}_{j}\right]$$
$$= \sum_{i=1}^{n} \left(E\left\{\mathbf{u}_{i}^{\mathrm{T}}\mathbf{d}\right\}\right)^{2}.$$
(15)

Using the well-known fact that the mean squared value of any random variable is always greater (or equal) than the square of its mean value, we have that

$$\|\mathrm{BIAS}(\hat{\mathbf{x}})\|^2 \le \sum_{i=1}^n E\left\{\left(\mathbf{u}_i^T \mathbf{d}\right)^2\right\}.$$
(16)

Comparing eqs (13) and (16), we can conclude that the minimization of the criterion function  $\varphi(\mathbf{\hat{x}})$  is equivalent to a weighted minimization of the maximum bias' Euclidean spectral components, with respect to the orthonormal basis formed by the eigenvectors of the normal matrix **N**.

# Remark 3:

The fact that the CV-adaptive estimator is optimized through a fitting to the classic LS solution may be perceived as a 'weak' point, due to the problems associated with the LS solution in ill-conditioned problems. However, a closer look at the formulation of the minimization principle in eq. (5) reveals that the fitting does not take place with a *particular* LS solution, but rather it takes place over the entire ensemble of possible solutions that we would obtain if we repeat the data collection and the LS inversion process an infinite number of times. Moreover, as seen from the previous analysis, the optimality of the CV-adaptive estimator in terms of achieving a best average quadratic fit with the LS estimator is essentially a means to control its inherent bias.

### 2.4 MSEE performance and admissible SNR range

The evaluation of the statistical accuracy of the CV-adaptive estimator  $\hat{\mathbf{x}}$  requires the analysis of its MSEE matrix

$$MSEE(\hat{\mathbf{x}}) = E\{(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^{\mathrm{T}}\}.$$
(17)

The above matrix is usually decomposed into two parts which reflect the additive contributions of the internal precision (variance and correlation structure) and the external accuracy (bias) to the total mean squared error of the estimated parameters. Taking into account the bias expression from eq. (9), we have

$$MSEE(\hat{\mathbf{x}}) = E\{(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^{\mathsf{T}}\}$$
  
=  $E\{(\hat{\mathbf{x}} - E\{\hat{\mathbf{x}}\} + E\{\hat{\mathbf{x}}\} - \mathbf{x})(\hat{\mathbf{x}} - E\{\hat{\mathbf{x}}\} + E\{\hat{\mathbf{x}}\} - \mathbf{x})^{\mathsf{T}}\}$   
=  $E\{(\hat{\mathbf{x}} - E\{\hat{\mathbf{x}}\} + \mathrm{BIAS}(\hat{\mathbf{x}})][\hat{\mathbf{x}} - E\{\hat{\mathbf{x}}\} + \mathrm{BIAS}(\hat{\mathbf{x}}))^{\mathsf{T}}\}$   
=  $E\{(\hat{\mathbf{x}} - E\{\hat{\mathbf{x}}\})(\hat{\mathbf{x}} - E\{\hat{\mathbf{x}}\})^{\mathsf{T}}\}$   
+ $E\{(\hat{\mathbf{x}} - E\{\hat{\mathbf{x}}\})[\mathrm{BIAS}(\hat{\mathbf{x}})]^{\mathsf{T}}\}$   
+ $E\{\mathrm{BIAS}(\hat{\mathbf{x}})(\hat{\mathbf{x}} - E\{\hat{\mathbf{x}}\})^{\mathsf{T}}\}$ 

$$+E\{[BIAS(\hat{\mathbf{x}})][BIAS(\hat{\mathbf{x}})]^{T}\}$$
  
=  $\mathbf{C}_{\hat{\mathbf{x}}} + BIAS(\hat{\mathbf{x}})[BIAS(\hat{\mathbf{x}})]^{T}$   
=  $\mathbf{C}_{\hat{\mathbf{x}}} + (\mathbf{R} - \mathbf{I})\mathbf{x}[(\mathbf{R} - \mathbf{I})\mathbf{x}]^{T}$   
=  $\mathbf{C}_{\hat{\mathbf{x}}} + (\mathbf{R} - \mathbf{I})\mathbf{x}\mathbf{x}^{T}(\mathbf{R} - \mathbf{I})^{T},$  (18)

where  $C_{\hat{x}}$  is the *a priori* chosen CV matrix of  $\hat{x}$ , and **R** is the associated regularization matrix obtained from eq. (6) or (7).

We shall consider a scalar measure of the total MSEE of  $\hat{\mathbf{x}}$ , namely the trace of its MSEE matrix

$$mse(\hat{\mathbf{x}}) = trMSEE(\hat{\mathbf{x}})$$
$$= trC_{\hat{\mathbf{x}}} + \mathbf{x}^{T}Q\mathbf{x}$$
(19)

which can be used to quantify the average (per parameter component) MSEE of the entire solution. Note that  ${\bf Q}$  denotes the symmetric matrix

$$\mathbf{Q} = (\mathbf{R} - \mathbf{I})^{\mathrm{T}} (\mathbf{R} - \mathbf{I})$$
(20)

which is non-negative definite due to its factorization as a product of a square matrix with its transpose (Harville 1997, p. 213). In this way, the following basic inequality holds (Strang 1988, pp. 348–349)

$$q_{\min} \mathbf{x}^{\mathrm{T}} \mathbf{x} \le \mathbf{x}^{\mathrm{T}} \mathbf{Q} \mathbf{x} \le q_{\max} \mathbf{x}^{\mathrm{T}} \mathbf{x}$$
(21)

for any parameter vector  $\mathbf{x}$ . The quantities  $q_{\min}$  and  $q_{\max}$  correspond to the minimum and maximum eigenvalues of  $\mathbf{Q}$ , and they depend directly on the regularization matrix  $\mathbf{R}$ . The maximum eigenvalue  $q_{\max}$ , in particular, plays an important role in the assessment of the performance of the CV-adaptive estimator. The quadratic form  $\mathbf{x}^{\mathrm{T}}$  $\mathbf{Q}\mathbf{x}$  is equal to the squared magnitude of the bias vector BIAS( $\hat{\mathbf{x}}$ ) =  $(\mathbf{R} - \mathbf{I})\mathbf{x}$ , and thus  $q_{\max}$  quantifies the maximum *relative* bias effect in the parameter solution  $\hat{\mathbf{x}}$ .

Applying the previous inequality into eq. (19), we get

$$tr\mathbf{C}_{\hat{\mathbf{x}}} + q_{\min}\mathbf{x}^{\mathrm{T}}\mathbf{x} \le mse(\hat{\mathbf{x}}) \le tr\mathbf{C}_{\hat{\mathbf{x}}} + q_{\max}\mathbf{x}^{\mathrm{T}}\mathbf{x}.$$
(22)

Let us now compare the CV-adaptive estimator  $\hat{\mathbf{x}}$  and the classic LS estimator  $\hat{\mathbf{x}}^{LS}$ , in terms of their MSEE performance. Due to its unbiasedness property, the MSEE matrix of the LS estimator is equal to its CV matrix

$$MSEE(\hat{\mathbf{x}}^{LS}) = \mathbf{C}_{\hat{\mathbf{x}}^{LS}} = \mathbf{N}^{-1}$$
(23)

and thus the scalar MSEE takes the simple form

$$mse(\hat{\mathbf{x}}^{LS}) = \operatorname{tr} \mathbf{N}^{-1} = \sum_{i=1}^{n} \frac{1}{\lambda_i}.$$
 (24)

In this case, the MSEE of the LS parameter estimator is bounded as follows

$$\frac{1}{\lambda_{\min}} \le mse(\mathbf{\hat{x}}^{LS}) \le \frac{\dim \mathbf{x}}{\lambda_{\min}},\tag{25}$$

where  $\lambda_{min}$  denotes the minimum eigenvalue of the normal matrix N.

By comparing eqs (22) and (25), we see that the MSEE of the LS estimator has an upper bound that is completely independent of the magnitude of the unknown model parameters, whereas the MSEE of the CV-adaptive estimator has an upper bound that depends directly on the Euclidean length of the parameter vector. Such a result indicates the potential of the biased estimator  $\hat{\mathbf{x}}$  to produce better results (with significantly smaller MSEE compared to the classic LS solution) in inverse problems with sufficiently low SNR. In fact, an admissible SNR range for an arbitrary linear model can be derived, within which the CV-adaptive estimator is guaranteed to outperform the LS estimator in the MSEE sense.

Let us first define the SNR of a linear model  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$  in terms of the ratio

$$SNR = \frac{\mathbf{x}^{\mathrm{T}}\mathbf{x}/\dim\mathbf{x}}{\mathrm{tr}\mathbf{C}_{\mathrm{v}}/\dim\mathbf{y}}.$$
(26)

The above quantity reflects the average (per parameter component) relative power of the parametric model with respect to the average noise level in its input data.

Taking into account eqs (22) and (24), the following inequality gives a sufficient condition to ensure MSEE improvement in linear model inversion when using the CV-adaptive estimator instead of the LS estimator [i.e.  $mse(\mathbf{\hat{x}}) < mse(\mathbf{\hat{x}}^{LS})$ ]

$$tr \mathbf{C}_{\hat{\mathbf{x}}} + q_{\max} \mathbf{x}^{\mathrm{T}} \mathbf{x} 
(27)$$

or equivalently

$$\mathbf{x}^{\mathrm{T}}\mathbf{x} < \frac{\mathrm{tr}\mathbf{N}^{-1} - \mathrm{tr}\mathbf{C}_{\hat{\mathbf{x}}}}{q_{\mathrm{max}}}.$$
(28)

Using the SNR definition from eq. (26), we finally obtain

$$SNR < \frac{\dim \mathbf{y}}{\dim \mathbf{x}} \frac{\operatorname{tr} \mathbf{N}^{-1} - \operatorname{tr} \mathbf{C}_{\hat{\mathbf{x}}}}{q_{\max} \operatorname{tr} \mathbf{C}_{\mathbf{y}}}$$
(29)

which defines the admissible SNR range for a linear model that justifies the use of the CV-adaptive estimator  $\mathbf{\hat{x}}$ . The smaller the SNR value of a particular model  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$  with respect to the upper bound given in eq. (29), the greater will be the reduction of the MSEE of  $\mathbf{\hat{x}}$ .

Note that the quantities dim **x**, dim **y**,  $trN^{-1}$  and  $trC_v$  are modeland data-dependent, whereas the terms  $q_{\text{max}}$  and  $trC_{\hat{x}}$  are controlled by the *a priori* choice of the CV matrix for the estimated parameters.

As a final comment, we should mention that in every linear model there will *always* exist an admissible SNR range for applying the CV-adaptive biased estimator, provided that the latter is set to have smaller total error variance than the LS solution. Indeed, in such a case we have that  $trN^{-1} > trC_s$  and the right side of the inequality condition (29) will always give a positive upper bound for the admissible SNR of the underlying linear model.

# 3 CHOICE OF C<sub>2</sub>-REGULARIZATION ASPECTS

### 3.1 General remarks

The *a priori* choice of the error CV matrix for the estimated parameters is a key issue for the implementation of the CV-adaptive estimator  $\hat{\mathbf{x}}$ . Although it may seem perplexing to formulate an inversion scheme which conforms to a user-specified error CV structure for the estimated model parameters, the introduction of an external constraint for  $C_{\hat{\mathbf{x}}}$  should be seen as a prospective regularization tool that gives us the flexibility to obtain and compare different solutions for the model parameters with particular statistical quality characteristics.

A relevant example is the case where the CV-adaptive estimator is designed to give a fully decorrelated solution with uniform precision for the model parameters,  $C_{\hat{x}} = \sigma^2 I$ . This particular choice (which is actually equivalent to a special implementation of GRR, as it will be explained later on) can be beneficial in inverse problems  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$  with a large number of unknown parameters that will be subsequently used for the determination of other quantities of interest  $\mathbf{z} = f(\mathbf{x})$ . The existence of a diagonal-scalar CV matrix  $\mathbf{C}_{\hat{\mathbf{x}}}$ simplifies to a large extent the accuracy assessment of  $\hat{\mathbf{z}} = f(\hat{\mathbf{x}})$ . at least with respect to the rigorous propagation of random-related errors.

From an algebraic viewpoint, the regularization process in linear model inversion is usually viewed in terms of its filtering effect on the spectral representation of the traditional LS estimator (Gerstl & Rummel 1981; Bouman 1998). The LS inversion of a linear model  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$  can be represented through the orthogonal expansion

$$\hat{\mathbf{x}}^{LS} = \sum_{i=1}^{n} \frac{\left(\mathbf{u}_{i}^{\mathrm{T}}\mathbf{b}\right)}{\lambda_{i}} \mathbf{u}_{i}, \qquad (30)$$

where b denotes the data-dependent vector

$$\mathbf{b} = \mathbf{A}^{\mathrm{T}} \mathbf{C}_{\mathbf{v}}^{-1} \mathbf{y} \tag{31}$$

and  $\{u_i\}, \{\lambda_i\}$  are the eigenvectors and eigenvalues of the normal matrix  $N = A^T C_v^{-1} A$ 

$$\mathbf{N} = \mathbf{U} \Lambda \mathbf{U}^{\mathrm{T}} = \sum_{i=1}^{n} \lambda_{i} \mathbf{u}_{i} \mathbf{u}_{i}^{\mathrm{T}}.$$
(32)

In ill-conditioned inverse problems one, or more, of the eigenvalues  $\{\lambda_i\}$  have very small magnitudes, thus causing problems of numerical instability and significant noise amplification in the LS solution from eq. (30).

In order to overcome these problems and obtain a stable solution with satisfactory accuracy, the classic LS estimator is modified according to the general scheme

$$\hat{\mathbf{x}}' = \sum_{i=1}^{n} \delta_i \frac{(\mathbf{u}_i^{\mathrm{T}} \mathbf{b})}{\lambda_i} \mathbf{u}_i.$$
(33)

The role of the filter factors  $\{\delta_i\}$  is to stabilize the inversion of the input data, while improving the overall accuracy of the estimated model parameters. The above filtering mechanism provides a common view on various regularization methods in ill-conditioned inverse problems, whose individual characteristics are dictated by the form of their corresponding filter factors. The analytical expressions of  $\{\delta_i\}$  for many popular regularization techniques used in geodetic and geophysical inverse problems, including Tikhonov–Philips regularization, GRR, LS collocation, TSVD, Landweber iteration method and conjugate gradient iteration method, can be found in Bouman (1998).

In the next section, it is shown that the CV-adaptive estimator from Section 2 corresponds to an eigenvalue filtering of the type given in eq. (33) for a particular 'restrictive' choice of the error CV matrix  $C_{\hat{s}}$  for the model parameters.

# 3.2 CV-adaptive regularization in the LS-based eigenvector basis

The problem of selecting the CV matrix of the estimated model parameters for the implementation of the CV-adaptive estimator  $\hat{\mathbf{x}}$  can be simplified if the choice of  $C_{\hat{\mathbf{x}}}$  is restricted to an *a priori* general form. Specifically, we adopt the following scheme:

$$\mathbf{C}_{\hat{\mathbf{x}}} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{T}},\tag{34}$$

where **U** is the unitary matrix associated with the eigenvalue decomposition of the normal matrix  $\mathbf{N} = \mathbf{A}^{T} \mathbf{C}_{\mathbf{v}}^{-1} \mathbf{A}$  (see eq. 32), and  $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$  is an arbitrary diagonal matrix with nonnegative elements. In this way, the CV matrix  $\mathbf{C}_{\mathbf{s}}$  is constrained to have the same eigenvectors as the normal matrix of the underlying linear model, and its *a priori* selection is reduced into a problem of specifying only its eigenvalues. Based on eq. (34), the optimal regularization matrix from eq. (6) or (7) takes the form

$$\mathbf{R} = \mathbf{U}(\mathbf{D}^{1/2} \Lambda^{1/2}) \mathbf{U}^{\mathrm{T}},\tag{35}$$

where  $\Lambda$  is the known eigenvalue matrix of the normal matrix N. In this case, the auxiliary matrix  $\mathbf{Q} = (\mathbf{R} - \mathbf{I})^{\mathrm{T}} (\mathbf{R} - \mathbf{I})$  is

$$\mathbf{Q} = \mathbf{U}(\mathbf{D}^{1/2} \Lambda^{1/2} - \mathbf{I})^2 \mathbf{U}^{\mathrm{T}}$$
(36)

and its eigenvalues  $\{q_i\}$  are given by the general analytic formula

$$q_i = \left(\sqrt{d_i\lambda_i} - 1\right)^2 = \left(1 - \sqrt{d_i\lambda_i}\right)^2. \tag{37}$$

Furthermore, the CV-adaptive estimator can be expressed in terms of the spectral expansion

$$\hat{\mathbf{x}} = \mathbf{R} \hat{\mathbf{x}}^{LS}$$

$$= \mathbf{R} \mathbf{N}^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{C}_{\mathbf{v}}^{-1} \mathbf{y}$$

$$= \mathbf{U} (\mathbf{D}^{1/2} \Lambda^{1/2}) \mathbf{U}^{\mathrm{T}} \mathbf{U} \Lambda^{-1} \mathbf{U}^{\mathrm{T}} \mathbf{A}^{\mathrm{T}} \mathbf{C}_{\mathbf{v}}^{-1} \mathbf{y}$$

$$= \mathbf{U} (\mathbf{D}^{1/2} \Lambda^{-1/2}) \mathbf{U}^{\mathrm{T}} \mathbf{b}$$

$$= \mathbf{U} (\mathbf{D}^{1/2} \Lambda^{-1/2}) \mathbf{U}^{\mathrm{T}} \mathbf{b}$$

$$= \sum_{i=1}^{n} \frac{\sqrt{d_{i}} (\mathbf{u}_{i}^{\mathrm{T}} \mathbf{b})}{\sqrt{\lambda_{i}}} \mathbf{u}_{i}$$
(38)

or equivalently

$$\mathbf{\hat{x}} = \sum_{i=1}^{n} \sqrt{d_i \lambda_i} \frac{(\mathbf{u}_i^{\mathrm{T}} \mathbf{b})}{\lambda_i} \mathbf{u}_i,$$
(39)

where the data-dependent vector  $\mathbf{b}$  has been defined in eq. (31).

It is seen that the general choice of eq. (34) leads to a CV-adaptive estimator which is equivalent to an eigenvalue filtering of the normal matrix in the underlying linear model. As such, it can cover all regularization techniques that are based on the spectral expansion of eq. (33).

The filter factors  $\{\delta_i\}$  associated with the CV-adaptive estimator have the general form

$$\delta_i = \sqrt{d_i \lambda_i} \tag{40}$$

and they depend uniquely on the chosen eigenvalues  $\{d_i\}$  for the CV matrix  $C_{\hat{s}}$ .

By inverting the previous formula, we get

$$d_i = \frac{\delta_i^2}{\lambda_i} \tag{41}$$

which establishes a one-to-one connection between existing regularized estimators that use specific parametrized forms for their filter factors  $\{\delta_i\}$ , and the CV-adaptive estimator resulting from the CV-matrix constraint in eq. (34).

A regularization scenario of particular interest within the context of the CV-adaptive estimator is to set all eigenvalues  $\{d_i\}$  equal to the same positive value  $\sigma^2$ . Such a choice implies an *a priori* diagonalization constraint for the CV matrix  $C_{\hat{s}}$  ('whitening' of the parameter estimation errors) and it results in a decorrelated solution with uniform precision for the model parameters. It is actually interesting that this type of solution can be equivalently obtained through the well-known method of GRR by properly selecting its built-in regularization parameters. In general, the filter factors associated with the GRR method are given by the formula (Bouman 1998)

$$\delta_i = \frac{\lambda_i}{\lambda_i + k_i},\tag{42}$$

Table 1.	The analytic forms	of the filter factors,	the eigenvalues of	of the CV m	atrix for the	estimated mo	odel parameters a	nd the maximum	eigenvalue of th
auxiliary	matrix Q, for some	common regularizati	ion techniques in	linear mode	l inversion.				

Regularization method	Filter factors	Eigenvalues of the CV matrix $C_{\hat{x}}$	Maximum eigenvalue of $\mathbf{Q}$ $q_{\text{max}} = \max(1 - \sqrt{d_i \lambda_i})^2$
Tikhonov–Philips <sup>a</sup>	$\delta_i = \frac{\lambda_i}{\lambda_i + k}$	$d_i = \frac{\lambda_i}{\lambda_i + k}$	$q_{\max} = \frac{k^2}{(\lambda_{\min}+k)^2}$
Generalized ridge regression <sup>b</sup>	$\delta_i = rac{\lambda_i}{\lambda_i + k_i}$	$d_i = rac{\lambda_i}{(\lambda_i + k_i)^2}$	$q_{\max} = \max_{k_i, \lambda_i} \frac{k_i^2}{(\lambda_i + k_i)^2}$
TSVD <sup>c</sup>	$\delta_i = \begin{cases} 1, i = 1, \dots, r \\ 0, i = r + 1, \dots, n \end{cases}$	$d_{i} = \begin{cases} 1/\lambda_{i}, i = 1, \dots, r\\ 0, i = r+1, \dots, n \end{cases}$	$q_{\max} = 1$
Shrunken LS estimation <sup>d</sup> Parameter decorrelation with uniform precision level <sup>e</sup>	$\delta_i = a \\ \delta_i = \sigma \sqrt{\lambda_i}$	$d_i = a^2 / \lambda_i d_i = \sigma^2$	$q_{\max} = (1 - a)^2$ $q_{\max} = (1 - \sigma \sqrt{\lambda_{\min}})^2$

<sup>a</sup>The symbol k denotes the scalar regularization parameter used in Tikhonov–Philips regularization (Schwarz 1979).

<sup>b</sup>The symbols  $\{k_i\}$  correspond to the set of regularization parameters associated with the implementation of generalized ridge regression (Xu & Rummel 1994).

<sup>c</sup>The symbol r corresponds to the rank truncation parameter in the TSVD regularization method (Xu 1998).

<sup>d</sup>The symbol *a* corresponds to the shrinkage factor of the LS solution for the implementation of the shrunken LS estimator (Mayer & Willke 1973).

<sup>*e*</sup> The symbol  $\sigma$  denotes the common standard deviation for the estimation errors in the parameter decorrelation regularization method (see this section).

where  $\{k_i\}$  is a set of regularization parameters that appear in the GRR solution

$$\hat{\mathbf{x}}^{\text{GRR}} = (\mathbf{A}^{\text{T}} \mathbf{C}_{\mathbf{v}}^{-1} \mathbf{A} + \mathbf{U} \mathbf{K} \mathbf{U}^{\text{T}})^{-1} \mathbf{A}^{\text{T}} \mathbf{C}_{\mathbf{v}}^{-1} \mathbf{y}$$
(43)

and  $\mathbf{K} = \operatorname{diag}(k_1, \ldots, k_n)$ ; see Xu & Rummel (1994).

If we set

$$k_i = \frac{\sqrt{\lambda_i}(1 - \sigma\sqrt{\lambda_i})}{\sigma} \tag{44}$$

for some chosen positive  $\sigma$  that satisfies  $\sigma < 1/\sqrt{\lambda_i}$  for every eigenvalue  $\lambda_i$  of the normal matrix **N**, then the GRR filter factors from eq. (42) become

$$\delta_i = \sigma \sqrt{\lambda_i}.\tag{45}$$

The equivalent filtering implementation of the CV-adaptive estimator  $\hat{\mathbf{x}}$ , according to eqs (39) and (40), requires that the eigenvalues  $\{d_i\}$  of  $\mathbf{C}_{\hat{\mathbf{x}}}$  should be chosen as

$$d_i = \frac{\delta_i^2}{\lambda_i} = \frac{\left(\sigma \sqrt{\lambda_i}\right)^2}{\lambda_i} = \sigma^2$$
(46)

which in turn implies the diagonalization constraint  $C_{\hat{x}} = UDU^{T} = \sigma^{2}I$ .

In Table 1, the analytic forms of the filter factors  $\{\delta_i\}$ , the corresponding eigenvalues  $\{d_i\}$  of the CV matrix  $\mathbf{C}_{\hat{\mathbf{x}}}$ , and the maximum eigenvalue  $q_{\text{max}}$  of the auxiliary matrix  $\mathbf{Q}$ , are provided for some known regularization methods (see also Bouman 1998). Note that all regularized estimators shown in Table 1 can be considered as special cases of the CV-adaptive estimator  $\hat{\mathbf{x}}$ , for a particular *a priori* constraint on the CV matrix of the estimated parameters drawn from the general choice  $\mathbf{C}_{\hat{\mathbf{x}}} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{T}}$ .

# Remark 4:

A more general regularization scheme can emerge from the use of the CV-adaptive estimator if the chosen eigenvectors of  $C_{\hat{x}}$  are different from the eigenvectors of the normal matrix  $N = A^T C_v^{-1} A$ . In this case, the optimal approximation of the unknown parameter vector is performed in a different orthogonal basis than the one used in traditional regularization schemes which rely on the eigenvalue filtering with respect to the eigenvector basis associated with the normal matrix N (see Section 3.1). In the context of the present paper, the use of such a different 'regularization basis' should be linked to the ability of the CV-adaptive estimator to impose an arbitrary structure on the error variances and covariances of the inversion results. As an example, we can mention the choice  $\mathbf{C}_{\hat{\mathbf{x}}} = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$  with  $\sigma_1^2 \neq \cdots \neq \sigma_n^2$ , which gives a decorrelated solution with different error variances for the estimated parameters.

## **4 NUMERICAL EXAMPLE**

A simulated numerical experiment has been performed to demonstrate the practical implementation and the usefulness of the proposed CV-adaptive regularization scheme. This particular example has been motivated by a similar experiment that is reported in Xu (1998) and it deals with a typical ill-posed inverse problem of downward continuation of gravity field functionals. More specifically, we consider the determination of a gravity anomaly grid on the surface of the Earth using gradiometric observables collected at satellite altitude of 200 km. The basic mathematical model that is used for the design of the experiment is given by the integral equation

$$T_{rr} = \frac{R}{4\pi} \iint_{\sigma_s} \Delta g S''(r, \psi) \mathrm{d}\sigma_s \tag{47}$$

where  $T_{rr}$  is the second radial derivative of the disturbing potential, R is the mean radius of the Earth,  $\Delta g$  is the gravity anomaly signal, and  $\sigma_s$  denotes the unit sphere. The kernel function  $S''(r, \psi)$  corresponds to the second derivative of the generalized Stokes function with respect to the geocentric radial distance r (for more details, see Xu 1998).

Assuming that we have available *n* discrete gradiometric observations of  $T_{rr}$  (e.g. via a satellite gravity mission such as GOCE), the integral formula in eq. (47) can be discretized according to the following linear model

$$(4\pi R/10)T_{rr}^{i} = \sum_{k} \bar{S}''(r_{i}, \psi_{ik})\Delta\sigma_{k}\Delta g_{k}, \quad i = 1, 2, \dots, n \quad (48)$$

(see e.g. Xu & Rummel 1995; Xu 1998), where  $\Delta g_k$  are blockmean gravity anomalies, the constant  $(4\pi R/10)$  originates from the generalized Stokes' formula and unit regulation, and  $\Delta \sigma_k$  and  $\bar{S}''(r, \psi)$  are given by

$$\Delta \sigma_k = \iint_{\sigma_k} \mathrm{d}\sigma \tag{49}$$

**Table 2.** Statistics of the true (simulated) block-mean gravity anomalies  $\Delta g$ , in comparison with the statistics of the estimated block-mean gravity anomalies obtained from: (i) standard LS inversion ( $\Delta g^{LS}$ ) and (ii) regularized decorrelation ( $\Delta g^R = \mathbf{R} \Delta g^{LS}$ ) for varying values of the *a priori* constrained error variance (all values in mgal).

	Max	Min	Mean	Sigma	rms
True field	83.970	-74.643	1.231	24.716	24.746
Standard LS inversion	118.651	-101.403	1.166	53.041	53.054
Regularized decorrelation with $\sigma^2 = 4 \text{ mgal}^2$	160.642	-83.584	9.250	58.856	59.578
Regularized decorrelation with $\sigma^2 = 1 \text{ mgal}^2$	80.321	-41.792	4.625	29.428	29.789
Regularized decorrelation with $\sigma^2 = 0.25 \text{ mgal}^2$	40.161	-20.896	2.312	14.714	14.895

$$\bar{S}''(r,\psi) = t^{3} \{(1-t\cos\psi) \left[ \frac{3(1-t^{2})}{D^{3}} - 4/D^{3} \right] -\frac{1+t^{2}}{D^{3}} - 10/D - 18D + 2 -3t\cos\psi \left( 15 + 6\ln\frac{1-t\cos\psi + D}{2} \right)$$
(50)

$$t = R/r \tag{51}$$

$$D = (1 - 2t\cos\psi + t^2)^{1/2}.$$
(52)

Using the discretized model of eq. (48), a linear system of observation equations  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$  can be formed, where  $\mathbf{y}$  is the data vector with elements  $(4\pi R/10)T_{rr}^{i}$ ,  $\mathbf{A}$  is the design matrix whose elements correspond to the products  $\overline{S''}(r_i, \psi_{ik})\Delta\sigma_k$ ,  $\mathbf{x}$  is the unknown vector of surface block-mean gravity anomalies  $\Delta g_k$ , and  $\mathbf{v}$  is the data noise vector.

For our experiment, true  $1^\circ\times1^\circ$  block-mean gravity anomaly values have been simulated within an  $10^\circ\times10^\circ$  test area, using the Hirvonen CV function model

$$C_{\Delta g}(\psi) = \frac{C_o}{1 + (\psi/a)^2},$$
 (53)

where  $C_o = 675 \text{ mgal}^2$ , and the parameter *a* has been selected such that the correlation length of the gravity anomaly field is  $\xi =$ 15.5 km. The number of satellite gradiometric observables  $T_{rr}^i$  is 400, taken on an 30' × 30' grid that is located directly above the area of interest at a constant altitude of r = 200 km. The accuracy of  $T_{rr}^i$  is set to 0.01 E (1 E = 10<sup>-9</sup> s<sup>-2</sup>), which was also employed in similar simulations by Koop (1993), Xu & Rummel (1995) and Xu (1998). The random errors **v** for the determination of the synthetic observations from the simulated gravity anomalies are generated according to  $(4\pi R/10) \text{ N}(T_{rr}^i)$ , where  $\text{N}(T_{rr}^i)$  is a Gaussian random number generator with mean zero and variance  $10^{-4}\text{E}^2$ .

We shall compare two different estimators for the downward continuation of the gradiometric data and the determination of the surface gravity anomaly grid. First, the inversion of the gradiometric measurements is performed using the standard LS methodology, without applying any special regularization tool. The condition number of the normal matrix is  $1.93 \times 10^6$ , and its eigenvalues range from  $1.58 \times 10^{-5}$  to 30.57. As expected, the accuracy of the LS inversion results is considerably poor, giving an average error standard deviation of  $\pm 50$  mgal in the estimated gravity anomaly values. The detailed statistics of the LS estimation errors are shown in Table 3, where it is seen that extreme errors of more than  $\pm 120$  mgal occur during the downward continuation of the gradiometric data. In Table 2, we see the statistics of the estimated  $\Delta g$  values which exhibit much stronger variability than the original true values, thus indicating the numerical instability of the downward continuation process through a simple LS inversion algorithm.

Alternatively, the inversion of the gradiometric data  $T_{rr}^{i}$  is performed using the CV-adaptive regularization scheme by constraining the CV matrix of the estimated gravity anomalies to a specific *a priori* model. In particular, we examine three different constraint choices all of which correspond to a diagonal matrix of the general form  $C_{\hat{s}} = \sigma^2 I$ . In this way, it is ensured that the parameter estimation errors during the downward continuation process become uncorrelated (i.e. 'whitening' of the estimated gravity anomalies) and they have a common user-dependent variance. The value of  $\sigma$ has been set equal to 2, 1 and 0.5 mgal, respectively. The results obtained from this CV-adaptive decorrelating inversion scheme are presented in Tables 2 and 3.

From the error statistics shown in Table 3, we see that there is a clear improvement in the downward continuation results when using the CV-adaptive regularized estimator. By setting the error standard deviation in the estimated gravity anomalies  $\sigma = 1$  mgal, for example, the rms error drops from 48.53 mgal (standard LS inversion) to 27.75 mgal. There is of course a remaining bias in the estimated gravity anomalies that amounts to about 3.4 mgal on average, which is nevertheless well below the statistical uncertainty of the results. It must be noted that, as the value of the a priori constrained error variance  $\sigma^2$  becomes smaller, the corresponding bias in the results of downward continuation decreases considerably (see Table 3). In this way, when we set  $\sigma = 0.5$  mgal, then the rms estimation error is reduced to 21.67 mgal and the average bias drops to approximately 1 mgal. However, the trade-off in this case lies on the oversmoothing that takes place in the resulting grid of block-mean gravity anomalies, as seen in the values of Table 2.

# **5 SUMMARY AND CONCLUSIONS**

A new approach for constructing optimal parameter estimators for the inversion of linear models has been presented in this paper. Our methodology is based on a linear transformation of the classic LS estimator and it adheres to two basic criteria. First, it provides a solution that is optimally fitted (in an average quadratic sense) to the classic LS parameter solution. Secondly, it complies with an external constraint that specifies *a priori* the error CV matrix for the estimated model parameters. The MSEE performance of the new estimator, as well as its optimality with respect to the maximum size of its inherent bias, have been explained and analysed. Furthermore, an expression for the admissible SNR range of the underlying linear model has been derived, within which the new estimator always outperforms (in the MSEE sense) the usual LS solution.

The paper does not claim the development of an autonomous regularization method that competes against other well-known inversion techniques, such as Tikhonov–Philips regularization, GRR or TSVD estimation. The aim of the paper was to present a certain theoretical formulation, based purely on statistical concepts and criteria, for the optimal inversion of linear models that leads to a general CV-adaptive biased estimator. It has been shown that the

**Table 3.** Statistics of the estimation errors in the block-mean gravity anomalies obtained from (i) standard LS inversion of the gradiometric data and (ii) regularized decorrelation for varying values of the *a priori* constrained error variance. The statistics of the estimation bias in the downward continuation results that are obtained from the regularized decorrelation inversion scheme are also shown. Note that the bias vector for the whole set of the block-mean gravity anomalies is determined as BIAS( $\Delta g^R$ ) = (**R** – **I**)**x**, where **x** contains the true (simulated) values and **R** corresponds to the optimal regularization matrix for each *a priori* choice of  $\sigma^2$  (all values in mgal).

	Max	Min	Mean	Sigma	rms
$\Delta g^{LS} - \Delta g$	122.227	-104.609	-0.064	48.530	48.530
Regularized decorrelation with a	priori constraine	ed error std $\sigma = 2$	mgal		
$\begin{array}{l} \Delta g^{R} - \Delta g \\ \Delta g^{R} - \Delta g - \text{BIAS}(\Delta g^{R}) \\ \text{BIAS}(\Delta g^{R}) \end{array}$	146.599 4.254 147.035	-88.916 -4.324 -87.695	8.019 -0.269 8.289	51.365 1.882 51.326	51.987 1.901 51.991
Regularized decorrelation with a	priori constraine	ed error std $\sigma = 1$	mgal		
$\begin{array}{l} \Delta g^{R} - \Delta g \\ \Delta g^{R} - \Delta g - \text{BIAS}(\Delta g^{R}) \\ \text{BIAS}(\Delta g^{R}) \end{array}$	70.607 2.127 70.826	-63.751 -2.162 -63.140	3.394 -0.135 3.529	27.543 0.941 27.532	27.751 0.951 27.757
Regularized decorrelation with a	priori constraine	ed error std $\sigma = 0$	.5 mgal		
$\Delta g^{R} - \Delta g$ $\Delta g^{R} - \Delta g - \text{BIAS}(\Delta g^{R})$ $\text{BIAS}(\Delta g^{R})$	56.137 1.063 57.093	-51.168 -1.081 -50.863	1.082 - 0.067 1.149	21.639 0.471 21.640	21.666 0.475 21.670

spectral representation of this parameter estimator offers a unified treatment for existing regularization methods in ill-posed inverse problems, particularly for those techniques that are based on the eigenvalue filtering of the inverse normal matrix. Our study provides an alternative viewpoint on the optimal statistical properties and the regularization mechanism of many inversion techniques commonly used in geodesy, by interpreting them as a family of 'CV-adaptive' parameter estimators that obey the same optimal criterion (i.e. optimal fitting with the LS solution in the mean squared sense) and differ only on the pre-selected form of their error CV matrix under a fixed model design.

Further studies and experimentation is obviously needed for the assessment of the practical potential associated with our proposed theoretical methodology in linear model inversion. The results obtained from a simulation experiment that was presented in this study certainly provide some encouragement for exploring new innovative strategies in ill-posed geodetic inverse problems (e.g. regularized decorrelation or optimal 'whitening' of the estimated model parameters).

# ACKNOWLEDGMENTS

The author would like to acknowledge the valuable suggestions and comments on the original manuscript by two anonymous reviewers, which led to an improved revised version of this paper.

# REFERENCES

- Aki, K. & Richards, P.G., 1980. Quantitative Seismology, Theory and Methods, W.H. Freeman, San Francisco.
- Backus, G.E., 1988. Bayesian inference in geomagnetism, *Geophys. J. Int.*, 92, 125–142.
- Bee, M. & Jacobson, R.S., 1984. Linear inversion of body wave data, *Geophysics*, 49, 2088–2093.
- Beutler, G., Brockmann, E., Gurtner, W., Hugentobler, U., Mervart, L., Rothacher, M. & Verdun, A., 1994. Extended orbit modeling techniques at the CODE Processing Center of the International GPS Service for Geodynamics (IGS): theory and initial results. *Manusc. Geod.*, 19, 367–385.

Bibby, J. & Toutenburg, H., 1977. Prediction and Improved Estimation in Linear Models, John Wiley & Sons, Chichester.

© 2007 The Author, *GJI*, **171**, 509–522 Journal compilation © 2007 RAS

Bouman, J., 1998. *Quality of Regularization Methods*. DEOS Report No. 98.2, Delft University Press, Delft, The Netherlands.

- Bouman, J. & Koop, R., 1997. Quality differences between Tikhonov regularization and generalized biased estimation in gradiometric analysis, *DEOS Progr. Lett.*, 97, 42–48.
- Chase, E.P. & Stuart, G.S., 1972. The N plate problem of plate tectonics, *Geophys. J. Roy. Astron. Soc.*, 29, 117–122.
- Crosson, R.S., 1976. Crustal structure modeling of earthquake data: simultaneous least squares estimation of hypocenter and velocity parameters, *J. Geophys. Res.*, **81**, 3036–3046.
- Dermanis, A. & Rummel, R., 2000. Data analysis methods in geodesy, in *Geomatic Methods for the Analysis of Data in Earth Sciences*, pp. 17–92, eds Dermanis, A., Grun, A. & Sanso, F., Springer-Verlag, Berlin.
- Ditmar, P., Kusche, J. & Klees, R., 2003. Computation of spherical harmonic coefficients from gravity gradiometry data to be acquired by the GOCE satellite: regularization issues, *J. Geod.*, 77, 465–477.
- Eldar, Y.C. & Oppenheim, A.V., 2003. Covariance shaping least-squares estimation. *IEEE Trans. Signal Proc.*, **51**(3), 686–697.
- Engelis, T. & Knudsen, P., 1989. Orbit improvement and determination of the ocean geoid and topography from 17 days of Seasat data, *Manusc. Geod.*, 14, 193–201.
- Fu, L-L. & Cazenave, A. eds, 2001. *Satellite Altimetry and Earth Sciences,* Academic Press, San Diego.
- Gerstl, M. & Rummel, R., 1981. Stability investigations of various representations of the gravity field, *Rev. Geophys. Space Phys.*, **19**, 415–420.
- Gire, C., Lemovel, J.L. & Madden, T., 1986. Motions at the core surface derived from SV data, *Geophys. J. Roy. Astron. Soc.*, 84, 1–29.
- Harville, D.A., 1997. *Matrix Algebra from a Statistician's Perspective,* Springer Verlag, Berlin Heidelberg.
- Heiskanen, W.A. & Moritz, H., 1967. *Physical Geodesy*, W. H. Freeman and Company, San Francisco.
- Hoerl, A.E. & Kennard, R.W., 1970. Ridge regression: biased estimation for nonorthogonal problems, *Technometrics*, 12, 55–67.
- Ilk, K.H., 1991. Inverse geodetic problems, in *Contributions to Geodetic Theory and Methodology*, pp. 127–161, ed. Grafarend, E.W., IAG Section IV, IUGG General Assembly, Vienna, August 11–24.
- Ilk, K.H., Kusche, J. & Rudolph, S., 2002. A contribution to data combination in ill-posed downward continuation problems, J. Geodyn., 33, 75– 99.
- Kleijer, F., 2004. Troposphere modeling and filtering for precise GPS leveling, in *Report of the Netherlands Geodetic Commission*, Vol. 56, Delft, The Netherlands.

- Knudsen, P., 1992. Altimetry for geodesy and oceanography, *Reports of the Finnish Geodetic Institute*, **115**, 87–129.
- Koch, M., 1985. A numerical study on the determination of the 3D structure of the lithosphere by linear and nonlinear inversion of teleseismic travel times, *Geophys. J. Roy. Astron. Soc.*, **80**, 73–93.
- Koch, K-R., 1999. Parameter Estimation and Hypothesis Testing in Linear Models, 2nd edn, Springer-Verlag, Berlin.
- Koch, K-R. & Kusche, J., 2002. Regularization of geopotential determination from satellite data by variance components, J. Geodyn., 76, 259–268.
- Koop, R., 1993. Global gravity field modeling using satellite gravity gradiometry. *PhD dissertation*. Neth Geod Comm Publ, 38, Delft, The Netherlands.
- Kunitake, M., 1996. Ionospheric tomography, in *Inverse Methods*, pp. 231– 238, eds Jacobsen, K., Mosegaard, B. & Sibani, P., Springer-Verlag, Berlin.
- Kusche, J. & Klees, R., 2002. Regularization of gravity field estimation from satellite gravity gradients, J. Geodyn., 76, 359–368.
- Marquardt, D.W., 1970. Generalized inverses, ridge regression, biased linear estimation and non-linear estimation. *Technometrics*, **12**(3), 591–612.
- Mayer, L.S. & Willke, T.A., 1973. On biased estimation in linear models, *Technometrics*, **15**(3), 497–508.
- Menke, W., 1989. Geophysical Data Analysis: Discrete Inverse Theory, Academic Press, San Diego.
- Minster, J.B., Jordan, T.H., Molnar, P. & Haines, E., 1974. Numerical modeling of instantaneous plate tectonics, *Geophys. J. Roy. Astron. Soc.*, 36, 541–576.
- Nataf, H.C., Nakanishi, I. & Anderson, D.L., 1986. Measurement of mantle wave velocities and inversion for lateral heterogeneity and anisotropy, J. Geophys. Res., 91, 7261–7308.
- Neyman, Y.M., 1985. Improperly posed problems in geodesy and methods for their solution, in *Proceedings of the Beijing International Summer School on Local Gravity Field Approximation*, pp. 499–566, ed Schwarz, K-P., Beijing, China, Aug 21–Sept 4, 1984.
- Papoulis, A., 1991. *Probability, Random Variables and Stochastic Processes,* 3rd edn, McGraw-Hill, New York.
- Parker, R.L., 1970. The inverse problem of the electrical conductivity of the mantle, *Geophys. J. Roy. Astron. Soc.*, 22, 121–138.
- Parker, R.L., Shure, L. & Hildebrand, J., 1987. An application of inverse theory to seamount magnetism, *Rev. Geophys.*, 25, 17–40.
- Pavlis, G.L. & Booker, J.R., 1980. The mixed discrete-continuous inverse problem: application to the simultaneous determination of earthquake hypocenters and velocity structures, *J. Geophys. Res.*, 85, 4801–4809.
- Rao, C.R., 1973. Linear Statistical Inference and its Applications, John Wiley & Sons, Chichester.

- Rao, C.R. & Toutenburg, H., 1999. Linear Models: Least Squares and Alternatives, 2nd edn, Springer-Verlag, Berlin.
- Rummel, R., 1985. From the observational model to gravity parameter estimation, in *Proceedings of the Beijing International Summer School on Local Gravity Field Approximation*, pp. 67–106, ed. Schwarz, K.-P., Beijing, China, Aug 21–Sept 4, 1984.
- Rummel, R., Hajela, D.P. & Rapp, R.H., 1976. Recovery of mean gravity anomalies from satellite-to-satellite range rate data using least-squares collocation, OSU Report No. 248, Department of Geodetic Science, Columbus, Ohio.
- Rummel, R., Schwarz, K.-P. & Gerstl, M., 1979. Least squares collocation and regularization, *Bull. Geodyn.*, 53, 343–361.
- Sanso, F. & Rummel, R. (eds), 1989. Theory of Satellite Geodesy and Gravity Field Determination, Springer-Verlag, Berlin.
- Schaffrin, B., 1980. Tikhonov-regularization in geodesy: an example, *Boll. Geodyn. Sci. Affi.*, **39**, 207–216.
- Schaffrin, B., 1985. Das geodätische datum mit stochastischer vorinformation, Habilitationsschrift, Universität Stuttgart, Deutsche Geodätische Kommission, Reihe C, Heft Nr. 313, München.
- Schaffrin, B., 1989. An alternative approach to robust collocation, Bull. Geodyn., 63, 395–404.
- Schwarz, K.P., 1979. Geodetic improperly posed problems and their regularization, Boll. Geodyn. Sci. Affi., 38, 389–416.
- Sen, A. & Srivastava, M., 1990. Regression Analysis: Theory, Methods and Applications, Springer-Verlag, Berlin.
- Strang, G., 1988. *Linear Algebra and its Applications*, 3rd edn, Harcourt Brace & Company, Orlando.
- Tapley, B.D., Schutz, B.E. & Born, G.H., 2004. Statistical Orbit Determination, Elsevier Academic Press, Burlington, MA.
- Tarantola, A., 1987. *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier, Amsterdam.
- Tikhonov, A.N. & Arsenin, V.Y., 1977. *Solutions of Ill-Posed Problems*, John Wiley & Sons, New York.
- Wunsch, C. & Minster, J.F., 1982. Methods for box models and ocean circulation tracers: mathematical programming and nonlinear inverse theory, *J. Geophys. Res.*, 87, 5647–5662.
- Xu, P., 1992. The value of minimum norm estimation in geopotential fields, *Geophys. J. Int.*, **111**, 170–178.
- Xu, P., 1998. Truncated SVD methods for discrete linear ill-posed problems, *Geophys. J. Int.*, 135, 505–514.
- Xu, P. & Rummel, R., 1994. Generalized ridge regression with applications in determination of potential fields, *Manusc. Geodyn.*, 20, 8–20.
- Xu, P. & Rummel, R., 1995. A simulation study of smoothness methods in recovery of regional gravity fields, *Geophys. J. Int.*, **117**, 472–486.

# APPENDIX A: DETERMINATION OF THE OPTIMAL REGULARIZATION MATRIX

Given the usual LS estimator  $\hat{\mathbf{x}}^{LS}$  and its associated CV matrix  $\mathbf{C}_{\hat{\mathbf{x}}^{LS}} = \mathbf{N}^{-1}$  for an unknown parameter vector  $\mathbf{x}$  in the linear model  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$ , we seek an optimal regularization matrix  $\mathbf{R}$  such that the *biased* estimator

$$\hat{\mathbf{x}} = \mathbf{R} \hat{\mathbf{x}}^{LS}$$
(A1)  
minimizes the quadratic form  

$$\varphi = E\{(\hat{\mathbf{x}} - \hat{\mathbf{x}}^{LS})^{T} \mathbf{N}(\hat{\mathbf{x}} - \hat{\mathbf{x}}^{LS})\}$$
(A2)  
and satisfies the following CV-adaptive constraint  

$$\mathbf{R} \mathbf{N}^{-1} \mathbf{R}^{T} = \mathbf{C}_{\hat{\mathbf{x}}},$$
(A3)

where  $C_{\hat{x}}$  is a given symmetric positive-definite matrix. To a large extent, the proof that follows adheres to a similar derivation given in Eldar & Oppenheim (2003).

Let us first apply an auxiliary invertible transformation to the vectors  $\hat{\mathbf{x}}$  and  $\hat{\mathbf{x}}^{LS}$ , as follows

$$\hat{\bar{\mathbf{x}}} = \mathbf{N}^{1/2} \hat{\mathbf{x}} \tag{A4}$$

$$\hat{\mathbf{x}}^{LS} = \mathbf{N}^{1/2} \hat{\mathbf{x}}^{LS}. \tag{A5}$$

In this way, the original quadratic form in eq. (A2) can be equivalently written as	
$\varphi = E\{(\mathbf{\hat{x}} - \mathbf{\hat{x}}^{LS})^{T}(\mathbf{\hat{x}} - \mathbf{\hat{x}}^{LS})\}.$	(A6)
The CV matrices of the transformed vectors $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}^{LS}$ will be given by the equations	
$\mathbf{C}_{\mathbf{\hat{x}}} = \mathbf{N}^{1/2} \mathbf{C}_{\mathbf{\hat{x}}} \mathbf{N}^{1/2}$	(A7)
$\mathbf{C}_{\mathbf{\hat{x}}^{LS}} = \mathbf{N}^{1/2} \mathbf{C}_{\mathbf{\hat{x}}^{LS}} \mathbf{N}^{1/2} = \mathbf{N}^{1/2} \mathbf{N}^{-1} \mathbf{N}^{1/2} = \mathbf{I}.$	(A8)
Following from eq. (A1), the relationship between the transformed estimators $\mathbf{\hat{x}}$ and $\mathbf{\hat{x}}^{LS}$ is expressed in terms of the linear equation	1
$\mathbf{\hat{x}} = \mathbf{\bar{R}}\mathbf{\hat{x}}^{LS},$	(A9)
where the matrix $\mathbf{\bar{R}}$ is related to the original regularization matrix $\mathbf{R}$ in terms of the transformation formula	
$\mathbf{\bar{R}} = \mathbf{N}^{1/2} \mathbf{R} \mathbf{N}^{-1/2}.$	(A10)
Let us consider the eigenvalue decomposition of the CV matrix given in eq. (A7)	
$\mathbf{C}_{\mathbf{\hat{x}}} = \mathbf{N}^{1/2} \mathbf{C}_{\mathbf{\hat{x}}} \mathbf{N}^{1/2} = \mathbf{V} \mathbf{D} \mathbf{V}^{\mathrm{T}},$	(A11)
where V is a unitary matrix whose columns correspond to the orthonormalized eigenvectors of $C_{\hat{x}}$ , and D is a diagonal matrix that the eigenvalues of $C_{\hat{x}}$	t contains
Using the unitary matrix V, an auxiliary invertible transformation will now be applied to the vectors $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}^{LS}$ , as follows	
$\hat{ ilde{\mathbf{x}}} = \mathbf{V}^{ op} \hat{ ilde{\mathbf{x}}}$	(A12)
$\hat{\mathbf{r}}^{LS} = \mathbf{v}^{T} \hat{\mathbf{r}}^{LS}$	(A13)
$\mathbf{x} = \mathbf{v} \cdot \mathbf{x}^T$ . Due to the orthogonality of the matrix $\mathbf{V}^T$ the quadratic form in eq. (A6) can be equivalently expressed as	(1115)
$\omega = E \left\{ \left( \hat{\mathbf{x}} - \hat{\mathbf{x}}^{LS} \right)^{T} \left( \hat{\mathbf{x}} - \hat{\mathbf{x}}^{LS} \right) \right\}$	(A14)
$\varphi = D\left[ \begin{pmatrix} \mathbf{x} & \mathbf{x} \end{pmatrix} \begin{pmatrix} \mathbf{x} & \mathbf{x} \end{pmatrix} \right].$ The CV matrices of both transformed vectors $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}^{LS}$ are now diagonal since	(2111)
$r = v^{T} c v^{Eq.(A1)} p$	(115)
$\mathbf{C}_{\hat{\mathbf{x}}} = \mathbf{V}^{\mathbf{x}} \mathbf{C}_{\hat{\mathbf{x}}} \mathbf{V} = \mathbf{D}$	(A15)
$\mathbf{C}_{\hat{\mathbf{x}}^{LS}} = \mathbf{V}^{\mathrm{T}} \mathbf{C}_{\hat{\mathbf{x}}^{LS}} \mathbf{V} \stackrel{\mathrm{Eq.}(\mathrm{A8})}{=} \mathbf{V}^{\mathrm{T}} \mathbf{I} \mathbf{V} = \mathbf{I}.$	(A16)
Following from eq. (A9), the relationship between the transformed vectors $\hat{\bar{x}}$ and $\hat{\bar{x}}^{LS}$ is given by the linear formula	
$\hat{\bar{\mathbf{x}}} = \bar{\mathbf{R}} \hat{\bar{\mathbf{x}}}^{LS},$	(A17)
where the new matrix $\mathbf{\bar{\bar{R}}}$ is related to the previous transformation matrix $\mathbf{\bar{R}}$ via the equation	
$\mathbf{\bar{\tilde{R}}} = \mathbf{V}^{\mathrm{T}}\mathbf{\bar{R}}\mathbf{V}.$	(A18)
Denoting by $\mathbf{\bar{m}}$ and $\mathbf{\bar{m}}^{LS}$ the expectation vectors of $\mathbf{\hat{\bar{x}}}$ and $\mathbf{\hat{\bar{x}}}^{LS}$ , respectively	
$\mathbf{\bar{m}} = E\{\mathbf{\bar{x}}\}$	(A19)
-LS	( )
$\mathbf{\bar{m}}^{LS} = E\{\mathbf{\bar{\bar{x}}}^{-}\}$	(A20)
the quadratic form in eq. (A14) can be decomposed as follows	
$\varphi = E\{(\bar{\mathbf{x}} - \bar{\mathbf{x}}^{LS})^{T}(\bar{\mathbf{x}} - \bar{\mathbf{x}}^{LS})\}$	
$= E\{(\bar{\mathbf{x}}^{LS})^{T}(\bar{\mathbf{x}}^{LS})\} + E\{(\bar{\mathbf{x}})^{T}(\bar{\mathbf{x}})\} - 2E\{(\bar{\mathbf{x}}^{LS})^{T}(\bar{\mathbf{x}})\}$	
$= (\mathbf{\tilde{m}}^{LS})^{1}(\mathbf{\tilde{m}}^{LS}) + \text{trace}\mathbf{C}_{\mathbf{\tilde{\tilde{x}}}^{LS}} + \rho,$	(A21)
where the ancillary term $\rho$ corresponds to the quantity	
$\rho = E\{(\mathbf{\tilde{x}})^{1}(\mathbf{\tilde{x}})\} - 2E\{(\mathbf{\tilde{x}}^{L3})^{1}(\mathbf{\tilde{x}})\}.$	(A22)
Furthermore, we have	
$(\mathbf{\tilde{m}}^{LS})^{\mathrm{T}}(\mathbf{\tilde{m}}^{LS}) = E\{(\mathbf{\tilde{x}}^{LS})^{\mathrm{T}}\}E\{\mathbf{\tilde{x}}^{LS}\}^{\mathrm{T}}= E\{(\mathbf{V}^{\mathrm{T}}\mathbf{\tilde{x}}^{LS})^{\mathrm{T}}\}E\{\mathbf{V}^{\mathrm{T}}\mathbf{\tilde{x}}^{LS}\}$	
$= E\{(\mathbf{\hat{x}}^{LS})^{T}\}\mathbf{V}\mathbf{V}^{T}E\{\mathbf{\hat{x}}^{LS}\} \stackrel{\text{Lq}(\mathcal{K}^{J})}{=} E\{(\mathbf{N}^{1/2}\mathbf{\hat{x}}^{LS})^{T}\}E\{\mathbf{N}^{1/2}\mathbf{\hat{x}}^{LS}\}$	(A23)
$= E\{(\hat{\mathbf{x}}^{LS})^{\mathrm{T}}\}\mathbf{N}^{1/2}\mathbf{N}^{1/2}E\{\hat{\mathbf{x}}^{LS}\}$	
$= \mathbf{X}^{*} \mathbf{I} \mathbf{N} \mathbf{X}.$	

## 520 C. Kotsakis

Taking into account eq. (A16), the quadratic form  $\varphi$  from eq. (A21) can be now expressed as

$$\varphi = \mathbf{x}^{\mathrm{T}} \mathbf{N} \mathbf{x} + \dim \mathbf{x} + \rho, \tag{A24}$$

where  $\mathbf{x}$  denotes the true vector of the unknown parameters.

The first two terms in the above equation do not depend on the required regularization matrix **R** (or on any of its transformed versions  $\mathbf{\tilde{R}}$  and  $\mathbf{\bar{R}}$ ), and thus the minimization of  $\varphi$  is equivalent to the minimization of the third term  $\rho$  which was defined in eq. (A22).

Taking into account the linearity of the expectation operator  $E\{\cdot\}$ , we can express the quantity  $\rho$  in the alternative form

$$\rho = \sum_{k=1}^{n} E\{(\hat{\bar{x}}[k])^2\} - 2\sum_{k=1}^{n} E\{\hat{\bar{x}}[k]\hat{\bar{x}}^{LS}[k]\},\tag{A25}$$

where  $\hat{\bar{x}}[k]$  and  $\hat{\bar{x}}^{LS}[k]$  denote the *k*th elements of the vectors  $\hat{\bar{\mathbf{x}}}$  and  $\hat{\bar{\mathbf{x}}}^{LS}$ , respectively. Since the auxiliary vector  $\hat{\bar{\mathbf{x}}}^{LS}$  also does not depend on the regularization matrix **R** (or on any of its transformed versions  $\bar{\mathbf{R}}$  and  $\bar{\bar{\mathbf{R}}}$ ), the minimization of  $\varphi$  is thus reduced in finding the elements  $\hat{\bar{x}}[k]$  that minimize the value of  $\rho$  as given in eq. (A25).

In order to find the minimum of  $\rho$  we will use the well-known cosine inequality (Papoulis 1991, p. 154)

$$E\{x_1x_2\} \le \sqrt{E\{x_1^2\}E\{x_2^2\}}$$
(A26)

which is valid for any pair  $(x_1, x_2)$  of real-valued random variables. Note that the equality sign in the previous inequality holds if and only if the random variables are linearly related,  $x_2 = ax_1$ , where *a* is an arbitrary real scalar.

Putting the elements  $\hat{\bar{x}}[k]$  and  $\hat{\bar{x}}^{LS}[k]$  in place of the arbitrary random variables  $x_1$  and  $x_2$ , the cosine inequality of eq. (A26) yields

$$E\{\hat{\bar{x}}[k]\hat{\bar{x}}^{LS}[k]\} \le \sqrt{E\{(\hat{\bar{x}}[k])^2\}E\{(\hat{\bar{x}}^{LS}[k])^2\}}$$
(A27)

or equivalently

$$-2E\{\hat{\bar{x}}[k]\hat{\bar{x}}^{LS}[k]\} \ge -2\sqrt{E\{(\hat{\bar{x}}[k])^2\}E\{(\hat{\bar{x}}^{LS}[k])^2\}}.$$
(A28)

By adding the term  $E\{(\hat{\bar{x}}[k])^2\}$  to both sides of the above inequality, we get

$$E\{(\hat{\bar{x}}[k])^2\} - 2E\{\hat{\bar{x}}[k]\hat{\bar{x}}^{LS}[k]\} \ge E\{(\hat{\bar{x}}[k])^2\} - 2\sqrt{E\{(\hat{\bar{x}}[k])^2\}}E\{(\hat{\bar{x}}^{LS}[k])^2\}}$$
(A29)

and by summing over all possible values of the index k, we finally obtain

$$\sum_{k=1}^{n} E\{(\hat{\bar{x}}[k])^2\} - 2\sum_{k=1}^{n} E\{\hat{\bar{x}}[k]\hat{\bar{x}}^{LS}[k]\} \ge \sum_{k=1}^{n} E\{(\hat{\bar{x}}[k])^2\} - 2\sum_{k=1}^{n} \sqrt{E\{(\hat{\bar{x}}[k])^2\}}E\{(\hat{\bar{x}}^{LS}[k])^2\}}$$
(A30a)

or equivalently

$$\rho \ge \sum_{k=1}^{n} E\{(\hat{\bar{x}}[k])^2\} - 2\sum_{k=1}^{n} \sqrt{E\{(\hat{\bar{x}}[k])^2\}} E\{(\hat{\bar{x}}^{LS}[k])^2\}}.$$
(A30b)

The minimum value of  $\rho$  is attained when the equality sign in eq. (A30b) holds, and this occurs *if and only if* the cosine inequality in eq. (A27) becomes an equality for every value of the index *k*, that is,

$$E\{\hat{\bar{x}}[k]\hat{\bar{x}}^{LS}[k]\} = \sqrt{E\{(\hat{\bar{x}}[k])^2\}E\{(\hat{\bar{x}}^{LS}[k])^2\}}, k = 1, 2, ..., n$$
(A31)

which obviously holds if and only if the random variables  $\hat{\bar{x}}[k]$  and  $\hat{\bar{x}}^{LS}[k]$  are linearly related

$$\bar{\bar{x}}[k] = \lambda_k \bar{\bar{x}}^{LS}[k], k = 1, 2, ..., n,$$
(A32)

where  $\lambda_k$  denotes an arbitrary scalar factor that depends on the index *k*.

It is important to keep in mind that the elements  $\hat{\bar{x}}^{LS}[k]$  are uncorrelated with each other with variances equal to one ( $\mathbf{C}_{\hat{\mathbf{x}}^{LS}} = \mathbf{I}$ ). The elements  $\hat{\bar{x}}[k]$  are also uncorrelated with each other, with variances specified by a given diagonal matrix ( $\mathbf{C}_{\hat{\mathbf{x}}} = \mathbf{D}$ ) which depends on the *a priori* choice of the constrained CV matrix  $\mathbf{C}_{\hat{\mathbf{x}}}$  for the original estimated parameters; see eqs (A11), (A15) and (A16). In this way, the proportionality constants  $\lambda_k$  in (A32) can be uniquely specified. Using vector notation, we have that  $\hat{\mathbf{x}}$  and  $\hat{\mathbf{x}}^{LS}$  should be related through a *diagonal* matrix  $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_k, \dots, \lambda_n)$ 

$$\hat{\bar{\mathbf{x}}} = \Lambda \hat{\bar{\mathbf{x}}}^{LS} \tag{A33}$$

such that  $\Lambda = \mathbf{D}^{1/2}$ .

Hence, based on the vector transformation formula in eq. (A17), the regularization matrix  $\mathbf{\tilde{R}}$  becomes

$$\mathbf{\bar{\tilde{R}}} = \mathbf{D}^{1/2}.$$
(A34)

From eq. (A18), and taking also into account eqs (A7) and (A11), we obtain the regularization matrix **R** 

$$\begin{split} \bar{\mathbf{R}} &= \mathbf{V} \bar{\mathbf{R}} \mathbf{V}^{1} = \mathbf{V} \mathbf{D}^{1/2} \mathbf{V}^{\mathrm{T}} \\ &= (\mathbf{C}_{\hat{\mathbf{x}}})^{1/2} \\ &= [\mathbf{N}^{1/2} \mathbf{C}_{\hat{\mathbf{x}}} \mathbf{N}^{1/2}]^{1/2}. \end{split}$$
(A35)

Using the last result, the required regularization matrix  $\mathbf{R}$  is finally determined from eq. (10) as

$$\mathbf{R} = \mathbf{N}^{-1/2} \bar{\mathbf{R}} \mathbf{N}^{1/2}$$
  
=  $\mathbf{N}^{-1/2} [\mathbf{N}^{1/2} \mathbf{C}_{\bullet} \mathbf{N}^{1/2}]^{1/2} \mathbf{N}^{1/2}$  (A36)

In order to conclude the proof, it is required to show that the above result satisfies the CV-adaptive constraint given in eq. (A3). We have

$$\mathbf{R}\mathbf{N}^{-1}\mathbf{R}^{\mathrm{T}} = \mathbf{N}^{-1/2} \left[ \mathbf{N}^{1/2} \, \mathbf{C}_{\hat{\mathbf{x}}} \, \mathbf{N}^{1/2} \right]^{1/2} \, \mathbf{N}^{1/2} \, \mathbf{N}^{-1} \left( \mathbf{N}^{-1/2} \left[ \mathbf{N}^{1/2} \, \mathbf{C}_{\hat{\mathbf{x}}} \, \mathbf{N}^{1/2} \right]^{1/2} \, \mathbf{N}^{1/2} \right)^{1} \\ = \mathbf{N}^{-1/2} \left[ \mathbf{N}^{1/2} \, \mathbf{C}_{\hat{\mathbf{x}}} \, \mathbf{N}^{1/2} \right]^{1/2} \, \mathbf{N}^{-1} \, \mathbf{N}^{1/2} \left[ \mathbf{N}^{1/2} \, \mathbf{C}_{\hat{\mathbf{x}}} \, \mathbf{N}^{1/2} \right]^{1/2} \, \mathbf{N}^{-1/2} \\ = \mathbf{N}^{-1/2} \left[ \mathbf{N}^{1/2} \, \mathbf{C}_{\hat{\mathbf{x}}} \, \mathbf{N}^{1/2} \right]^{1/2} \left[ \mathbf{N}^{1/2} \, \mathbf{C}_{\hat{\mathbf{x}}} \, \mathbf{N}^{1/2} \right]^{1/2} \, \mathbf{N}^{-1/2} \\ = \mathbf{N}^{-1/2} \left[ \mathbf{N}^{1/2} \, \mathbf{C}_{\hat{\mathbf{x}}} \, \mathbf{N}^{1/2} \right]^{1/2} \, \mathbf{N}^{-1/2} \\ = \mathbf{C}_{\hat{\mathbf{x}}}$$
(A37)

which concludes our proof.

The auxiliary matrices **D** and  $\Lambda$  which have been used in the previous derivations are not related to the diagonal eigenvalue matrices **D** and  $\Lambda$  that were introduced in the main body of the paper. In particular, the symbol **D** in the main paper corresponds to a diagonal matrix containing the eigenvalues of the *a priori* constrained CV matrix of the estimated parameters **C**<sub>s</sub>; see (34). The same symbol, however, is also used in Appendix A to denote an auxiliary matrix that plays a supporting role in the presentation of the proof. Similarly, the symbol  $\Lambda$  in the main paper corresponds to a diagonal matrix containing the eigenvalues of the normal matrix **N** (see (12)), whereas in Appendix A the symbol  $\Lambda$  is employed to denote an auxiliary matrix that plays a supporting role in the presentation of the proof.

# APPENDIX B: AN ALTERNATIVE FORMULA FOR THE OPTIMAL REGULARIZATION MATRIX

In Appendix A it was proven that the optimal regularization matrix **R** in the CV-adaptive estimator  $\hat{\mathbf{x}} = \mathbf{R} \hat{\mathbf{x}}^{LS}$  is given by the equation

$$\mathbf{R} = \mathbf{N}^{-1/2} \left[ \mathbf{N}^{1/2} \, \mathbf{C}_{\hat{\mathbf{x}}} \, \mathbf{N}^{1/2} \right]^{1/2} \, \mathbf{N}^{1/2}. \tag{B1}$$

Here we shall derive an alternative equivalent expression for the above regularization matrix. From eq. (B1) we have

$$\mathbf{C}_{\hat{\mathbf{x}}}^{-1/2} \mathbf{R} \mathbf{C}_{\hat{\mathbf{x}}}^{1/2} = \mathbf{C}_{\hat{\mathbf{x}}}^{-1/2} \mathbf{N}^{-1/2} \left[ \mathbf{N}^{1/2} \, \mathbf{C}_{\hat{\mathbf{x}}} \, \mathbf{N}^{1/2} \right]^{1/2} \mathbf{N}^{1/2} \, \mathbf{C}_{\hat{\mathbf{x}}}^{1/2} \tag{B2}$$

and

$$\begin{bmatrix} \mathbf{C}_{\hat{\mathbf{x}}}^{-1/2} \mathbf{R} \mathbf{C}_{\hat{\mathbf{x}}}^{1/2} \end{bmatrix}^2 = \mathbf{C}_{\hat{\mathbf{x}}}^{-1/2} \mathbf{N}^{-1/2} \begin{bmatrix} \mathbf{N}^{1/2} \mathbf{C}_{\hat{\mathbf{x}}} \mathbf{N}^{1/2} \end{bmatrix}^{1/2} \mathbf{N}^{1/2} \mathbf{C}_{\hat{\mathbf{x}}}^{1/2} \\ \times \mathbf{C}_{\hat{\mathbf{x}}}^{-1/2} \mathbf{N}^{-1/2} \begin{bmatrix} \mathbf{N}^{1/2} \mathbf{C}_{\hat{\mathbf{x}}} \mathbf{N}^{1/2} \end{bmatrix}^{1/2} \mathbf{N}^{1/2} \mathbf{C}_{\hat{\mathbf{x}}}^{1/2} \\ = \mathbf{C}_{\hat{\mathbf{x}}}^{-1/2} \mathbf{N}^{-1/2} \begin{bmatrix} \mathbf{N}^{1/2} \mathbf{C}_{\hat{\mathbf{x}}} \mathbf{N}^{1/2} \end{bmatrix}^{1/2} \begin{bmatrix} \mathbf{N}^{1/2} \mathbf{C}_{\hat{\mathbf{x}}} \mathbf{N}^{1/2} \end{bmatrix}^{1/2} \mathbf{N}^{1/2} \mathbf{C}_{\hat{\mathbf{x}}}^{1/2} \\ = \mathbf{C}_{\hat{\mathbf{x}}}^{-1/2} \mathbf{N}^{-1/2} \mathbf{N}^{1/2} \mathbf{C}_{\hat{\mathbf{x}}} \mathbf{N}^{1/2} \mathbf{N}^{1/2} \mathbf{C}_{\hat{\mathbf{x}}}^{1/2} \\ = \mathbf{C}_{\hat{\mathbf{x}}}^{-1/2} \mathbf{C}_{\hat{\mathbf{x}}} \mathbf{N} \mathbf{C}_{\hat{\mathbf{x}}}^{1/2} \\ = \mathbf{C}_{\hat{\mathbf{x}}}^{-1/2} \mathbf{C}_{\hat{\mathbf{x}}} \mathbf{N} \mathbf{C}_{\hat{\mathbf{x}}}^{1/2} \\ = \mathbf{C}_{\hat{\mathbf{x}}}^{-1/2} \mathbf{N} \mathbf{C}_{\hat{\mathbf{x}}}^{1/2}. \end{aligned}$$
(B3)

© 2007 The Author, *GJI*, **171**, 509–522 Journal compilation © 2007 RAS 522 C. Kotsakis

Taking into account the result of the last formula, we can write that

$\mathbf{C}_{\mathbf{\hat{x}}}^{-1/2} \mathbf{R} \mathbf{C}_{\mathbf{\hat{x}}}^{1/2} = \left[\mathbf{C}_{\mathbf{\hat{x}}}^{1/2} \mathbf{N} \mathbf{C}_{\mathbf{\hat{x}}}^{1/2}\right]^{1/2}$	(B4)
and finally	
$\mathbf{R} = \mathbf{C}_{\hat{\mathbf{x}}}^{1/2} \left[ \mathbf{C}_{\hat{\mathbf{x}}}^{1/2}  \mathbf{N}  \mathbf{C}_{\hat{\mathbf{x}}}^{1/2} \right]^{1/2} \mathbf{C}_{\hat{\mathbf{x}}}^{-1/2}$	(B5)

which gives an alternative expression of the regularization matrix in eq. (B1).