Geophysical Prospecting **36**, 899–918, 1988

BAYESIAN ESTIMATION IN SEISMIC INVERSION. PART II: UNCERTAINTY ANALYSIS[1]

A. J. W. DUIJNDAM[2]

ABSTRACT

DUIJNDAM, A.J.W. 1988. Bayesian estimation in seismic inversion. Part II: Uncertainty analysis. *Geophysical Prospecting* **36**, 899–918.

A parameter estimation or inversion procedure is incomplete without an analysis of uncertainties in the results. In the fundamental approach of Bayesian parameter estimation, discussed in Part I of this paper, the *a posteriori* probability density function (pdf) is the solution to the inverse problem. It is the product of the *a priori* pdf, containing *a priori* information on the parameters, and the likelihood function, which represents the information from the data. The maximum of the *a posteriori* pdf is usually taken as a point estimate of the parameters. The shape of this pdf, however, gives the full picture of uncertainty in the parameters. Uncertainty analysis is strictly a problem of information reduction. This can be achieved in several stages. Standard deviations can be computed as overall uncertainty measures of the parameters, when the shape of the *a posteriori* pdf is not too far from Gaussian. Covariance and related matrices give more detailed information. An eigenvalue or principle component analysis allows the inspection of essential linear combinations of the parameters.

The relative contributions of *a priori* information and data to the solution can be elegantly studied. Results in this paper are especially worked out for the non-linear Gaussian case. Comparisons with other approaches are given. The procedures are illustrated with a simple two-parameter inverse problem.

INTRODUCTION

Inversion techniques based on parameter estimation are becoming increasingly important in seismics. In Part I of this paper (Duijndam 1988, hereafter referred to as 'Part I') the principles of Bayesian estimation are discussed. They are briefly reviewed below. In the Bayesian approach parameters are estimated by combining information from data with *a priori* information on the parameters. The first type of information is reflected in the likelihood function, the second in the *a priori* probability density function (pdf). The product of the two determines the *a posteriori* pdf, which is the solution to the inverse problem. Because it is practically impossible for all but the most trivial problems to inspect the *a posteriori* pdf through the whole of

[1] Received May 1987, revision accepted December 1987.

[2] Delft Geophysical B.V., P.O. Box 148, 2600 AC Delft, The Netherlands.

parameter space, its maximum is taken as a point estimate of the parameters. Such a point estimate (which consists of a number for each parameter) represents a limited amount of information. One needs to be aware of when the *a posteriori* pdf is not sharply peaked and a large number of parameters have almost equal likelihood of corresponding to the true model. An analysis of uncertainties is, therefore, considered important.

The two basic questions to be answered in such an analysis procedure are: (1) How well is the estimate determined by the *a posteriori* pdf, i.e. by the combination of data and *a priori* information? (2) What are the respective contributions of data and *a priori* information?

The answer to the first question can be found by inspection of the form of the *a posteriori* pdf. The shape of the pdf around the maximum determines how well the estimate is resolved from the information available. The second question can be answered by comparison of the *a priori* pdf with the likelihood function.

Uncertainty analysis is closely related to what, in geophysical literature, is usually referred to as resolution analysis. The terms 'resolution from the total state of information', or briefly 'resolution' and 'resolution from the data' are used. The first term expresses how well the estimate is determined by the combination of *a priori*, observational and theoretical knowledge. The latter expresses to what extent the parameter estimates are determined by the combination of theoretical and observational information (or, in more popular terms, 'the data') in comparison with the *a priori* information. Note that with these definitions parameters can be well resolved by the total state of information while being poorly resolved from the data.

The uncertainty analysis problem is essentially one of information reduction. This reduction can be achieved in several ways depending on how much information is desired. First the functions can be inspected along directions of interest. These will in practice be directions along which the parameters are poorly resolved. A stronger reduction of information can be obtained by computation of the *a posteriori* covariance matrix. From this covariance matrix the standard deviations of the parameters can be computed. These can be interpreted as overall uncertainty bounds on the parameters and will often be the most convenient and simplest, albeit limited, form of information.

Part II shows that these approaches of uncertainty or resolution analysis are strongly related to methods of resolution analysis as proposed by other authors for linear or linearized forward models. For Gaussian assumptions and non-linear models these procedures can be performed more accurately when second derivatives are taken into account in the computation of the so-called Hessian matrix.

The above possibilities for uncertainty analysis are illustrated with the two-parameter example used in Part I. It is briefly reviewed below.

BAYESIAN ESTIMATION

Let the vector $\mathbf{y}$ contain discretized data and the vector $\mathbf{x}$ contain parameters describing a physical model. The solution to the inverse problem is given by Bayes'

rule

$$p(\mathbf{x} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}. \tag{1}$$

The *a posteriori* pdf $p(\mathbf{x} \mid \mathbf{y})$ is the pdf of the parameters $\mathbf{x}$ given $\mathbf{y}$. The function $p(\mathbf{y} \mid \mathbf{x})$ viewed as a function of $\mathbf{x}$ is called the likelihood function. *A priori* information on the parameters is contained in $p(\mathbf{x})$. The denominator $p(\mathbf{y})$ is a constant in the inverse problem and can be considered immaterial. When a forward model $\mathbf{g}(\mathbf{x})$ is available for the computation of synthetic data the so-called standard reduced model,

$$\mathbf{y} = \mathbf{g}(\mathbf{x}) + \mathbf{n}, \tag{2}$$

can be used to construct the likelihood function. The vector $\mathbf{n}$ contains the noise. In Part I it is shown that for Gaussian distributions for noise and *a priori* information the MAP estimator, which is the maximum of the *a posteriori* pdf is the vector $\hat{\mathbf{x}}$ that minimizes the weighted l_2 norm:

$$2F(\mathbf{x}) = (\mathbf{y} - \mathbf{g}(\mathbf{x}))^T \mathbf{C}_n^{-1} (\mathbf{y} - \mathbf{g}(\mathbf{x})) + (\mathbf{x}^i - \mathbf{x})^T \mathbf{C}_x^{-1} (\mathbf{x}^i - \mathbf{x}), \tag{3}$$

where $\mathbf{C}_n$ is the covariance matrix of the noise, $\mathbf{x}^i$ is the *a priori* model (mean of the *a priori* pdf) and $\mathbf{C}_x$ is the *a priori* covariance matrix.

Part I illustrates the principles of Bayesian inversion for a two-parameter problem. Because the uncertainty analysis is also illustrated with this example it is quickly reviewed here. The problem is a one-dimensional (1D) seismic inverse problem and concerns the estimation of the acoustic impedance and the thickness in traveltime of a thin layer embedded in a homogeneous medium. The first parameter is referred to as the difference in acoustic impedance ΔZ of the thin layer with the background. The second parameter is the thickness $\Delta\tau$ in traveltime. The forward model used is the convolutional model (primaries only). The vector $\mathbf{g}(\mathbf{x})$ contains the samples of the synthetic trace $s(t)$, given by

$$s(t) = \frac{\Delta Z}{2Z + \Delta Z} \{w(t - \tau_1) - w(t - (\tau_1 + \Delta\tau))\}, \tag{4}$$

where Z is the acoustic impedance of the background, $w(t)$ the wavelet and τ_1 the position of the upper boundary of the layer. The true model, the wavelet and the noisy data used are shown in Part I, Fig. 2. The signal-to-noise ratio is 3 dB and assumed to be known in the problem. The *a priori* information used on the parameters are independent Gaussian pdf s with means (*a priori* model) given by

$$\Delta Z^i = 3.4 \times 10^6 \text{ kg m}^{-2} \text{ s}^{-1}, \tag{5}$$

$$\Delta\tau^i = 3 \text{ ms}$$

and standard deviations by

$$\sigma_{\Delta Z} = 0.5 \times 10^6 \text{ kg m}^{-2} \text{ s}^{-1},$$

$$\sigma_{\Delta\tau} = 2 \text{ ms}. \tag{6}$$

The 2D pdfs for the problem are given in Part I, Fig. 4. The maximum of the *a posteriori* pdf is found by minimizing (3), using an optimization algorithm.

Inspection of the PDFs along Principal Components

The *a posteriori* pdf $p(\mathbf{x} \mid \mathbf{y})$ determines the resolution from the total state of information. The most relevant information can be obtained in a practical way and without approximations by determining essential directions in the parameter space and by plotting the relevant functions along these directions.

Principles

Suppose F is the objective function to be minimized in order to find the maximum of the *a posteriori* pdf. For the particular examples in this paper $F(\mathbf{x}) = -\ln(p(\mathbf{x} \mid \mathbf{y}))$. We are interested in the region where $F(\mathbf{x})$ is close to $F(\hat{\mathbf{x}})$. In particular we may want to know in what region the difference $F(\mathbf{x}) - F(\hat{\mathbf{x}})$ is smaller than some specified number ε. This is also called an ε-indifference region (Bard 1974). Following Bard (1974), $F(\mathbf{x})$ is expanded in a Taylor series around $\hat{\mathbf{x}}$, retaining only the first terms,

$$F(\mathbf{x}) \approx F(\hat{\mathbf{x}}) + \mathbf{q}^T \Delta\mathbf{x} + \tfrac{1}{2}\Delta\mathbf{x}^T \mathbf{H} \Delta\mathbf{x}, \tag{7}$$

where

$$\Delta\mathbf{x} = \mathbf{x} - \hat{\mathbf{x}}. \tag{8}$$

The gradient $\mathbf{q}$ is defined by:

$$q_i = \frac{\partial F}{\partial x_i}, \tag{9}$$

and $\mathbf{H}$ is the so-called Hessian Matrix of second derivatives:

$$H_{ij} = \frac{\partial^2 F}{\partial x_i \partial x_j}. \tag{10}$$

When $F(\hat{x})$ is a minimum, $\mathbf{H}$ is positive semi-definite. In the sequel only an unconstrained minimum is discussed. The constrained minimum is much more complex (Bard 1974) and will not be treated here. At an unconstrained minimum the gradient vanishes ($\mathbf{q} = 0$) and (7) becomes

$$F(\mathbf{x}) = F(\hat{\mathbf{x}}) + \tfrac{1}{2}\Delta\mathbf{x}^T \mathbf{H} \Delta\mathbf{x}. \tag{11}$$

For the boundary of the ε-indifference region we find

$$2\varepsilon = 2(F(\mathbf{x}) - F(\hat{\mathbf{x}})) = \Delta\mathbf{x}^T \mathbf{H} \Delta\mathbf{x}. \tag{12}$$

This is the equation of an N-dimensional ellipsoid. The ellipsoids corresponding to different values of ε are concentric and similar in shape; see Part I, Fig. 5c, where

the contours of the *a posteriori* pdf for the two-parameter example are plotted. Far away from the maximum the approximation (11) is no longer valid for non-linear problems and the ellipses deform. This is clearly visible in Part I, Fig. 5c. The so-called canonical form of (12) can be obtained by using the eigenvalue decomposition of the positive semi-definite matrix **H**.

$$\mathbf{H} = \mathbf{V}\Lambda^2\mathbf{V}^T, \tag{13}$$

where **V** is the matrix which columns are the normalized eigenvectors of **H** and the diagonal matrix Λ contains the non-negative eigenvalues, denoted by λ_i^2. The principle axes of the ellipsoid are in the direction of the so-called vector of canonical parameters which are defined as $\psi = \mathbf{V}^T\mathbf{x}$. With this definition (12) can be written as

$$\varepsilon = \psi^T\Lambda^2\psi = \sum_i \lambda_i^2 \psi_i^2. \tag{14}$$

This is the so-called canonical form. The principle axes of the ellipsoids correspond with the coordinate axes in the ψ space, the eigenvectors of **H**. From (14) it follows that the lengths of the principle axes are inversely proportional to the square roots λ_i of the eigenvalues. This means that the ellipsoids are stretched in the directions corresponding to low eigenvalues. Along these directions the parameters are poorly resolved.

A practical and still detailed uncertainty and resolution analysis can be carried out by scanning the parameter space along the principle axes and plotting the pdfs or other desired functions like the data mismatch, etc. Especially those directions along which the parameters are poorly resolved will be important in practice.

Scaling or transformation of parameters

It should be realized that in general the eigenvalue decomposition of the Hessian matrix as expressed in (13) is physically meaningless. No physical units can be assigned to the elements of the matrices **V** and Λ such that (13) is consistent. This has been pointed out by Tarantola (1987) for the closely-related eigenvalue decomposition of a covariance matrix. The decomposition can, nevertheless, be carried out as simply a numerical procedure acting on the values that are contained in the matrix **H** in some unit system. However, it is, in general, not invariant with respect to linear transformations of the parameters. Simply expressing some of the parameters in different units will change the results of the eigenvalue decomposition. After such a change the eigenvectors correspond to different physical parameter models than before the change. Consider the linear transformation from $\Delta\mathbf{x}$ to $\Delta\tilde{\mathbf{x}}$

$$\Delta\mathbf{x} = \mathbf{D}\Delta\tilde{\mathbf{x}}, \tag{15}$$

where **D** is a non-singular square matrix. Substitution of the parameter transformation in (11) yields

$$F(\mathbf{D}\mathbf{x}) = F(\mathbf{D}\hat{\mathbf{x}}) + \tfrac{1}{2}\Delta\tilde{\mathbf{x}}^T\mathbf{D}^T\mathbf{H}\mathbf{D}\Delta\tilde{\mathbf{x}}. \tag{16}$$

Let the eigenvalue decomposition of the transformed Hessian matrix be given by:

$$\tilde{\mathbf{H}} = \mathbf{D}^T\mathbf{H}\mathbf{D} = \mathbf{W}\Pi^2\mathbf{W}^T. \tag{17}$$

It can be shown using (13) and (17) that the eigenvectors of $\tilde{\mathbf{H}}$ correspond to those of $\mathbf{H}$ according to the linear transformation (15),

$$\mathbf{V} = \mathbf{DW}, \tag{18}$$

if and only if, $\mathbf{D}$ is a so-called unitary matrix defined by

$$\mathbf{D}^T = \mathbf{D}^{-1}. \tag{19}$$

The directions of the eigenvectors are preserved for the less restrictive condition

$$\mathbf{DD}^T = \mathbf{D}^T\mathbf{D} = c\mathbf{I}, \tag{20}$$

with c an arbitrary constant. The difference between (20) and (19) is an irrelevant constant factor for all parameters.

The question arises in which units to express the parameters or, starting from some unit system, which linear transformation to apply such that the eigenvalue decomposition is most meaningful. Simply expressing the parameters in SI units can be a bad choice because parameters of different types can be of different numerical orders. More useful options are: (1) to express the parameters in units that correspond with the parameter ranges of interest for the problem; (2) to take the *a priori* standard deviations as units; and (3) to statistically normalize the parameters with the transformation $\mathbf{D} = \mathbf{C}_x^{1/2}$, so that the transformed parameters have the identity matrix as *a priori* covariance matrix. Each of these options renders the transformed parameters dimensionless. Option 3 is an interesting one, especially when considering the least-squares problem which is analysed below.

The least-squares problem

In the least-squares problem the objective function F is given by (3) and is written here in the form

$$F = \tfrac{1}{2}\mathbf{e}^T\mathbf{e}, \tag{21}$$

with $\mathbf{e}$ the partitioned vector:

$$\mathbf{e} = \begin{pmatrix} \mathbf{C}_n^{-1/2}(\mathbf{d} - \mathbf{g}(\mathbf{x})) \\ \mathbf{C}_x^{-1/2}(\mathbf{x}^i - \mathbf{x}) \end{pmatrix}. \tag{22}$$

The Jacobian matrix $\mathbf{J} = \partial\mathbf{e}/\partial\mathbf{x}$ is partitioned accordingly

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_y \\ \mathbf{J}_x \end{bmatrix} = -\begin{bmatrix} \mathbf{C}_n^{-1/2}\mathbf{A} \\ \mathbf{C}_x^{-1/2} \end{bmatrix}, \tag{23}$$

with $\mathbf{A} = \partial\mathbf{g}/\partial\mathbf{x}$. It can easily be verified that the Hessian $\mathbf{H}$ can be written as the sum

$$\mathbf{H} = \mathbf{H}_y + \mathbf{H}_x, \tag{24}$$

where $\mathbf{H}_y$ is the contribution of the data. It stems from the likelihood function at this point x, and is given by (see Duijndam 1987):

$$\mathbf{H}_y = \mathbf{J}_y^T\mathbf{J}_y + \sum_{i=1}^{N_d} e_i\mathbf{T}_i, \tag{25}$$

where N_d is the number of data points and $\mathbf{T}_i$ is the Hessian matrix of the residual point e_i at this point. $\mathbf{H}_x$ is due to the *a priori* information and reads,

$$\mathbf{H}_x = \mathbf{J}_x^T \mathbf{J}_x = \mathbf{C}_x^{-1}. \tag{26}$$

After a linear transformation the transformed Hessian will be

$$\begin{aligned} \tilde{\mathbf{H}} &= \mathbf{D}^T \mathbf{H} \mathbf{D}, \\ &= \mathbf{D}^T \mathbf{H}_y \mathbf{D} + \mathbf{D}^T \mathbf{H}_x \mathbf{D}. \end{aligned} \tag{27}$$

Using the statistical normalization $\mathbf{D} = \mathbf{C}_x^{1/2}$ in (27) leads to

$$\tilde{\mathbf{H}} = \mathbf{C}_x^{1/2} \mathbf{H}_y \mathbf{C}_x^{1/2} + \mathbf{I}. \tag{28}$$

Let the eigenvalue decomposition of the first term be given by

$$\begin{aligned} \tilde{\mathbf{H}}_y &= \mathbf{C}_x^{1/2} \mathbf{H}_y \mathbf{C}_x^{1/2}, \\ &= \mathbf{V} \Lambda_y \mathbf{V}^T. \end{aligned} \tag{29}$$

The eigenvalue decomposition of $\tilde{\mathbf{H}}$ is then

$$\tilde{\mathbf{H}} = \mathbf{V}(\Lambda_y + \mathbf{I})\mathbf{V}^T. \tag{30}$$

The eigenvalue spectrum of $\tilde{\mathbf{H}}$ is built up by two terms

$$\lambda_i^2 = \lambda_{yi} + 1. \tag{31}$$

The first term is due to the data, the second due to the *a priori* information. Because $\tilde{\mathbf{H}}$ is positive definite (positive semi-definite in exceptional cases) all eigenvalues are non-negative and can therefore be written as squares. It follows from (31) that the diagonal values λ_{yi} of Λ_y cannot be less than -1. The larger value on the right-hand side of (31) most strongly determines the curvature of the objective function F along the direction of the eigenvector $\mathbf{v}_i$. Therefore, directions for which $\lambda_{yi} < 1$ can be defined as ill-resolved (with respect to the data!) or weak directions. In these directions the *a priori* information more strongly determines the *a posteriori* pdf than the likelihood function does.

For low residuals and/or for sufficiently linear models the second term of the right-hand side of (25) vanishes. Then we have, for the transformed data part of the Hessian,

$$\begin{aligned} \tilde{\mathbf{H}}_y &= \mathbf{C}_x^{1/2} \mathbf{J}^T \mathbf{J} \mathbf{C}_x^{1/2}, \\ &= \mathbf{C}_x^{1/2} \mathbf{A}^T \mathbf{C}_n^{-1} \mathbf{A} \mathbf{C}_x^{1/2}, \\ &= \tilde{\mathbf{A}}^T \tilde{\mathbf{A}}, \end{aligned} \tag{32}$$

where $\tilde{\mathbf{A}}$ is the sensitivity or forward matrix of the linear or linearized problem, weighted with the uncertainties of noise and *a priori* information:

$$\tilde{\mathbf{A}} = \mathbf{C}_n^{-1/2} \mathbf{A} \mathbf{C}_x^{1/2}. \tag{33}$$

The approximation of $\tilde{\mathbf{H}}_y$ as given in (32) is positive semi-definite and its eigenvalue decomposition can, therefore, be written as

$$\tilde{\mathbf{H}}_y = \mathbf{V} \mathbf{S}^2 \mathbf{V}^T, \tag{34}$$

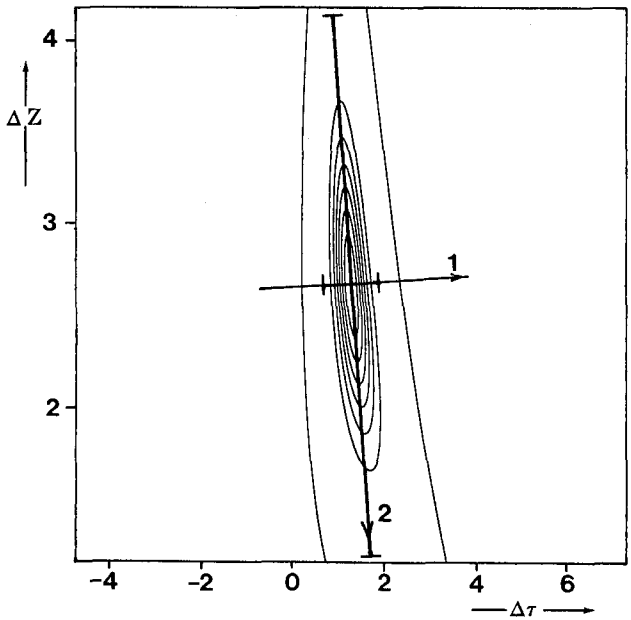

FIG. 1. The eigenvectors in the two-parameter example. The ranges used for plotting of the pdfs are indicated.

where the matrices **V** and **S** are (by definition) exactly the same as the ones occurring in the singular value decomposition (SVD) of $\tilde{\mathbf{A}}$:

$$\tilde{\mathbf{A}} = \mathbf{USV}^T. \tag{35}$$

These results clearly establish the link between this approach of uncertainty and resolution analysis and those as given for linear models using the SVD by for example, Jackson (1973) and Van Riel and Berkhout (1985). Using (25) in full form is of course more accurate for non-linear models.

The two-parameter example

The procedure can be nicely illustrated for the two-parameter example as discussed above. The parameters are statistically normalized with the transformation matrix $\mathbf{D} = \mathbf{C}_x^{1/2}$. The eigenvalue analysis is based on (28), with the approximated form of the data part as given in (32). The singular values s_i of the reweighted forward matrix are

$$s_1 = 9.083,$$
$$s_2 = 0.127. \tag{36}$$

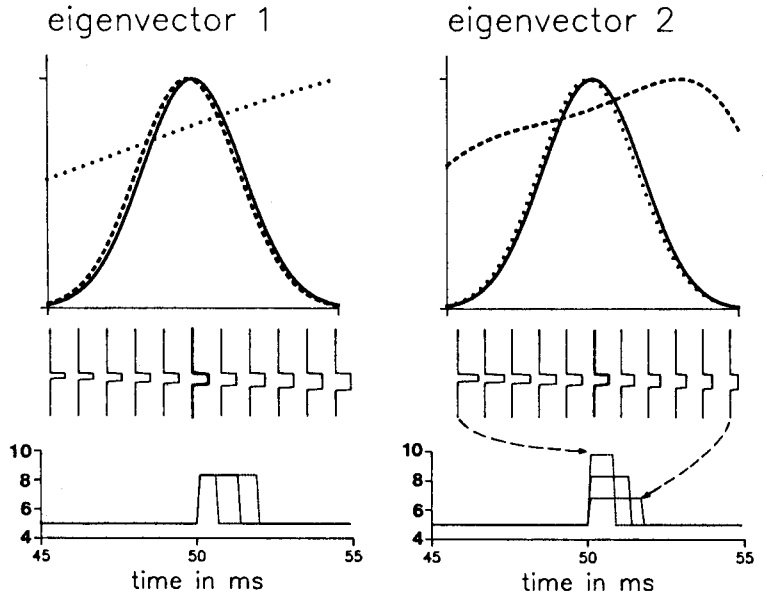

FIG. 2. Pdfs along the eigenvectors. The corresponding models are plotted along the x-axis, with the estimated model in thick line. The estimated and the extreme models are also plotted in the bottom pictures. (· · · ·) *a priori* pdf; (– – –) likelihood function; (——) *a posteriori* pdf.

The square roots of the eigenvalues of $\tilde{\mathbf{H}}$ follow from (32):

$$\lambda_1 = 9.138,$$
$$\lambda_2 = 1.008. \tag{37}$$

The principal axis in direction 2 is approximately 9 times as long as the one in direction 1. This can be verified in Fig. 1, where the contours of the *a posteriori* pdf are again given. The ranges plotted along the x and y axis are equal in terms of *a priori* standard deviations. The eigenvectors are given by the columns of the matrix $\mathbf{V}$:

$$\mathbf{V} = \begin{bmatrix} 0.0684 & -0.9977 \\ 0.9977 & 0.0684 \end{bmatrix}. \tag{38}$$

It can easily be verified that $\mathbf{V}$ is orthonormal. The directions of the eigenvectors are drawn in Fig. 1. The second direction, corresponding with the small eigenvalue, is in the direction of decreasing impedance and increasing layer thickness. Along this direction the synthetic data $\mathbf{g}(\mathbf{x})$ and hence the likelihood function varies only slowly.

The uncertainty analysis along eigenvectors can be done by plotting the *a priori* pdf, the likelihood function and the *a posteriori* pdf around the maximum of the *a posteriori* pdf. The ranges that are visualized are indicated in Fig. 1. The plots of the pdfs themselves are given in Fig. 2. The ranges (three units in the canonical system) are chosen such that the *a posteriori* pdf fits nicely in the plot. Along the x-axes the

corresponding impedance models are plotted. For a more precise evaluation of the ranges of the models the middle (estimated) and the 'extreme' models are again plotted at the bottom of the pictures. Jackson's (1973, 1976) so-called edgehog and most-squares methods for computing extreme models are closely related to the procedure described here. It is clear that along direction 1 the likelihood function determines the *a posteriori* pdf or, in more popular terms, the data determines the answer. Along direction 2 on the other hand, the *a priori* information determines the answer.

In this example the plots hardly add any information to what we already know from the full 2D plots. In, for example, a 10D problem, however, the full pdfs cannot be plotted while ten plots along the eigenvectors are still practical and may give a good idea of the functions involved.

THE *A POSTERIORI* COVARIANCE MATRIX

Computation

An analysis of uncertainties or resolution can be given in more concise form by covariance matrices. For non-linear models an exact computation of the *a posteriori* covariance matrix would involve a full computation of the *a posteriori* pdf through the whole of parameter space. This is computationally too demanding for all but the most trivial cases. Fortunately a good approximation is possible when the forward model is linear enough. Equation (11) gives the approximation of the objective function F around the maximum $\hat{\mathbf{x}}$

$$F(\mathbf{x}) \approx F(\hat{\mathbf{x}}) + \tfrac{1}{2}(\mathbf{x} - \hat{\mathbf{x}})^T\mathbf{H}(\mathbf{x} - \hat{\mathbf{x}}). \tag{39}$$

F is the negative logarithm of the *a posteriori* pdf, so that

$$p(\mathbf{x}\,|\,\mathbf{y}) \approx p(\hat{\mathbf{x}}\,|\,\mathbf{y}) \exp\{-\tfrac{1}{2}(\mathbf{x} - \hat{\mathbf{x}})^T\mathbf{H}(\mathbf{x} - \hat{\mathbf{x}})\}. \tag{40}$$

Thus around the maximum the *a posteriori* pdf looks like a Gaussian distribution with mean $\mathbf{x} = \hat{\mathbf{x}}$ and covariance matrix $\mathbf{C}_{\hat{x}} = \mathbf{H}^{-1}$. This holds irrespective of the distributions involved for noise and *a priori* information, provided of course that the approximation (39) exists. For linear models and Gaussian assumptions (39) and (40) are exact. The *a posteriori* covariance is then given by

$$\begin{aligned}\mathbf{C}_{\hat{x}} &= \mathbf{H}^{-1},\\ &= (\mathbf{J}^T\mathbf{J})^{-1},\\ &= (\mathbf{A}^T\mathbf{C}_n^{-1}\mathbf{A} + \mathbf{C}_x^{-1})^{-1},\end{aligned} \tag{41}$$

where (23) has been used. This is a well-known result, see e.g. Bard (1974). For non-linear models relations (24) and (25) have to be used. Note that only when the *a posteriori* pdf is low enough far from the maximum and approximation (40) holds for the whole of parameter space, the inverse Hessian $\mathbf{H}^{-1}$ is a good approximation for the *a posteriori* covariance matrix.

It is enlightening to consider two extreme situations:

1. The data fully determines the answer; the shape of the *a posteriori* pdf is fully determined by the likelihood function. We then have

$$\mathbf{C}_{\hat{x}} = \mathbf{H}_y^{-1}, \tag{42}$$

with $\mathbf{H}_y$ given in (25). For linear models the covariance matrix is equal to that of the well-known Gauss-Markov estimator:

$$\mathbf{C}_{\hat{x}} = (\mathbf{A}^T\mathbf{C}_n^{-1}\mathbf{A})^{-1}. \tag{43}$$

Note, however, that the *a posteriori* pdf differs conceptually from the sampling distribution from which, in classical statistics, the covariance matrix for the Gauss-Markov estimator is derived.

2. When the *a priori* information fully determines the answer we have

$$\mathbf{C}_{\hat{x}} = \mathbf{H}_x^{-1} = \mathbf{C}_x, \tag{44}$$

meaning that the uncertainty after inversion is as big as before inversion. The eigenvalue decomposition can provide some more insight into the nature of the *a posteriori* covariance matrix. Consider the parameters to be weighted with $\mathbf{D} = \mathbf{C}_x^{1/2}$ (statistically normalized), and $\mathbf{H}_y$ to be positive definite (true for linear models). We then have, using (28) and (34)

$$\begin{aligned}\tilde{\mathbf{C}}_{\hat{x}} &= \tilde{\mathbf{H}}^{-1}, \\ &= \mathbf{V}(\mathbf{S}^2 + \mathbf{I})^{-1}\mathbf{V}^T, \\ &= \sum_i \frac{1}{s_i^2 + 1}\mathbf{v}_i\mathbf{v}_i^T.\end{aligned} \tag{45}$$

The values s_i are inversely proportional to the square root of the noise power. It can be seen from (45) that in the extreme of the noise power going to zero, the *a posteriori* covariance matrix will still not vanish when there are s_i with values of zero. There will always remain a term

$$\tilde{\mathbf{C}}_{\hat{x}} = \sum_{i \in O} \mathbf{v}_i\mathbf{v}_i^T, \tag{46}$$

where O denotes the set of eigenvectors for which $s_i = 0$. Along these directions the data does not provide any information, no matter how low the noise level. It also follows that without *a priori* information, when the term 1 in the denominator is not present, the covariance matrix becomes very large when some of the s_i are very small.

The two-parameter example

Comparison of the *a posteriori* covariance matrix with the *a priori* covariance matrix reveals how much information has been gained by using the data. An illustration is again given for the two-parameter example. The *a priori* and *a posteriori* covariance matrices are:

$$\mathbf{C}_x = \begin{bmatrix} 2.5 \times 10^{11} & 0 \\ 0 & 4 \times 10^{-6} \end{bmatrix}, \qquad \mathbf{C}_{\hat{x}} = \begin{bmatrix} 2.45 \times 10^{11} & -66 \\ -66 & 6.8 \times 10^{-8} \end{bmatrix}$$

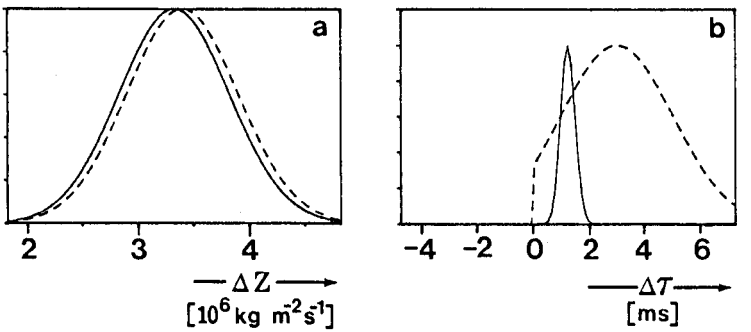

FIG. 3. *A priori* and approximated *a posteriori* pdfs for the acoustic impedance ΔZ (a) and the time thickness $\Delta\tau$ (b). (· · · · ·) *a priori* pdf; (——) *a posteriori* pdf.

with the SI units:

$$\begin{bmatrix} \text{kg}^2\,\text{m}^{-4}\,\text{s}^{-2} & \text{kg}\,\text{m}^{-2} \\ \text{kg}\,\text{m}^{-2} & \text{s}^2 \end{bmatrix}.$$

An absolute interpretation of the *a posteriori* covariance in such a unit system is not easy. A direct comparison with the *a priori* covariance matrix is easier to interpret. This especially holds for the diagonal values which give the variances of the parameters. Their square roots—the standard deviations—are given by:

$$\sigma_{\Delta Z} = 0.5 \times 10^6 \text{ kg m}^{-2} \text{ s}^{-1}, \qquad \sigma_{\hat{\Delta Z}} = 0.495 \times 10^6 \text{ kg m}^{-2} \text{ s}^{-1},$$

$$\sigma_{\Delta\tau} = 2 \times 10^{-3} \text{ s}, \qquad \sigma_{\hat{\Delta\tau}} = 0.26 \times 10^{-3} \text{ s}.$$

They can be interpreted as overall uncertainty measures (although this is not always appropriate when the form of the pdf is far from Gaussian). In Fig. 3 the *a priori* pdfs and the approximated marginal *a posteriori* pdfs are given. In the region where the latter are non-zero they are Gaussian pdfs with the estimated values as means and the standard deviations as given above. A convenient display for this problem is to plot the *a priori* and *a posteriori* models with the standard deviations as uncertainty bouds (see Fig. 4). From the numbers and the figures it is clear that hardly any information has been gained on the acoustic impedance. The standard deviation of the thickness on the other hand has been reduced by a factor of nearly 8. The *a posteriori* standard deviations provide uncertainty intervals for the true model. Figure 4c shows that for this specific example the true model indeed lies within these intervals.

The interpretation of *a posteriori* covariance matrix may be easier when the parameters are reweighted with the *a priori* covariance matrix. We then have for the transformed covariance matrices:

$$\mathbf{C}_x = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \tilde{\mathbf{C}}_{\hat{x}} = \begin{bmatrix} 0.98 & -0.066 \\ -0.066 & 0.017 \end{bmatrix}.$$

The elements are dimensionless. The *a posteriori* covariance matrix can now be compared directly with the identity matrix which is advantageous when, for larger

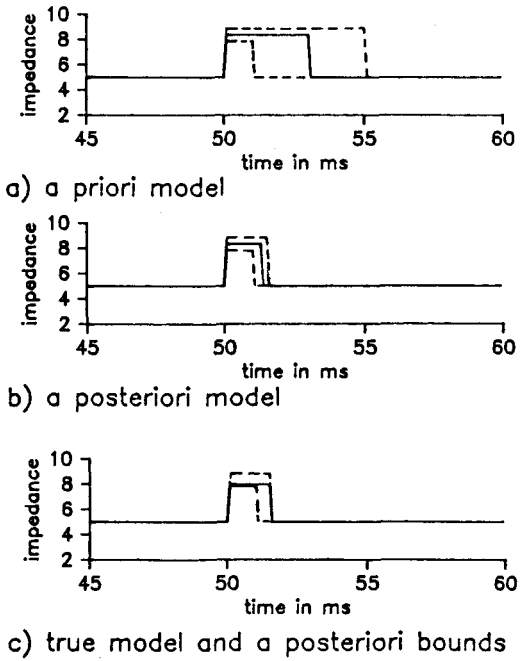

FIG. 4. *A priori* (a) and *a posteriori* (b) model with uncertainty intervals. In (c) the true model is given with the *a posteriori* uncertainty intervals.

problems, the *a posteriori* covariance matrix is plotted rather than given as explicit numbers (see Duijndam 1987). The off-diagonal elements, the covariances, are most easily studied when the covariance matrices are normalized on their variances. The correlation matrix is the result. We get

$$\begin{matrix}\textit{a priori}\\ \text{correlation}\\ \text{matrix}\end{matrix} = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, \qquad \begin{matrix}\textit{a posteriori}\\ \text{correlation}\\ \text{matrix}\end{matrix} = \begin{bmatrix}1 & 0.51\\ 0.51 & 1\end{bmatrix}.$$

There was no correlation in the *a priori* information. The *a posteriori* correlation coefficient of 0.51 indicates that, due to the exchange effect, the parameters are better resolved in one direction than another.

Linearity check

In such an analysis one would like to know over what range the quadratic approximation (39) is valid. Sometimes it can be shown analytically that the forward model is linear enough. A practical numerical procedure is to scan the parameter space along the eigenvectors and to plot the exact and approximated *a posteriori* pdfs. For the two-parameter example this is shown in Fig. 5. Along eigenvector 2 the approximation is excellent. Along eigenvector 1 there are some small differences.

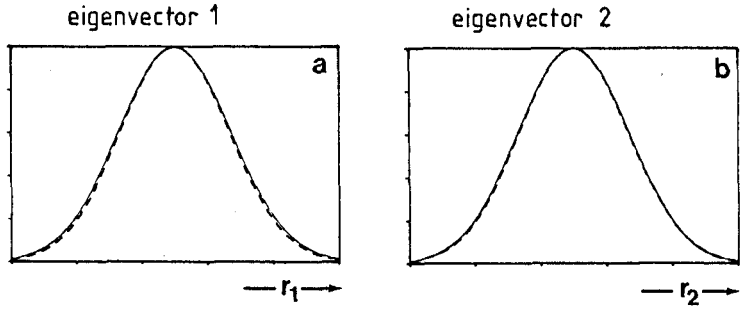

FIG. 5. Linearity analysis along the eigenvectors. (——) true *a posteriori* pdf; (· · · · ·) approximated *a posteriori* pdf.

Standard deviations and marginal pdf's

The standard deviation σ_i of a parameter is defined as the square root of

$$\sigma_i^2 = \int (x_i - Ex_i)^2 p(x_i)\, dx_i, \tag{47}$$

where $p(x_i)$ is the marginal pdf for x_i. The marginal pdf reflects the 'average' information for x_i. The standard deviation σ_i is therefore an 'average' or 'overall' uncertainty measure. For linear models, (40) for the *a posteriori* pdf is exact. The marginal pdf of a Gaussian pdf with mean $\hat{\mathbf{x}}$ and covariance $\mathbf{C}_{\hat{x}}$ is itself Gaussian with mean $\hat{x}_i$ and standard deviation σ_i and is therefore exactly known.

For non-linear models the true marginal pdf may differ from the approximated one as derived from (40). In Fig. 6 this is shown for the two-parameter problem. The true marginal pdf's are computed from the 2D pdf. The approximation is good for the time thickness. The true marginal pdf of the acoustic impedance is slightly shifted to lower values. This can be explained from the contours for lower values of the 2D pdf as shown in Part I, Fig. 5c. A point estimate of ΔZ based on the maximum of the marginal pdf would yield a lower value than the MAP estimate given in Part I (which is too high) and would therefore be a better estimate. This

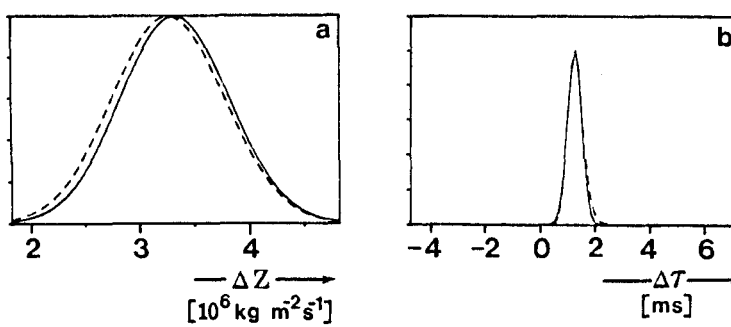

FIG. 6. True and approximated marginal pdf's for the acoustic impedance (a) and the time thickness (b). (——) approximated marginal pdf; (· · · · ·) true marginal pdf.

illustrates that parameters in which one is not interested, so-called nuisance parameters, should ideally be integrated out. In this example, integrating over $\Delta\tau$ yields a better estimate for ΔZ. Unfortunately integrating out the nuisance parameters is often not possible analytically and too costly to be done numerically.

Sampling distributionss

A concept that hardly ever occurs in a Bayesian context but is essential for the frequentist school in statistics is that of the sampling distribution. It arises when the point estimator is considered as a random variable that takes different values for different realizations of the data. The sampling distribution describe the variations of the estimates. The corresponding covariance matrix $\mathbf{C}_{sy}$ can be approximated by (Bard 1974):

$$\mathbf{C}_{sy} \approx \mathbf{H}^{-1}\mathbf{A}^T\mathbf{C}_n^{-1}\mathbf{A}\mathbf{H}^{-1}, \tag{48}$$

for Gaussian assumptions and sufficiently linear models. The Hessian $\mathbf{H}$ and the sensitivity matrix $\mathbf{A}$ are evaluated at the estimated model for the actual data set under consideration. Accordingly, a sampling distribution due to variations in the *a priori* information can also be defined. Its covariance matrix can be approximated by

$$\mathbf{C}_{sx} \approx \mathbf{H}^{-1}\mathbf{C}_x^{-1}\mathbf{H}^{-1}. \tag{49}$$

It has the following interesting intepretation. When the *a priori* information is varied with covariance $\mathbf{C}_x$, the resulting point estimate will vary with covariance $\mathbf{C}_{sx}$. For the linear Gaussian case Jackson and Matsu'ura (1985) derived the combination of these two expressions as follows (although they do not mention the concept of sampling distribution). The Bayesian estimator for the linear Gaussian case can be written as the sum of two operators, working on data and '*a priori* data' respectively:

$$\hat{\mathbf{x}} = \mathbf{K}\mathbf{y} + \mathbf{L}\mathbf{x}^i \tag{50}$$

with

$$\mathbf{K} = \mathbf{H}^{-1}\mathbf{A}^T\mathbf{C}_n^{-1} \tag{51}$$

and

$$\mathbf{L} = \mathbf{H}^{-1}\mathbf{C}_x^{-1}. \tag{52}$$

$\mathbf{H}$ is the Hessian for the linear case:

$$\mathbf{H} = \mathbf{A}^T\mathbf{C}_n^{-1}\mathbf{A} + \mathbf{C}_x^{-1}. \tag{53}$$

Jackson and Matsu'ura compute the '*a posteriori*' covariance, which is in fact the sampling covariance $\mathbf{C}_s$, due to data and *a priori* information. It follows from (50)

$$\begin{aligned}\mathbf{C}_s &= \mathbf{K}\mathbf{C}_n^{-1}\mathbf{K}^T + \mathbf{L}\mathbf{C}_x^{-1}\mathbf{L}^T, \\ &= \mathbf{H}^{-1}\mathbf{A}^T\mathbf{C}_n^{-1}\mathbf{A}\mathbf{H}^{-1} + \mathbf{H}^{-1}\mathbf{C}_x^{-1}\mathbf{H}^{-1}.\end{aligned} \tag{54}$$

Although the right-hand side, the sum of the two sampling covariance matrices, is indeed equal to the *a posteriori* covariance as given in (41), this is not obvious. The *a posteriori* pdf is a fundamentally different concept from the one of sampling distribution. It is nevertheless interesting to study the extremes of (54). For very low noise levels the first term on the right-hand side of (54) will dominate the second one, assuming **A** has no zero singular values. This is not surprising. For low noise levels the data determines the answer so that varying the *a priori* information does not alter the solution much. For very accurate *a priori* information the second term will dominate the first one.

As an illustration the terms appearing in (54) are given for the two-parameter example in a statistically-normalized parameter system:

$$\begin{bmatrix} 0.980 & -0.066 \\ -0.066 & 0.017 \end{bmatrix} = \begin{bmatrix} 0.016 & 0 \\ 0 & 0.012 \end{bmatrix} + \begin{bmatrix} 0.964 & -0.066 \\ -0.066 & 0.005 \end{bmatrix}.$$

Thus, for the acoustic impedance, the '*a priori* sampling distribution' is dominant while for the thickness the data sampling distribution is dominant. Within the numerical accuracy in this example, the off-diagonal elements of the total sampling covariance matrix $\mathbf{C}_s$ are entirely due to those of the sampling distribution of the *a priori* information. Again this need not be surprising. Providing randomly uncorrelated *a priori* models yield a correlation in the sampling distribution of the estimates because the shape of the likelihood function forces a preferred linear combination of parameter estimates. On the other hand, when keeping the *a priori* information fixed and changing the data at random, no correlation enters the distribution of the estimates, because the *a priori* information does not force preferred linear combinations.

In a true Bayesian interpretation sampling distributions are unnatural concepts. Their use is not recommended. The *a priori* and *a posteriori* covariance matrices provide all the information required.

The Resolution Matrix

Another concept, the resolution matrix, is often introduced in estimation problems. Like the sampling distributions it is not a natural concept in a Bayesian analysis and its interpretation is not without problems in the non-linear case. To derive it an (approximate) linear relation between the estimated and the true parameter values has to be found. The Gaussian case is analysed here. The position of the maximum of the *a posteriori* pdf is denoted by $\mathbf{x}_m$. We can write for the estimated model, see Duijndam (1987):

$$\mathbf{H}(\hat{\mathbf{x}} - \mathbf{x}_m) = -\mathbf{J}^T\mathbf{e}, \tag{55}$$

where **H** and **J** are evaluated at $\mathbf{x}_m$. The actual data **d** as it occurs in **e** can be written as

$$\mathbf{d} = \mathbf{g}(\mathbf{x}_t) + \mathbf{n}_t, \tag{56}$$

where $\mathbf{x}_t$ and $\mathbf{n}_t$ denote the true model and the true noise respectively. If $\mathbf{g}(\mathbf{x})$ can be approximated by the first terms of a Taylor expansion around the maximum $\mathbf{x}_m$, we have for $\mathbf{x} = \mathbf{x}_t$

$$\mathbf{g}(\mathbf{x}_t) \approx \mathbf{g}(\mathbf{x}_m) + \mathbf{A}(\mathbf{x}_t - \mathbf{x}_m), \tag{57}$$

where, again, $\mathbf{A}$ is evaluated at $\mathbf{x}_m$. We can substitute this in (57) and (56) to obtain

$$\mathbf{H}(\hat{\mathbf{x}} - \mathbf{x}_m) = -\mathbf{J}^T \begin{pmatrix} \mathbf{C}_n^{-1/2}\mathbf{A}(\mathbf{x}_t - \mathbf{x}_m) + \mathbf{C}_n^{-1/2}\mathbf{n}_t \\ \mathbf{C}_x^{-1/2}(\mathbf{x}^i - \mathbf{x}_m) \end{pmatrix}. \tag{58}$$

The resolution matrix $\mathbf{R}$ is defined through the linear relation between the estimated and the true model:

$$\hat{\mathbf{x}} - \mathbf{x}_m = \mathbf{R}(\mathbf{x}_t - \mathbf{x}_m) + \mathbf{c}, \tag{59}$$

with $\mathbf{c}$ a vector of constants. The combination of (58) with (59) and using (23) yields

$$\mathbf{R} = \mathbf{H}^{-1}\mathbf{A}^T\mathbf{C}_n^{-1}\mathbf{A}. \tag{60}$$

Again, for non-linear models this expression is more accurate than the expression usually given for the linear case, see e.g. Jackson (1979):

$$\mathbf{R} = (\mathbf{A}^T\mathbf{C}_n^{-1}\mathbf{A} + \mathbf{C}_x^{-1})^{-1}\mathbf{A}^T\mathbf{C}_n^{-1}\mathbf{A}. \tag{61}$$

It is stressed that the expressions for the resolution matrix (60) and (61) are only valid when approximation (57) is valid, i.e. when the true model is close enough to the maximum of the *a posteriori* pdf. This is, of course, never known in practice, so we never know whether $\mathbf{R}$ is the true operator that maps the true parameters on the estimated ones. The *a posteriori* pdf, however, gives the probability that the true model *is* close to the maximum. Linearizing at another model makes no sense. If it has a higher probability of being close to the true model it should replace the maximum as the point estimate.

From (41) and (61) an interesting relation between the resolution matrix and the *a posteriori* covariance matrix can be derived for the linear Gaussian case:

$$\mathbf{R} = \mathbf{I} - \mathbf{C}_{\hat{x}}\mathbf{C}_x^{-1}. \tag{62}$$

This relation has also been found by Tarantola (1987). The extreme situations can again provide some insight. When the data fully determines the answer we have $\mathbf{C}_{\hat{x}} \ll \mathbf{C}_x$ and thus $\mathbf{R} = \mathbf{I}$, meaning full resolution from the data. When the *a priori* information fully determines the answer we have $\mathbf{C}_{\hat{x}} = \mathbf{C}_x$ and thus $\mathbf{R} = 0$, meaning no resolution from the data at all. From (62) it can be seen that the resolution matrix is typically a measure of the balance between the information supplied by the data and the *a priori* information.

When the parameters are statistically normalized on the *a priori* information we have the simple relation

$$\tilde{\mathbf{R}} = \mathbf{I} - \tilde{\mathbf{C}}_{\hat{x}}. \tag{63}$$

The interesting features of the resolution matrix are the deviations from the identity matrix and these turn out to be given exactly by the *a posteriori* covariance matrix

$\tilde{\mathbf{C}}_x$! The relation with the eigenvalue analysis is immediately clear for the linear normalized case. Using (45) we can rewrite (63) as

$$\tilde{\mathbf{R}} = \mathbf{I} - \sum_i \frac{1}{s_i^2 + 1} \mathbf{v}_i \mathbf{v}_i^T. \tag{64}$$

The low singular values s_i determine the deviations of $\tilde{\mathbf{R}}$ from the ideal shape $\mathbf{I}$. Jackson and Matsu'ura (1985) and Tarantola (1987) suggested the following interpretation of the traces of matrices $\mathbf{I}$, $\tilde{\mathbf{R}}$ and $\mathbf{I} - \tilde{\mathbf{R}}$:

$$\text{trace } (\mathbf{I}) = n, \qquad \text{the total number of parameters,}$$

$$\text{trace } (\tilde{\mathbf{R}}) = \sum_i \frac{s_i^2}{s_i^2 + 1}, \text{ the number of parameters determined by the data,}$$

$$\text{trace } (\mathbf{I} - \tilde{\mathbf{R}}) = \sum_i \frac{1}{s_i^2 + 1}, \text{ the number of parameters determined by the a priori information.}$$

The expressions with the singular values hold only for the linear normalized case.

The resolution matrix for the two-parameter example is

$$\mathbf{R} = \begin{bmatrix} 0.02 & 1.65 \times 10^7 \text{ kg m}^{-2} \text{ s}^{-2} \\ 2.64 \times 10^{-10} \text{ kg}^{-1} \text{ m}^2 \text{ s}^2 & 0.983 \end{bmatrix}.$$

The off-diagonal elements are not dimensionless, unless all parameters are of the same type. It is clear that the resolution matrix in the normalized system is much easier to interpret

$$\tilde{\mathbf{R}} = \begin{bmatrix} 0.02 & 0.066 \\ 0.066 & 0.983 \end{bmatrix}.$$

The acoustic impedance is poorly resolved from the data but the thickness is well resolved.

Summary of Uncertainty Analysis

In an uncertainty or resolution analysis four types of quantities can be distinguished, from simple to complex: (1) standard deviations, (2) covariance matrices, (3) eigenvalue spectra, (4) pdf s.

Standard deviations are the simplest type of information. They can be interpreted as overall uncertainty bounds on the parameters when the *a posteriori* pdf does not deviate too much from the Gaussian form, and will be comprehensible to a person with little knowledge of statistics. Often they can be displayed in an easy to comprehend way. *A priori* and *a posteriori* standard deviations can be compared to assess how much information has been gained from the data. Remember, however, that standard deviations form a limited amount of information. Strong correlations between parameters are *not* visualized. Furthermore, the problem has to be linear

enough to allow the computation of the *a posteriori* covariance matrix from which they are derived.

More detailed information can be derived from covariance and correlation matrices. As above, *a priori* and *a posteriori* covariance matrices can be compared for an assessment of the information brought in by the data. For a comprehensible display, the parameters can best be statistically normalized on the *a priori* information. The *a priori* covariance matrix is then equal to the identity matrix. For the *a posteriori* correlations, the correlation matrix may be more useful than the covariance matrix. Note that for non-linear problems only an *approximation* of the covariance matrix can be computed. The problem has to be linear enough for the approximation to be accurate. If it is not, the computed matrix may still be useful, because it describes the curvature of the *a posteriori* pdf around the maximum.

The third type of information are eigenvalue spectra. The square roots of eigenvalues are most useful because they are inversely proportional to the lengths of the ellipsoids that are the contours of the *a posteriori* pdf around the maximum. Again it is most useful to statistically transform the parameters according to the *a priori* information, especially for the Gaussian problem. The *a priori* spectrum then has a value of 1. Where the eigenvalues of the data part of the Hessian matrix are lower than 1, the *a priori* information determines the answer for the corresponding linear combination of parameters. From these plots an impression can be obtained along how many directions the data determines the answer.

The information from the eigenvalue spectra can be augmented by plotting the functions (*a priori* pdf, likelihood function and *a posteriori* pdf) along desired directions. This gives some impression of the actual shape of the functions around the estimated model. Unlike the covariance matrix, these plots can be computed without approximations.

It is stated again that for Gaussian problems the Hessian matrix can be computed more accurately than is done in most studies by taking the second derivatives of the forward model into account. It is also emphasized that the eigenvalue analysis is still useful when one rejects the probabilistic basis of the inversion approach. Instead of pdfs, quantities like energy of data mismatch, etc., can then be plotted. The procedures sketched here are used and illustrated in two companion papers on the detailed inversion of post-stack seismic data (see also Duijndam 1987).

ACKNOWLEDGEMENTS

I wish to thank Professor A. J. Berkhout and G. J. Lörtzer for critically reviewing the manuscript and Miss T. van Lier for typing it. Thanks are also due to Delft Geophysical for permission to publish this paper.

REFERENCES

BARD, Y. 1974. *Nonlinear Parameter Estimation.* Academic Press, Inc.

DUIJNDAM, A.J.W. 1987. Detailed Bayesian inversion of seismic data. Ph.D. thesis, Delft University of Technology, Delft, The Netherlands.

DUIJNDAM, A.J.W. 1988. Bayesian estimation in seismic inversion. Part I: Principles. *Geophysical Prospecting* **36**, 878–898.

JACKSON, D.D. 1973. Marginal solutions to quasi-linear inverse problems in geophysics: the edgehog method. *Geophysical Journal of the Royal Astronomical Society* **35**, 121–136.

JACKSON, D.D. 1976. Most-squares inversion. *Journal of Geophysical Research* **81**, 1027–1030.

JACKSON, D.D. 1979. The use of *a priori* data to resolve non-uniqueness in linear inversion. *Geophysical Journal of the Royal Astronomical Society* **57**, 137–157.

JACKSON, D.D. and MATSU'URA, M. 1985. A Bayesian approach to nonlinear inversion. *Journal of Geophysical Research* **90**, 581–591.

TARANTOLA, A. 1987. *Inverse Problem Theory, Methods for Data Fitting and Model Parameter Estimation.* Elsevier Science Publishing Co.

VAN RIEL, P. and BERKHOUT, A.J. 1985. Resolution in seismic trace inversion by parameter estimation. *Geophysics* **50**, 1440–1455.