Geophys. J. Int. (1994) **119**, 521–528

Biased estimation: a simple framework for inversion and uncertainty analysis with prior information

Max A. Meju
Environmental and Industrial Geophysics Research, Department of Geology, University of Leicester, Leicester LE1 7RH, *UK*

Accepted 1994 April 13. Received 1994 January 24

SUMMARY

The use of *a priori* information to resolve non-uniqueness in geophysical inversion is well known, but the kinds of constraining conditions required for the solution to an inverse problelm to be uniquely assured as well as the problem of extremal inversion with *a priori* information may still be explored further. An attempt has been made to address some aspects of these problems in inversion and uncertainty analysis within a unifying framework of biased estimation using a simple matrix algebra and taking advantage of the explicit distinction between the *a priori* information and the starting model in non-linear estimation. The adopted approach is flexible and allows the use of either reliable or diffuse *a priori* information making it a useful procedure for exploiting the peculiarities of different geophysical situations. It is shown that the more rigorous inversion algorithms can be derived easily from this framework as special cases and a digestible analysis is provided to increase our understanding of the undergirding principles of these classical algorithms.

Key words: *a priori* information, most-squares inversion, parameter estimation, uncertainty analysis.

1 INTRODUCTION

The inversion of practical geophysical measurements is beset by the problem of non-uniqueness due to a host of factors of which the typically inconsistent and bandlimited nature of the measurements, and the inherent non-linearity of the processes are the most problematical. In interpreting a scanty set of inexact data, conventional wisdom is to seek models that are in agreement with *a priori* data derived from say, previous geophysical studies, borehole or geological data. These extraneous information help single out a plausible solution from amongst all possible ones admitted by the inexact practical data. It is well known that the use of *a priori* data in linear inversion helps resolve the problem of non-uniqueness (e.g. Jackson 1979). Consequently, many ill-posed linear problems have been successfully solved by specifying *ab initio* some values of the desired parameters. However, while much is known about linear inverse problems, there are still a few outstanding issues in parameter estimation and error analysis. For example, it may be practically expedient to know whether the solution to an ill-posed linear inverse problem can be uniquely assured simply by specifying sufficient constraints only on the form of the solution, or by specifying values other than those of the actual parameters. In linear error analysis with prior information, the model uncertainties may be described in a concise form by the *a posteriori* covariance matrix (e.g. Jackson 1979; Jackson & Matsu'ura 1985; Duijndam 1988b). However, in the presence of observational uncertainties it may be preferable to calculate extreme parameters sets or bounding models incorporating the available *a priori* information since no single model can then satisfy all the observations. Jackson (1979) employed *a priori* information in extremal inversion. However, the extremal inversion algorithm given by Jackson (1979, eqs 53 and 54) is a weighted variant of the original most-squares formalism (Jackson 1976) and did not formally incorporate *a priori* parameter estimates. There is, therefore, a need to formalize this favoured approach to model appraisal. The above issues will form the starting point of this paper, and the discussion will then be extended to non-linear estimation problems.

In non-linear inversion, there is no unequivocal technique as yet for resolving the problem of non-uniqueness. Three common approaches to addressing this problem are the Backus–Gilbert (Backus & Gilbert 1968, 1970) and Monte Carlo methods and the direct use of *a priori* information. A composite scheme that shares the main features of all the above approaches may be developed within the adopted framework using the most-squares criterion (e.g. Meju & Hutton 1992). In parameter estimation with prior information, a common but naïve approach involves arbitrarily holding constant the values of some of the sought parameters within an iterative inversion scheme. The practical advantages of accruing from this approach are operational simplicity, efficiency and the fact that the interpreter retains the desired prior data in the optimal model by subjectively setting to nought the values of the parameter perturbations. Such a practice could be formalized by incorporating the *a priori* data directly in the problem formulation. However, most of the published formal treatment of *a priori* information in non-linear inversion adopts a probabilistic approach (e.g. Gol'tsman 1971, 1975; Tarantola & Valette 1982; Jackson & Matsu'ura 1985; Backus 1988; Duijndam 1988a,b) which, it may be argued, best characterizes the huge variability in practical geophysical measurements, but (with the notable exception of Jackson & Matsu'ura) the superior mathematical skills required to fully exploit all the benefits accruing from such schemes often makes the uninitiated data analyst (whose guiding philosophies are simplicity and flexibility) fall back on the informal approach. A flexible somewhat simpler approach involving an easily digestible matrix algebra and minimal statistical commitment is offered in this paper. An explicit distinction is made between starting models and *a priori* information to enable one to explore the peculiarities of different geophysical situations in accord with the aforementioned interpretational philosophies.

To achieve these objectives, different forms of *a priori* constraints or solution simplicity measures are considered and the observational errors are taken into account. Finally, the relationships between the ensuing algorithms and various classical inversion algorithms are shown. This strategy, it is hoped, will enable us to understand better the more rigorous landmark algorithms for non-linear inversion (e.g. Tarantola & Valetta 1982). Note, however, that the biased-estimation process as presented here is essentially an extension of the Marquadt (1970) method. The undergirding principles lean upon the excellent work on constrained linear inversion by Twomey (1977) and Jackson (1979).

2 BIASED LINEAR ESTIMATION REVISITED

2.1 Undergirding philosophy

The sought $p \times 1$ model parameters, $\mathbf{m}$ and our $n \times 1$ observational data $\mathbf{d}$ from which they are to be estimated are related in the explicit form $\mathbf{d} = \mathbf{Gm} + \mathbf{e}$, where $\mathbf{G}$ is the $n \times p$ design matrix of known coefficients and $\mathbf{e}$ is the $n \times 1$ vector of additive noise to the field measurements. Classical least-squares inversion attempts to determine the best estimates of our desired model parameters by minimizing the quadratic measure (see e.g. Twomey 1977)

$$\mathbf{e}^{\mathrm{T}}\mathbf{e} = (\mathbf{d} - \mathbf{Gm})^{\mathrm{T}}(\mathbf{d} - \mathbf{Gm}) \tag{1}$$

where the notation $\mathbf{a}^{\mathrm{T}}$ denotes the transpose of $\mathbf{a}$. The final result thus obtained is an unbiased estimate of the sought parameters. However, due to the nature of the problem, there will be other solutions that are also consistent with the data and hence the result is said to be non-unique. Fortunately, it is possible to narrow down the range of possible solutions to a problem by introducing some form of constraints at the outset. Such constraints may result from our previous knowledge of the form of the solution or some other related data and serve to single out a unique solution from among those which satisfy the data. In this situation, the solution process is said to be constrained and the results are said to be biased towards our *a priori* assumptions.

Our goal in biased estimation is to retain any desirable prior estimates or structural forms in the final solution to an inverse problem. It is well known that the general constrained linear problem may be formulated mathematically as the least-squares solution of a set of equations of the form (e.g. Jackson 1979)

$$\begin{bmatrix} \mathbf{G} \\ \cdots \\ \mathbf{D} \end{bmatrix} [\mathbf{m}] \approx \begin{bmatrix} \mathbf{d} \\ \cdots \\ \mathbf{h} \end{bmatrix} \tag{2}$$

where the diagonally dominant matrix $\mathbf{D}$ is a form of filter that emphasizes or suppresses some features of $\mathbf{m}$ depending on the values assigned to $\mathbf{h}$. Note that if $\mathbf{h}$ is the null vector, then some form of the filtered version of $\mathbf{m}$ is caused to vanish. The reverse is the case if $\mathbf{h}$ is a non-null vector. The bottom block in the above partitioned set of eqs (2) represents our constraints. These *a priori* constraints may also serve to stabilize the inversion process and the two forms to be considered here are

$$\mathbf{Dm} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_l \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_l \end{bmatrix} \tag{3a}$$

when it is desired to force our solution into conformity with some given values of $\mathbf{h}$ or

$$\mathbf{Dm} = \begin{bmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & & \\ & & & 1 & -1 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_l \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_l \end{bmatrix} \tag{3b}$$

if we are interested in retaining a particular form of the solution (see e.g. Twomey 1977; Menke 1984; Constable, Parker & Constable 1987; Meju & Hutton 1992) where $l \le p$ and implies that we can actively constrain some parameters leaving the rest to be determined by the data.

We also strive to maintain statistical stability in our solution process. Thus, owing to the nature of practical data and if we assume that the standard errors σ_i associated with our data are statistically independent, an $n \times n$ diagonal weighting matrix $\mathbf{W} = \sigma^{-1}\mathbf{I}$, may be defined and used for scaling our data equations so that undue importance is not given to poorly estimated data. This scaling operation can be interpreted as a process of standardization as the data equations are thus rendered dimensionless and uncorrelated (see e.g. Kaula 1966, Chapt. 5; Twomey 1977, Chapt. 7).

2.2 The estimation problem

Our goal here is to bias $\mathbf{m}$ towards $\mathbf{h}$. We simply state the problem as: 'Given a finite collection of inexact observational data, find the solution amongst all the equivalent ones (on account of data and model errors) that explains the observations and satisfies the given desirable estimates of the model parameters', or equivalently, minimize the Lagrangian function

$$\mathscr{L} = (\mathbf{Wd} - \mathbf{WGm})^{\mathrm{T}}(\mathbf{Wd} - \mathbf{WGm}) + \boldsymbol{\beta}(\mathbf{Dm} - \mathbf{h})^{\mathrm{T}}\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{Dm} - \mathbf{h}) \tag{4}$$

where $\boldsymbol{\beta}$ is a diagonal matrix of Langrange multipliers that enables us to pick out those parameters that should be forced into conformity with the prior estimates leaving the rest unconstrained; the elements of $\boldsymbol{\beta}$ corresponding to the prior data are assigned a constant positive value while the rest are set to nought and may not be considered in the interpretation process. The solution to this problem is

$$\mathbf{m} = [(\mathbf{WG})^{\mathrm{T}}\mathbf{WG} + \mathbf{D}^{\mathrm{T}}\mathbf{BD}]^{-1}[(\mathbf{WG})^{\mathrm{T}}\mathbf{Wd} + \mathbf{D}^{\mathrm{T}}\mathbf{Bh}] \tag{5}$$

where $\mathbf{B} = \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta}$. The somewhat identical role of the matrices $\mathbf{W}$ and $\boldsymbol{\beta}$ is obvious—both serve to weight the two kinds of data considered in the problem formulation and we shall examine the statistical implications later.

If the *a priori* information is diffuse (or unreliable), then it may be desirable to set the constraints equal to zero (i.e. $\mathbf{h} = [0, 0, \dots,]^{\mathrm{T}}$) and all the elements of $\boldsymbol{\beta}$ to some constant value $(0 < \beta_i < 1)$ so that all the parameters have equal importance. In this case, $\boldsymbol{\beta}$ may be conveniently replaced by a single undetermined multiplier β and the solution reduces to

$$\mathbf{m} = [(\mathbf{WG})^{\mathrm{T}}\mathbf{WG} + \beta^2\mathbf{D}^{\mathrm{T}}\mathbf{D}]^{-1}(\mathbf{WG})^{\mathrm{T}}\mathbf{Wd}. \tag{6}$$

Notice that if we define another matrix $\mathbf{E} = \mathbf{W}^{\mathrm{T}}\mathbf{W}$, then eq. (5) may be expressed in another enlightening form

$$\mathbf{m}_b = (\mathbf{G}^{\mathrm{T}}\mathbf{EG} + \mathbf{D}^{\mathrm{T}}\mathbf{BD})^{-1}[\mathbf{G}^{\mathrm{T}}\mathbf{Ed} + \mathbf{D}^{\mathrm{T}}\mathbf{Bh}] \tag{7}$$

and is thus equivalent to the so-called linear Bayesian estimator (see e.g. Jackson & Matsu'ura 1985; Duijndam 1988a).

The parameter resolution matrix (Jackson 1972) for this solution is simply

$$\begin{aligned}\mathbf{R} &= (\mathbf{G}^{\mathrm{T}}\mathbf{EG} + \mathbf{D}^{\mathrm{T}}\mathbf{BD})^{-1}(\mathbf{WG})^{\mathrm{T}} \cdot \mathbf{WG} \\ &\quad + (\mathbf{G}^{\mathrm{T}}\mathbf{EG} + \mathbf{D}^{\mathrm{T}}\mathbf{BD})^{-1}\boldsymbol{\beta}\mathbf{D}^{\mathrm{T}} \cdot \boldsymbol{\beta}^{\mathrm{T}}\mathbf{D} \\ &= (\mathbf{G}^{\mathrm{T}}\mathbf{EG} + \mathbf{D}^{\mathrm{T}}\mathbf{BD})^{-1}(\mathbf{G}^{\mathrm{T}}\mathbf{EG} + \mathbf{D}^{\mathrm{T}}\mathbf{BD}) \\ &= \mathbf{I}.\end{aligned} \tag{8}$$

Thus, as is well known, the constrained solution incorporating *a priori* parameter estimates has perfect resolution.

If the experimental data are uncorrelated and of equal variance σ^2, the covariance matrix of the estimated solution may be easily determined (by law of propagation of errors; see Meyer 1977) for eq. (7) as

$$\begin{aligned}\mathrm{Cov}\,(\mathbf{m}) = {} & (\mathbf{G}^{\mathrm{T}}\mathbf{EG} + \mathbf{D}^{\mathrm{T}}\mathbf{BD})^{-1}(\mathbf{WG})^{\mathrm{T}}\{\mathbf{E}[\mathrm{Cov}\,(\mathbf{d})]\} \\ & \times \mathbf{WG}(\mathbf{G}^{\mathrm{T}}\mathbf{EG} + \mathbf{D}^{\mathrm{T}}\mathbf{BD})^{-1} \\ & + (\mathbf{G}^{\mathrm{T}}\mathbf{EG} + \mathbf{D}^{\mathrm{T}}\mathbf{BD})^{-1}(\boldsymbol{\beta}\mathbf{D})^{\mathrm{T}}\{\mathbf{B}[\mathrm{Cov}\,(\mathbf{h})]\} \\ & \times \boldsymbol{\beta}\mathbf{D}(\mathbf{G}^{\mathrm{T}}\mathbf{EG} + \mathbf{D}^{\mathrm{T}}\mathbf{BD})^{-1}\end{aligned}$$

in this standardized framework, or simply

$$\begin{aligned}\mathrm{Cov}\,(\mathbf{m}) &= (\mathbf{G}^{\mathrm{T}}\mathbf{EG} + \mathbf{D}^{\mathrm{T}}\mathbf{BD})^{-1}\{\mathbf{G}^{\mathrm{T}}\mathbf{EG} + \mathbf{D}^{\mathrm{T}}\mathbf{BD}\} \\ &\quad \times (\mathbf{G}^{\mathrm{T}}\mathbf{EG} + \mathbf{D}^{\mathrm{T}}\mathbf{BD})^{-1} \\ &= (\mathbf{G}^{\mathrm{T}}\mathbf{EG} + \mathbf{D}^{\mathrm{T}}\mathbf{BD})^{-1}\end{aligned} \tag{9a}$$

if we allow for variations in $\mathbf{h}$. The above expression coincides with the *a posteriori* covariance given by other workers (e.g. Jackson 1979; Jackson & Matsu'ura 1985; Duijndam 1998b). However, it may be contended that the variances and covariances of parameters constrained to equal fixed values are zero, in which case

$$\mathrm{Cov}\,(\mathbf{m}) = (\mathbf{G}^{\mathrm{T}}\mathbf{EG} + \mathbf{D}^{\mathrm{T}}\mathbf{BD})^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{EG}(\mathbf{G}^{\mathrm{T}}\mathbf{EG} + \mathbf{D}^{\mathrm{T}}\mathbf{BD})^{-1} \tag{9b}$$

suggesting that the covariances of the free parameters are also reduced by the use of such active constraints in linear estimation.

2.3 Extremal inversion with *a priori* information

The most-squares method of Jackson (1976) is suited to the problem of determining extreme parameter sets from experimental data bedeviled by observational uncertainties and will be reformulated here within the framework of biased estimation. The constrained problem is defined simply as: given $\mathbf{m}$, an optimal least-squares solution to an inverse problem incorporating *a priori* information with residuals q_{LS}, find (on account of the observational and model uncertainties) other solutions retaining the *a priori* data that fit the experimental data to a specified threshold residual q_T; or equivalently, minimize

$$\begin{aligned}\mathscr{L} = {} & \mathbf{m}^{\mathrm{T}}\mathbf{b} + \frac{1}{2\mu}\{(\mathbf{Wd} - \mathbf{WGm})^{\mathrm{T}}(\mathbf{Wd} - \mathbf{WGm}) \\ & + (\mathbf{Dm} - \mathbf{h})^{\mathrm{T}}\mathbf{B}(\mathbf{Dm} - \mathbf{h}) - q_{\mathrm{T}}\}\end{aligned} \tag{10}$$

where the premultiplier $1/2\mu$ is undetermined and $\mathbf{b}$ is a projection vector (see Jackson 1979). The biased most-squares solution to the above problem is

$$\mathbf{m}_b = [\mathbf{G}^{\mathrm{T}}\mathbf{EG} + \mathbf{D}^{\mathrm{T}}\mathbf{BD}]^{-1}[\mathbf{G}^{\mathrm{T}}\mathbf{Ed} + \mathbf{D}^{\mathrm{T}}\mathbf{Bh} - \mu\mathbf{b}] \tag{11}$$

and must satisfy the condition

$$\begin{aligned}q_{\mathrm{T}} &= |\mathbf{Wd} - \mathbf{WGm}|^2 + \mathbf{B}\,|\mathbf{Dm} - \mathbf{h}|^2 \\ &= \mathbf{d}^{\mathrm{T}}\mathbf{Ed} + \mathbf{h}^{\mathrm{T}}\mathbf{Bh} - (\mathbf{d}^{\mathrm{T}}\mathbf{EG} + \mathbf{h}^{\mathrm{T}}\mathbf{BD})\mathbf{m} \\ &\quad - \mathbf{m}^{\mathrm{T}}(\mathbf{G}^{\mathrm{T}}\mathbf{Ed} + \mathbf{D}^{\mathrm{T}}\mathbf{Bh}) \\ &\quad + \mathbf{m}^{\mathrm{T}}(\mathbf{G}^{\mathrm{T}}\mathbf{EG} + \mathbf{D}^{\mathrm{T}}\mathbf{BD})\mathbf{m}.\end{aligned} \tag{12}$$

Substituting $\mathbf{m}_b$ for $\mathbf{m}$ in eq. (12), we have that

$$\begin{aligned}q_{\mathrm{T}} &= \mathbf{d}^{\mathrm{T}}\mathbf{Ed} + \mathbf{h}^{\mathrm{T}}\mathbf{Bh} - (\mathbf{d}^{\mathrm{T}}\mathbf{EG} + \mathbf{h}^{\mathrm{T}}\mathbf{BD}) \\ &\quad \times (\mathbf{G}^{\mathrm{T}}\mathbf{EG} + \mathbf{D}^{\mathrm{T}}\mathbf{BD})^{-1}(\mathbf{G}^{\mathrm{T}}\mathbf{Ed} + \mathbf{D}^{\mathrm{T}}\mathbf{Bh}) \\ &\quad + \mu\mathbf{b}^{\mathrm{T}}(\mathbf{G}^{\mathrm{T}}\mathbf{EG} + \mathbf{D}^{\mathrm{T}}\mathbf{BD})^{-1}\mu\mathbf{b}\end{aligned}$$

giving

$$\mu = \pm\left(\frac{q_{\mathrm{T}} - q_{LS}}{\mathbf{b}^{\mathrm{T}}[\mathbf{G}^{\mathrm{T}}\mathbf{EG} + \mathbf{D}^{\mathrm{T}}\mathbf{BD}]^{-1}\mathbf{b}}\right)^{1/2} \tag{13}$$

where

$$\begin{aligned}q_{LS} &= \mathbf{d}^{\mathrm{T}}\mathbf{Ed} + \mathbf{h}^{\mathrm{T}}\mathbf{Bh} - (\mathbf{d}^{\mathrm{T}}\mathbf{EG} + \mathbf{h}^{\mathrm{T}}\mathbf{BD}) \\ &\quad \times (\mathbf{G}^{\mathrm{T}}\mathbf{EG} + \mathbf{D}^{\mathrm{T}}\mathbf{BD})^{-1}(\mathbf{G}^{\mathrm{T}}\mathbf{Ed} + \mathbf{D}^{\mathrm{T}}\mathbf{Bh}).\end{aligned}$$

If the constraints are set equal to zero for the smoothest solution (as gauged by the measure $\mathbf{m}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{BDm}$), then

$$\mathbf{m}_b = [\mathbf{G}^{\mathrm{T}}\mathbf{EG} + \beta^2\mathbf{D}^{\mathrm{T}}\mathbf{D}]^{-1}[\mathbf{G}^{\mathrm{T}}\mathbf{Ed} - \mu\mathbf{b}] \tag{14}$$

and

$$\mu = \pm\left(\frac{q_T - q_{LS}}{\mathbf{b}^T[\mathbf{G}^T\mathbf{E}\mathbf{G} + \beta^2\mathbf{D}^T\mathbf{D}]^{-1}\mathbf{b}}\right)^{1/2} \tag{15}$$

where

$$q_{LS} = \mathbf{d}^T\mathbf{E}\mathbf{d} - \mathbf{d}^T\mathbf{E}\mathbf{G}(\mathbf{G}^T\mathbf{E}\mathbf{G} + \beta^2\mathbf{D}^T\mathbf{D})^{-1}\mathbf{G}^T\mathbf{E}\mathbf{d}.$$

Note that if $\mathbf{D} = \mathbf{I}$, then the above constraining operations are equivalent to adding a positive constant bias to the main diagonal of the matrix $\mathbf{G}^T\mathbf{E}\mathbf{G}$, a well-known stabilizing procedure. The expected value of q_T is $n - l$, where there are l active constraints in the problem. The appropriate bounding models may be determined using the above most-squares solution. It may be noted that Jackson (1979) in his seminal paper used *a priori* information in extremal inversion but gave the unconstrained most-squares formula (Jackson 1979, eqs 53 and 54)

$$\mathbf{m} = [\mathbf{G}^T\mathbf{E}\mathbf{G}]^{-1}[\mathbf{G}^T\mathbf{E}\mathbf{d} - \mu\mathbf{b}] \quad \text{with} \quad \mu = \pm\left(\frac{q_T - q_{LS}}{\mathbf{b}^T[\mathbf{G}^T\mathbf{E}\mathbf{G}]^{-1}\mathbf{b}}\right)^{1/2}$$

which is different from that given above. The present development may thus be interpreted as an extension of his method.

It is instructive to proceed in the same fashion as above for a different class of inverse problems when attempting to develop a naïve unifying framework for parameter estimation and uncertainty analysis as in the following formulations.

3 BIASED NON-LINEAR ESTIMATION

Many problems in geophysics are non-linear in nature and the observed data and the sought model parameters are related in the form $\mathbf{d} = \mathbf{f}(\mathbf{m}) + \mathbf{e}$, where the functional $\mathbf{f}$ allows us to predict the observable data for a given numerical representation of a geophysical system—the forward theory. To bias out parameter estimation towards $\mathbf{h}$, an iterative process must be used. Now, since retaining some given values of the sought parameters in the solution assumes that they are independent of the data, an explicit distinction is made between *a priori* information, $\mathbf{h}$ and initial (or starting) models, $\mathbf{m}^0$ in this approach. Note, however, that the components of the starting model may also include the desirable *a priori* estimates towards which we wish to bias the final solution, but this is not obligatory. As before, we define the problem as minimize

$$\begin{aligned}\mathscr{L} &= (\mathbf{W}\mathbf{d} - \mathbf{W}\mathbf{f}(\mathbf{m}))^T(\mathbf{W}\mathbf{d} - \mathbf{W}\mathbf{f}(\mathbf{m})) \\ &\quad + (\boldsymbol{\beta}[\mathbf{D}\mathbf{m} - \mathbf{h}])^T(\boldsymbol{\beta}[\mathbf{D}\mathbf{m} - \mathbf{h}])\end{aligned} \tag{16}$$

where the symbols have their usual meanings.

3.1 Linearized estimation

If $\mathbf{f}(\mathbf{m})$ is continuous and differentiable, expanding it about an initial model $\mathbf{m}^0$ using Taylor's theorem gives the linearized approximation of eq. (16) as

$$\begin{aligned}\mathscr{L} &= (\mathbf{W}\mathbf{y} - \mathbf{W}\mathbf{A}\mathbf{x})^T(\mathbf{W}\mathbf{y} - \mathbf{W}\mathbf{A}\mathbf{x}) \\ &\quad + \{[\mathbf{D}(\mathbf{m}^0 + \mathbf{x}) - \mathbf{h}]^T\boldsymbol{\beta}^T\boldsymbol{\beta}[\mathbf{D}(\mathbf{m}^0 + \mathbf{x}) - \mathbf{h}\}\end{aligned} \tag{17}$$

where $\mathbf{y}_i = \mathbf{d}_i - \mathbf{f}_i(\mathbf{m}^0)$, $\mathbf{A}$ is the $n \times p$ matrix with elements $\mathbf{A}_{ij} = \partial\mathbf{f}_i(\boldsymbol{m}^0)/\partial\boldsymbol{m}_j$, and we have used the approximation $\mathbf{x} = \mathbf{m} - \mathbf{m}^0$.

Minimization is effected by setting to zero the derivatives of $\mathscr{L}$ with respect to $\mathbf{x}$, i.e.

$$\begin{aligned}\frac{\partial\mathscr{L}}{\partial\mathbf{x}} &= -\mathbf{y}^T\mathbf{E}\mathbf{A} - \mathbf{A}^T\mathbf{E}\mathbf{y} + \mathbf{A}^T\mathbf{E}\mathbf{A}\mathbf{x} + \mathbf{x}^T\mathbf{A}^T\mathbf{E}\mathbf{A} \\ &\quad + \mathbf{m}^{0T}\mathbf{D}^T\mathbf{B}\mathbf{D} + \mathbf{D}^T\mathbf{B}\mathbf{D}\mathbf{m}^0 + \mathbf{D}^T\mathbf{B}\mathbf{D}\mathbf{x} \\ &\quad + \mathbf{x}^T\mathbf{D}^T\mathbf{B}\mathbf{D} - \mathbf{D}^T\mathbf{B}\mathbf{h} - \mathbf{h}^T\mathbf{B}\mathbf{D} = 0\end{aligned} \tag{18}$$

yielding the normal equations

$$\mathbf{A}^T\mathbf{E}\mathbf{A}\mathbf{x} + \mathbf{D}^T\mathbf{B}\mathbf{D}\mathbf{x} = \mathbf{A}^T\mathbf{E}\mathbf{y} + \mathbf{D}^T\mathbf{B}\mathbf{h} - \mathbf{D}^T\mathbf{B}\mathbf{D}\mathbf{m}^0 \tag{19}$$

whose solution for the parameter corrections to be applied to $\mathbf{m}^0$ is

$$\mathbf{x} = [\mathbf{A}^T\mathbf{E}\mathbf{A} + \mathbf{D}^T\mathbf{B}\mathbf{D}]^{-1}[\mathbf{A}^T\mathbf{E}\mathbf{y} + \mathbf{D}^T\mathbf{B}\{\mathbf{h} - \mathbf{D}\mathbf{m}^0\}]. \tag{20}$$

Non-linearity is dealt with using an iterative formula of the form

$$\mathbf{m}^{k+1} = \mathbf{m}^k + [\mathbf{A}^T\mathbf{E}\mathbf{A} + \mathbf{D}^T\mathbf{B}\mathbf{D}]^{-1}[\mathbf{A}^T\mathbf{E}\mathbf{y} + \mathbf{D}^T\mathbf{B}\{\mathbf{h} - \mathbf{D}\mathbf{m}^k\}] \tag{21}$$

where the Jacobian matrix $\mathbf{A}$ and the data misfit vector $\mathbf{y}$ are evaluated at $\mathbf{m}^k$ and the iteration is begun at $k = 0$. The term $\mathbf{D}^T\mathbf{B}\{\mathbf{h} - \mathbf{D}\mathbf{m}^k\}$ on the right-hand side of eq. (21) serves to force the solution into conformity with the specified *a priori* parameter estimates while the $\boldsymbol{\beta}\mathbf{D}^T\mathbf{D}$ term in the inverted system serves to regularize the solution process (Tikhonov 1963; Twomey 1963; Tikhonov & Arsenin 1977). This is in effect a maximum-bias algorithm as it can be used to generate acceptable solutions which retain the values of the parameters specified *ab initio* in accord with the interpreter's wisdom. If the *a priori* information is diffuse (or inaccurate), then as in the linear case the problem is thus redefined as the search for the smoothest model (as gauged by the smoothness measure $\beta^2\mathbf{m}^T\mathbf{D}^T\mathbf{D}\mathbf{m}$) that will explain the given experimental data or equivalently, we *minimize* the function (*cf.* Constable *et al.* 1987; Meju 1993)

$$\mathscr{L} = (\mathbf{W}\mathbf{d} - \mathbf{W}\mathbf{f}(\mathbf{m}))^T(\mathbf{W}\mathbf{d} - \mathbf{W}\mathbf{f}(\mathbf{m})) + \beta^2\mathbf{m}^T\mathbf{D}^T\mathbf{D}\mathbf{m}. \tag{22}$$

The solution for the parameter corrections is

$$\mathbf{x}_s = [\mathbf{A}^T\mathbf{E}\mathbf{A} + \beta^2\mathbf{D}^T\mathbf{D}]^{-1}[\mathbf{A}^T\mathbf{E}\mathbf{y} - \beta^2\mathbf{D}^T\mathbf{D}\mathbf{m}^0] \tag{23}$$

and the iteration formula is identically

$$\mathbf{m}^{k+1} = \mathbf{m}^k + [\mathbf{A}^T\mathbf{E}\mathbf{A} + \beta^2\mathbf{D}^T\mathbf{D}]^{-1}[\mathbf{A}^T\mathbf{E}\mathbf{y} - \beta^2\mathbf{D}^T\mathbf{D}\mathbf{m}^k]. \tag{24}$$

In this case, the term $\beta^2\mathbf{D}^T\mathbf{D}$ serves to regularize the solution process while the term $-\beta^2\mathbf{D}^T\mathbf{D}\mathbf{m}^k$ helps control the departure of the solution from the stipulated form (i.e. $\mathbf{h} = [0, \ldots, 0]^T$); and we may refer to the procedure as inversion with smoothness measures or the minimum-bias algorithm.

Note that all the above algorithms simplify greatly if $\mathbf{D} = \mathbf{I}$ which is the case when we want to retain the specified prior estimates in the final solution. The algorithms also hold good if we are interested in retaining known or desirable forms of the sought model as in some practical situations where there are gradational changes in physical properties in the subsurface say, and we wish to obtain the smoothest solution with minimized differences between physically adjacent parameters (e.g. Constable *et al.* 1987; Meju & Hutton 1992; Meju 1993); generally speaking, in such cases, we may not have any quantifiable estimates of $\mathbf{h}$ and conventional wisdom dictates that we set the elements of $\mathbf{h}$ equal to nought.

3.2 Consistency analysis of linearized solutions

It is instructive to ascertain that the above non-linear solution process will yield the linear solution as a special case if the problem is sufficiently linear, i.e. that the formulations are consistent with the well-understood linear analogues (see also, Tarantola & Valette 1982, p. 224). Recall that in linear estimation the sought model parameters are directly retrieved from the data. We may also elect to solve explicitly for the sought parameters instead of the usual parameter corrections in non-linear estimation (see Constable *et al.* 1987; Meju 1993) and the solution thus obtained can be forced to give the linear solution if the problem tackled is only weakly non-linear. This approach is also useful when calculating parameter covariances in non-linear estimation as shown later. To achieve this goal, we simply replace $\mathbf{x}$ with the quantity $(\mathbf{m}-\mathbf{m}^0)$ in the appropriate normal equations and solve directly for $\mathbf{m}$.

The normal equations for the maximum-bias estimator (eq. 19) may be written as

$$\mathbf{A}^T\mathbf{E}\mathbf{A}\mathbf{m} + \mathbf{D}^T\mathbf{B}\mathbf{D}\mathbf{m} = \mathbf{A}^T\mathbf{E}\mathbf{y} + \mathbf{D}^T\mathbf{B}\mathbf{h} - \mathbf{D}^T\mathbf{B}\mathbf{D}\mathbf{m}^0 + \{\mathbf{A}^T\mathbf{E}\mathbf{A}\mathbf{m}^0 + \mathbf{D}^T\mathbf{B}\mathbf{D}\mathbf{m}^0\}$$

so that

$$\mathbf{m} = (\mathbf{A}^T\mathbf{E}\mathbf{A} + \mathbf{D}^T\mathbf{B}\mathbf{D})^{-1}[(\mathbf{W}\mathbf{A})^T\{\hat{\mathbf{d}}\} + \mathbf{D}^T\mathbf{B}\mathbf{h}] \tag{25a}$$

where $\{\hat{\mathbf{d}}\} = \{\mathbf{W}\mathbf{y} + \mathbf{W}\mathbf{A}\mathbf{m}^0\}$ is a kind of data and the formula yields the biased linear solution (eq. 5) if the problem is linear. The minimum-bias (or smooth) solution given by eq. (23) is identically equal to

$$\mathbf{m} = (\mathbf{A}^T\mathbf{E}\mathbf{A} + \beta^2\mathbf{D}^T\mathbf{D})^{-1}(\mathbf{W}\mathbf{A})^T\{\hat{\mathbf{d}}\} \tag{26}$$

which is also consistent with the linear analogue represented by eq. (6). The simplicity of the biased estimation approach is clearly seen when we restate eq. (25a) in the form (*cf.* eq. 2)

$$[(\mathbf{W}\mathbf{A})^T \vdots (\boldsymbol{\beta}\mathbf{D})^T]\begin{Bmatrix}\mathbf{W}\mathbf{A}\\ \cdots \\ \boldsymbol{\beta}\mathbf{D}\end{Bmatrix}\mathbf{m} = [(\mathbf{W}\mathbf{A})^T \vdots (\boldsymbol{\beta}\mathbf{D})^T]\begin{Bmatrix}\hat{\mathbf{d}}\\ \cdots \\ \mathbf{h}\end{Bmatrix}. \tag{25b}$$

3.3 Relationships with standard non-linear inversion algorithms

The stability characteristics and the effectiveness of the biased estimation methods depend largely on $\boldsymbol{\beta}$ and $\mathbf{D}$. A comparison between the algorithms derived above and a few classical inversion schemes would perhaps throw more light on the principle and techniques of biased estimation. For the case $\mathbf{D} = \mathbf{I}$, the algorithm given in eq. (21) simplifies to

$$\mathbf{m}^{k+1} = \mathbf{m}^k + \{[\mathbf{A}^T\mathbf{E}\mathbf{A} + \mathbf{B}]^{-1}[\mathbf{A}^T\mathbf{E}\mathbf{y} + \mathbf{B}(\mathbf{h} - \mathbf{m}^k)]\} \tag{27}$$

which is equivalent to the Bayesian estimation scheme of Jackson & Matsu'ura (1985, eq. 65–67) and comparable to Tarantola & Valette's (1982, eq. 49) non-linear algorithm if $\mathbf{B}$ is statistically interpreted as the inverse *a priori* parameter covariance matrix. Thus, using a simple algebra and a practical interpretive strategy, we have in effect developed a somewhat similar scheme to the more rigorous schemes based on a probabilistic treatment of non-linear inversion with prior data. However, it is re-iterated that the inversion philosophy and the usage of *a priori* information in Tarantola & Valette's landmark method differ significantly from ours. We are primarily interested in forcing the final solution into the close conformity with those parameter estimates known *a priori* and thus the last term on the right-hand side of eq. (27) is non-zero since $\mathbf{h}$ may not be the same as $\mathbf{m}^0$ and only a few parameters estimates may be known beforehand (or we would not be attempting to estimate them). In Tarantola & Valette's algorithm, $\mathbf{h}$ is simply the actual initial model $\mathbf{m}^0$ and is effectively non-informative and it would appear that the main thrust is in constructing solutions that are independent of the prior data. In the present biased estimation scheme, $\mathbf{h}$ is treated as being separate from $\mathbf{m}^0$, more or less as in Jackson & Matsu'ura (1985) but with increased attention to those parameters that have been determined *a priori* and deemed worthy of retention in the final solution. Put simply, this approach stresses the independence of $\mathbf{h}$ on the observational data. A notable structural simplicity of this approach is that one can easily constrain some of the parameters leaving the rest free to be determined by the data. Also, the matrix $\mathbf{B}$ is determined *post facto* (as that value that yields a solution satisfying the prior and observed data).

Note that in the algorithm of Jackson & Matsu'ura, the quantity in braces in eq. (27) is multiplied by a suitably chosen factor of convenience b $(0<b<1)$; this is a useful practical constraint and is tantamount to decreasing the solution's length without changing its direction.

Notice that eq. (23) differs from the classical damped least squares or ridge regression (Marquardt 1963, 1970; Hoerl & Kennard 1970) update formula (see e.g. Meju 1992)

$$\mathbf{x} = [\mathbf{A}^T\mathbf{E}\mathbf{A} + \beta\mathbf{I}]^{-1}\mathbf{A}^T\mathbf{E}\mathbf{y} \tag{28}$$

by the $-\beta^2\mathbf{m}^0$ term where $\mathbf{D} = \mathbf{I}$. Now, the above formula could be obtained by minimizing

$$\mathscr{L} = (\mathbf{W}\mathbf{d} - \mathbf{W}\mathbf{f}(\mathbf{m}))^T(\mathbf{W}\mathbf{d} - \mathbf{W}\mathbf{f}(\mathbf{m})) + \beta(\mathbf{m} - \mathbf{m}^0)^T(\mathbf{m} - \mathbf{m}^0) \tag{29}$$

where we are mainly searching for that bounded perturbation to our initial model that is optimal for fitting our data (*cf.* Marquardt 1963). To fully understand the Marquardt solution and to highlight any differences from that proposed here, let us analyse it for consistency. Using the normal equations for (28), we have that

$$(\mathbf{A}^T\mathbf{E}\mathbf{A} + \beta\mathbf{I})\mathbf{m} = (\mathbf{W}\mathbf{A})^T(\mathbf{W}\mathbf{y} + \mathbf{W}\mathbf{A}\mathbf{m}^0) + \beta\mathbf{I}\mathbf{m}^0$$

so that

$$\mathbf{m} = (\mathbf{A}^T\mathbf{E}\mathbf{A} + \beta\mathbf{I})^{-1}\{(\mathbf{W}\mathbf{A})^T\hat{\mathbf{d}} + \beta\mathbf{m}^0\} \tag{30}$$

where the quantity $\hat{\mathbf{d}}$ is as previously defined. It is obvious that the above relation is structurally similar to eq. (25a) if the term $\beta\mathbf{m}^0$ is equated to $\boldsymbol{\beta}^T\mathbf{h}$, i.e. some 'assumed' *a priori* model. Thus the main difference between the traditional

Marquardt approach and that advocated here is that the former uses a new *a priori* model (the previous iterate) at each iteration. This strategy is not in accord with our philosophy of biased estimation and, as also mooted by Jackson (1979), treating the previous iterate as the *a priori* model is not consistent with our assumption that **h** is independent of the observational data under consideration, and may not be a favoured approach to addressing the problem of non-uniqueness in geophysical inversion.

It is also obvious that the use of the constraints defined in eq. (3b) in eq. (26) leads to the hugely popular smooth model ('occam') algorithm of Constable *et al.* (1987) which may therefore be regarded as a special case of biased estimation. As elaborated upon in Meju (1993), the optimally smooth models derived from smooth (or supposedly non-informative) initial models using this approach are dependent on the *a priori* assumptions in the solution process.

These caveats aside, the biased estimation methods presented in this paper have been shown to be comparable to many of the conventional non-linear inversion methods. However, unlike most other algorithms we can explore different geophysical situations and this approach is therefore general. Let us now examine the topical issue of error analysis in non-linear inversion.

4 UNCERTAINTY ANALYSIS IN NON-LINEAR ESTIMATION

Error analysis in non-linear inversion is an unresolved fundamental problem. In this section, some suggestions are offered for estimating the approximate reliability of the non-linear solutions and are based on the tools developed in Section 2 for linear model appraisal.

4.1 Parameter resolution matrix

The resolution matrix of a linear system (e.g. Jackson 1972) can be easily calculated and this was one motivation for re-casting our non-linear solution in equivalent linear forms for the consistency analyses presented in Section 3.2. Applying the straightforward rule $\mathbf{R} = \mathbf{HA}$ (where $\mathbf{H}$ is the generalized inverse used) to eq. (25a), we have that

$$\begin{aligned}\mathbf{R} &= \{(\mathbf{A}^T\mathbf{EA} + \mathbf{D}^T\mathbf{BD})^{-1}\mathbf{A}^T\mathbf{E}\cdot\mathbf{A}\} \\ &\quad + \{(\mathbf{A}^T\mathbf{EA} + \mathbf{D}^T\mathbf{BD})^{-1}\mathbf{D}^T\cdot\mathbf{BD}\} \\ &= (\mathbf{A}^T\mathbf{EA} + \mathbf{D}^T\mathbf{BD})^{-1}(\mathbf{A}^T\mathbf{EA} + \mathbf{D}^T\mathbf{BD}) = \mathbf{I} \qquad (31)\end{aligned}$$

showing that we may have perfect resolution by incorporating *a priori* data in our solution process (see also, Jackson & Matsu'ura 1985). In general, $\mathbf{R}$ is evaluated at an acceptable model and may be interpreted as a gauge of the balance between the information provided by the data and those assumed *a priori*. If $\mathbf{R} = \mathbf{I}$, then each model parameter may be well determined. As in the linear case, the deviation of the rows of $\mathbf{R}$ from those of the identity matrix measures the lack of resolution for the corresponding model parameters.

4.2 Parameter covariance matrix

Model errors in non-linear inversion are commonly derived from the parameter covariance matrix. Following the established rule for linear estimation, the covariance of the maximum-bias solution (eq. 25a) may be defined as

$$\begin{aligned}\mathrm{Cov}(\mathbf{m}) &= \{(\mathbf{A}^T\mathbf{EA} + \mathbf{D}^T\mathbf{BD})^{-1}(\mathbf{WA})^T\}\mathbf{E}[\mathrm{Cov}(\hat{\mathbf{d}})] \\ &\quad \times \{(\mathbf{A}^T\mathbf{EA} + \mathbf{D}^T\mathbf{BD})^{-1}(\mathbf{WA})^T\}^T \\ &\quad + \{(\mathbf{A}^T\mathbf{EA} + \mathbf{D}^T\mathbf{BD})^{-1}(\boldsymbol{\beta}\mathbf{D})^T\}\mathbf{B}[\mathrm{Cov}(\mathbf{h})] \\ &\quad \times \{(\mathbf{A}^T\mathbf{EA} + \mathbf{D}^T\mathbf{BD})^{-1}(\boldsymbol{\beta}\mathbf{D})^T\}^T. \qquad (32)\end{aligned}$$

Here, we may take advantage of the explicit distinction between **h** and $\mathbf{m}^0$. If the parameters are actively constrained to equal fixed values (i.e. **h** is assumed to be independent of the fitted data), then eq. (32) reduces to

$$\mathrm{Cov}(\mathbf{m}) = (\mathbf{A}^T\mathbf{EA} + \mathbf{D}^T\mathbf{BD})^{-1}\mathbf{A}^T\mathbf{EA}(\mathbf{A}^T\mathbf{EA} + \mathbf{D}^T\mathbf{BD})^{-1}. \qquad (33)$$

However, the matrix $\boldsymbol{\beta}$ is largely undetermined and controls the degree to which the optimal solution retains the specified **h**. It follows that the resulting point estimates may be different for each set of values used as $\boldsymbol{\beta}$ (a fact exploited in Meju 1993 for defining the effective depth to which any given data set constrains a model). It is possible to describe these variations statistically and determine the corresponding covariance matrix. However, noting that both the $\mathbf{W}$ and $\boldsymbol{\beta}$ matrices serve the same function for the two kinds of data considered in the problem formulation and that Cov (**d**) is given by $\sigma^2\mathbf{I} = (\mathbf{W}^T\mathbf{W})^{-1}$ or simply $\mathbf{E}^{-1}$, we make the intuitive deduction that the covariance of the variations in **h** may be given by $\mathbf{B}^{-1} = [\boldsymbol{\beta}^T\boldsymbol{\beta}]^{-1}$. The effect of these variations on Cov (**m**) may then be roughly approximated by

$$\mathrm{Cov}(\mathbf{m}_h^{est}) = (\mathbf{A}^T\mathbf{EA} + \mathbf{D}^T\mathbf{BD})^{-1}\mathbf{D}^T\mathbf{BD}(\mathbf{A}^T\mathbf{EA} + \mathbf{D}^T\mathbf{BD})^{-1} \qquad (34)$$

so that the full covariance relation given by eq. (32) can be evaluated as the sum of the two sets of covariances defined in eqs (33) and (34), that is

$$\begin{aligned}\mathrm{Cov}(\mathbf{m}) &= (\mathbf{A}^T\mathbf{EA} + \mathbf{D}^T\mathbf{BD})^{-1}[\mathbf{A}^T\mathbf{EA} + \mathbf{D}^T\mathbf{BD}] \\ &\quad \times (\mathbf{A}^T\mathbf{EA} + \mathbf{D}^T\mathbf{BD})^{-1} \\ &= (\mathbf{A}^T\mathbf{EA} + \mathbf{D}^T\mathbf{BD})^{-1} \qquad (35)\end{aligned}$$

which is similar to that derived from different viewpoints by other workers (Jackson & Matsu'ura 1985; Duijndam 1988b; Backus 1988).

For the minimum-bias solution given by eq. (26), we find that

$$\mathrm{Cov}(\mathbf{m}) = (\mathbf{A}^T\mathbf{EA} + \beta^2\mathbf{I})^{-1}\mathbf{A}^T\mathbf{EA}(\mathbf{A}^T\mathbf{EA} + \beta^2\mathbf{I})^{-1}. \qquad (36)$$

4.3 Extreme parameter sets

Meju & Hutton (1992) recently showed how Jackson's (1976) most-squares method can be applied in the determination of bounding models in non-linear inversion. The method will be extended here to allow for the retention of *a priori* parameter estimates in the extremal inversion process. Essentially, having obtained an optimal least-squares solution to an inverse problem **m**, we extremize the objective function $\mathbf{m}^T\mathbf{b}$ subject to the constraint that the residuals $q = |\mathbf{Wd} - \mathbf{Wf}(\mathbf{m})|^2 + |\boldsymbol{\beta}^T(\mathbf{Dm} - \mathbf{h})|^2$ are not greater than some threshold value q_T, where $\mathbf{b}$ is the parameter projection vector. The expected value of q_T is $n-l$ (where the solution is effectively constrained to satisfy l known parameter estimates).

Since it is desired to retain the specified prior estimates in the extremal inversion, we minimize the function

$$\mathbf{m}^T\mathbf{b} + \frac{1}{2\mu}\{(\mathbf{Wd} - \mathbf{Wf(m)})^T(\mathbf{Wd} - \mathbf{Wf(m)}) + \boldsymbol{\beta}[\mathbf{Dm} - \mathbf{h}]^T\boldsymbol{\beta}^T[\mathbf{Dm} - \mathbf{h}] - q_T\}. \tag{37}$$

Linearizing the problem as usual leads to the approximate function

$$\mathscr{L} = (\mathbf{m}^0 + \mathbf{x})^T\mathbf{b} + \frac{1}{2\mu}\{(\mathbf{Wy} - \mathbf{WAx})^T(\mathbf{Wy} - \mathbf{WAx}) + \boldsymbol{\beta}[\mathbf{D}(\mathbf{m}^0 + \mathbf{x}) - \mathbf{h}]^T\boldsymbol{\beta}^T[\mathbf{D}(\mathbf{m}^0 + \mathbf{x}) - \mathbf{h}] - q_T\}. \tag{38}$$

Differentiating $\mathscr{L}$ with respect to $\mathbf{x}$ and setting equal to zero yields the solution for the parameter corrections

$$\mathbf{x} = [\mathbf{A}^T\mathbf{EA} + \mathbf{D}^T\mathbf{BD}]^{-1}\{\mathbf{A}^T\mathbf{Ey} + \mathbf{D}^T\mathbf{B}(\mathbf{h} - \mathbf{Dm}^0) - \mu\mathbf{b}\} \tag{39}$$

or the direct parameter estimates

$$\mathbf{m}_b = [\mathbf{A}^T\mathbf{EA} + \mathbf{D}^T\mathbf{BD}]^{-1}\{(\mathbf{WA})^T\hat{\mathbf{d}} + \mathbf{D}^T\mathbf{Bh} - \mu\mathbf{b}\} \tag{40}$$

where

$$\mu = \pm\left(\frac{q_T - q_{LS}}{\mathbf{b}^T[\mathbf{A}^T\mathbf{EA} + \mathbf{D}^T\mathbf{BD}]^{-1}\mathbf{b}}\right)^{1/2} \tag{41}$$

and

$$q_{LS} = \mathbf{y}^T\mathbf{Ed} + \mathbf{B}\{\mathbf{h}^T(\mathbf{h} - \mathbf{Dm}^0) + \mathbf{m}^{0T}\mathbf{D}^T(\mathbf{Dm}^0 - \mathbf{h})\} - (\mathbf{y}^T\mathbf{EA} - \mathbf{Bm}^{0T}\mathbf{DD}^T)(\mathbf{A}^T\mathbf{EA} + \mathbf{D}^T\mathbf{BD})^{-1} \times (\mathbf{A}^T\mathbf{Ey} - \mathbf{D}^T\mathbf{BDm}^0)$$

and the other symbols have their usual meanings. The quantity μ is determined by the constraint $q = q_T$.

For the special case where $\mathbf{h}$ is the null vector, we have that

$$\mathbf{x} = [\mathbf{A}^T\mathbf{EA} + \beta^2\mathbf{D}^T\mathbf{D}]^{-1}\{\mathbf{A}^T\mathbf{Ey} - \beta^2\mathbf{D}^T\mathbf{Dm}^0 - \mu\mathbf{b}\} \tag{42}$$

or

$$\mathbf{m} = [\mathbf{A}^T\mathbf{EA} + \beta^2\mathbf{D}^T\mathbf{D}]^{-1}\{(\mathbf{WA})^T\hat{\mathbf{d}} - \mu\mathbf{b}\} \tag{43}$$

where $\hat{\mathbf{d}}$ is as previously defined, and the undetermined quantity μ is given by

$$\mu = \pm\left(\frac{q_T - q_{LS}}{\mathbf{b}^T[\mathbf{A}^T\mathbf{EA} + \beta^2\mathbf{D}^T\mathbf{D}]^{-1}\mathbf{b}}\right)^{1/2} \tag{44}$$

where

$$q_{LS} = \mathbf{y}^T\mathbf{Ed} + \beta^2\mathbf{m}^{0T}\mathbf{D}^T\mathbf{Dm}^0 - (\mathbf{y}^T\mathbf{EA} - \beta^2\mathbf{m}^{0T}\mathbf{DD}^T) \times (\mathbf{A}^T\mathbf{EA} + \beta^2\mathbf{D}^T\mathbf{D})^{-1}(\mathbf{A}^T\mathbf{Ey} - \beta^2\mathbf{D}^T\mathbf{Dm}^0).$$

For the sake of completeness, we will revisit the Marquardt-type damped most-squares problem solved recently by Meju & Hutton (1992, eqs 19–27) and provide a slightly refined definition within the framework of biased estimation. The problem is restated simply as *minimize* (*cf.* Meju & Hutton 1992)

$$\mathbf{m}^T\mathbf{b} + \frac{1}{2\mu}\{[\mathbf{Wd} - \mathbf{Wf(m)}]^T[\mathbf{Wd} - \mathbf{Wf(m)}] + \beta(\mathbf{m} - \mathbf{m}^0)^T(\mathbf{m} - \mathbf{m}^0) - q_T\} \tag{45}$$

and the statistical measure q is defined as

$$q = |\mathbf{Wd} - \mathbf{Wf(m)}|^2 + \beta\,|\mathbf{m} - \mathbf{m}^0|^2. \tag{46}$$

The linearized quantity to be minimized is

$$(\mathbf{m}^0 + \mathbf{x})^T\mathbf{b} + \frac{1}{2\mu}\{(\mathbf{y} - \mathbf{Ax})^T(\mathbf{y} - \mathbf{Ax}) + \beta\mathbf{x}^T\mathbf{x} - q_T\}. \tag{47}$$

Now, eq. (47) is subtly different from that of Meju & Hutton (1992, eq. 22) who sought the extreme models with the maximum permissible data misfit. Here, we are interested in the extreme models with the maximum combined data and model prediction errors. Effecting the minimization in the usual manner, we have that

$$\mathbf{x} = [\mathbf{A}^T\mathbf{EA} + \beta\mathbf{I}]^{-1}\{\mathbf{A}^T\mathbf{Ey} - \mu\mathbf{b}\} \tag{48}$$

or

$$\mathbf{m} = [\mathbf{A}^T\mathbf{EA} + \beta\mathbf{I}]^{-1}\{(\mathbf{WA})^T\hat{\mathbf{d}} + \beta\mathbf{m}^0 - \mu\mathbf{b}\}. \tag{49}$$

When the quadratic constraint ($q = q_T$) is satisfied, we find—as in Meju & Hutton (1992)—that

$$q_T = \mathbf{y}^T\mathbf{Ey} + \mu^2\mathbf{b}^T(\mathbf{A}^T\mathbf{EA} + \beta\mathbf{I})^{-1}\mathbf{b} - \mathbf{y}^T\mathbf{EA}(\mathbf{A}^T\mathbf{EA} + \beta\mathbf{I})^{-1}\mathbf{A}^T\mathbf{Ey} \tag{50}$$

giving

$$\mu = \pm\left(\frac{q_T - q_{LS}}{\mathbf{b}^T[\mathbf{A}^T\mathbf{EA} + \beta\mathbf{I}]^{-1}\mathbf{b}}\right)^{1/2} \tag{51}$$

where

$$q_{LS} = \mathbf{y}^T\mathbf{Ey} - \mathbf{y}^T\mathbf{EA}(\mathbf{A}^T\mathbf{EA} + \beta\mathbf{I})^{-1}\mathbf{A}^T\mathbf{Ey}.$$

Non-linearity is dealt with using an iterative formula of the form

$$\mathbf{m}^{k+1} = \mathbf{m}^k + (\mathbf{A}^T\mathbf{EA} + \beta\mathbf{I})^{-1}(\mathbf{A}^T\mathbf{Ey} - \mu\mathbf{b})$$

or

$$\mathbf{m}^{k+1} = (\mathbf{A}^T\mathbf{EA} + \beta\mathbf{I})^{-1}\{(\mathbf{WA})^T\hat{\mathbf{d}} + \beta\mathbf{m}^k - \mu\mathbf{b}\} \tag{52}$$

where $\mathbf{A}$ and $\mathbf{y}$ are evaluated at $\mathbf{m}^k$.

It is easy to deduce that although eqs (49) and (40) are structurally similar, their underlying philosophies are different. These extremal inversion schemes employ a compensating relationship between the parameters and may be used to calculate the confidence limits of the optimal least-squares model or extreme parameter sets that are maximally consistent with the given field data (e.g. Meju & Hutton 1992). The scheme may be easily shown to share the main features of the common approaches to reducing non-uniqueness in non-linear inversion.

Note that the above techniques are applicable to a variety of geophysical problems. For instance, Meju & Hutton (1993) applied the biased estimation algorithm with null priors in 2-D magnetotelluric inversion and Meju (1994b) show practical examples of how any desirable information can be retained in 1-D inversion of electromagnetic depth sounding data. Using a simple refraction seismology time-term practical problem, Meju (1994b) also demonstrated that a unique solution to an ill-posed linear problem may be obtained simply by providing sufficient constraints on the form of the solution without specifying *ab initio* the true value of any of the sought model parameters. A computer program based on the schemes discussed in this paper for linear parameter estimation and uncertainty analysis is given in Meju (1994a).

CONCLUSIONS

A naïve attempt has been made to develop a unified framework for linear and non-linear inversion and uncertainty analysis with *a priori* information using an easily digestible matrix algebra. The main emphasis in this framework is on the retention of known or desirable prior information in the optimal solutions as a means of reducing non-uniqueness in the interpretive models. Thus, in the estimation part, the goal is to develop a flexible scheme for forcing the optimal solution into conformity with the features of the sought parameters specified *ab initio* by the interpreter. In non-linear estimation, it was deemed necessary to analyse the solutions for consistency with their linear analogues and to clarify the relationships with the more rigorous landmark algorithms for geophysical inversion. The biased estimation approach was shown to yield these least-squares solutions as special cases. The consistency analysis procedure was also used to shed some light on the nature of some of the classical estimation algorithms discussed in the paper.

In the uncertainty analysis part, we exploit the fact that the prior data are independent of the observational data and attempt to develop the appropriate covariance relations for gauging the reliability of the optimal solutions. In the favoured most-squares approach to model appraisal, a general formalism incorporating prior data is presented as an extension of Jackson's (1976, 1979) original treatise on extremal inversion.

ACKNOWLEDGMENTS

The critical reviews and helpful suggestions made by Pascal Tarits and two other reviewers are gratefully acknowledged. This work was greatly influenced by the doctrines of S. Twomey and D. D. Jackson, and I thank E. R. G. Hill for showing me how to transcribe their scriptures and V. R. S. Hutton for suggesting this vocation in my days at Edinburgh University.

REFERENCES

Backus, G.E., 1988. Bayesian reference in geomagnetism, *Geophys. J. Int.*, **92,** 125–142.

Backus, G.E. & Gilbert, J.F., 1970. Uniqueness in the inversion of inaccurate gross earth data, *Phil. Trans. R. Soc.*, A, **266,** 123–192.

Backus, G.E. & Gilbert, J.F., 1968. The resolving power of gross earth data, *Geophys. J. R. astr. Soc.*, **16,** 169–205.

Constable, S.C., Parker, R.L. & Constable, C.G., 1987. Occam's inversion: a practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics*, **52,** 289–300.

Duijndam, A.J.W., 1988a. Bayesian estimation in seismic inversion. Part I: Principles, *Geophys. Prospect.*, **36,** 878–898.

Duijndam, A.J.W., 1988b. Bayesian estimation in seismic inversion. Part II: Uncertainty Analysis, *Geophys. Prospect.*, **36,** 899–918.

Gol'tsman, F.M., 1971. *Statistical Models of Interpretation*, Science (Nauka) Press, Moscow.

Gol'tsman, F.M., 1975. Statistical theory for the interpretation of geophysical fields, *Izv., Earth Phys.*, **1,** 29–53 (translated by M. N. Pillai).

Hoerl, A.E. & Kennard, R.W., 1970. Ridge regression: biased estimation for nonorthogonal problems, *Technometrics*, **12,** 55–67.

Jackson, D.D., 1972. Interpretation of inaccurate, insufficient, and inconsistent data, *Geophys. J. R. astr. Soc.*, **28,** 97–109.

Jackson, D.D., 1976. Most squares inversion, *J. geophys. Res.*, **81,** 1027–1030.

Jackson, D.D., 1979. The use of *a priori* data to resolve non-uniqueness in linear inversion, *Geophys. J. R. astr. Soc.*, **57,** 137–157.

Jackson, D.D. & Matsu'ura, M., 1985. A Bayesian approach to nonlinear inversion, *J. geophys. Res.*, **90,** 581–591.

Kaula, W.M., 1966. *Theory of Satellite Geodesy*, Blaisdell, Waltham, MA.

Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters, *J. Soc. Ind. appl. Math.*, **11,** 431–441.

Marquardt, D.W., 1970. Generalized inverses, ridge regression, biased linear estimation and nonlinear estimation, *Technometrics*, **12,** 591–612.

Meju, M.A., 1992. An effective ridge regression procedure for resistivity inversion, *Comput. Geosci.*, **18,** 99–118.

Meju, M.A., 1993. Inversion for smooth geoelectromagnetic models and the effective depth of inference: an interpretative analysis, *Geophysics*, submitted.

Meju, M.A., 1994a. A general program for linear parameter estimation and uncertainty analysis, *Comput. Geosci.*, **20,** 197–220.

Meju, M.A., 1994b. *Geophysical Data Analysis: understanding inverse problem theory and practice*, Soc. Expl. Geophys., Course Notes Series, SEG Publ., Tulsa, OK, in press.

Meju, M.A. & Hutton, V.R.S., 1992. Iterative most-squares inversion: application to magnetotelluric data, *Geophys. J. Int.*, **108,** 758–766.

Meju, M.A. & Hutton, V.R.S., 1993. The geoelectric structure across the Great Glen Fault, northern Scotland, *Geophys. J. Int.*, submitted.

Menke, W., 1984. *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press, Orlando, FL.

Meyer, S.L., 1977. *Data Analysis for Scientists and Engineers*, John Wiley and Sons, New York.

Tarantola, A. & Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys. Space Phys.*, **20,** 219–232.

Tikhonov, A.N., 1963. Regularization of Ill-posed Problems, *Dokl. Akad. Nauk SSSR*, **153,** 1–6.

Tikhonov, A.N. & Arsenin, V.Y., 1977. *Solutions of Ill-posed Problems*, John Wiley and Sons, New York.

Twomey, S., 1963. On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature, *J. Assoc. Comput. Man.*, **10,** 97–101.

Twomey, S., 1977. *An Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements*, Elsevier Scientific Publ. Co., Amsterdam.