Geophys. J. R. astr. Soc. (1984) **76**, 299–306

Three-dimensional inversion without blocks

A. Tarantola and A. Nercessian *Laboratoire de Sismologie (LEGSP LA195 CNRS), Institut de Physique du Globe de Paris, Université Pierre et Marie Curie, 4, place Jussieu, Tour 14, 75230 Paris Cédex 05, France*

Received 1983 February 21; in original form 1982 May 17

Summary. We propose a method for solving non-linear inverse problems in the case where the unknown is a function of the spatial coordinates and the data set is discrete and finite. The method is based on a generalized least-squares criterion, it is defined directly for non-linear problems (without previous linearization of the forward problem), and in the particular linear case it gives the same results (although slightly more general) than the Backus & Gilbert approach. As an example, we apply the method to the three-dimensional seismic velocity inverse problem, using as data the arrival times of seismic waves. The following paper (Nercessian *et al.*) shows some results obtained using the present method.

1 Introduction

In many geophysical inverse problems the unknown is a function of the spatial coordinates. One way of solving such inverse problems is by parameterizing the unknown function. This parameterization may be done in several ways, as for example by developing the function on a truncated basis, or, more commonly, by assuming that the function takes constant values in blocks of an *a priori* division of the space. In both cases the unknown function is replaced by a discrete and finite number of unknown parameters.

An alternative approach consists of taking the *function* itself as the unknown, without *a priori* parameterization, and defining the inverse problem as the search for a particular function verifying some *a priori* assumptions and allowing computed values of data which are reasonably close to the observed values. A first and general approach to this problem was done, in the linear case, by Backus & Gilbert (1970). In particular, they argued that the solution of such a functional, linear, inverse problem, can always be interpreted as a filtered version of the 'true solution'. The preferred solution is then the one for which the corresponding filter is 'the closest' to the identity filter. This means that from the Backus & Gilbert point of view, we can define preferred solutions for whichever different criteria of closeness we use. The simplest results are obtained when the traditional least-squares criterion of closeness is used.

The aim of this paper is to show that when using the least-squares criterion at a much more preliminary step of the argument, a theory is obtained that contains as a special case

the Backus & Gilbert solution, but which is more general in the sense that: (1) it takes naturally into account the estimated error distribution in the data set – and also a wide class of *a priori* assumptions on the character of the solution; and (2) it is directly defined for non-linear problems (with the corollary that although our solution and the Backus & Gilbert solution are coincident for linear problems, the solution which is obtained by an iterative use of the Backus & Gilbert solution for a *linearized* forward problem is not coincident with our non-linear solution).

We now develop some results demonstrated in Tarantola & Valette (1982), which will be referred to as Paper I.

2 The general inverse problem

Let $m(\mathbf{r})$ represent a *model* of the Earth (i.e. a function of the spatial coordinates). We assume that $m(\mathbf{r})$ is defined inside a volume V which will represent the entire Earth or a part of it. It is well known that if we restrict the space of functions to be considered to the set such that for any two functions $m_1(\mathbf{r})$ and $m_2(\mathbf{r})$ the integral $\int d\mathbf{r}\, m_1(\mathbf{r})\, m_2(\mathbf{r})$ is defined, then this set of functions forms a vector Hilbert space for which some general theorems are available. The functions $m_1(\mathbf{r})$ and $m_2(\mathbf{r})$, when considered as elements of the vectorial space will be simply noted $\mathbf{m}_1$ and $\mathbf{m}_2$. The *scalar product* of $\mathbf{m}_1$ and $\mathbf{m}_2$ is noted $\mathbf{m}_1^* \mathbf{m}_2$ and defined by

$$\mathbf{m}_1^* \mathbf{m}_2 = \int d\mathbf{r}\, m_1(\mathbf{r})\, m_2(\mathbf{r}). \tag{1}$$

Let d^i represent a set of observable quantities whose values depend on the actual value of the model $\mathbf{m}$ through an equation of the form

$$d^i = g^i(\mathbf{m}) \qquad (i = 1, \ldots, n) \tag{2}$$

where the g^i are a set of non-linear functionals which are assumed to be known. Equation (2) represents the solution of the *forward problem*. The quantities $(d^1, \ldots, d^n)$ are named the *data set* and are noted $\mathbf{d}$ for short. Equation (2) is then rewritten

$$\mathbf{d} = \mathbf{g}(\mathbf{m}). \tag{3}$$

We assume that we have performed a physical experiment that has furnished the 'observed values'

$$\mathbf{d}_0 = (d_0^1, \ldots, d_0^n)$$

for the data set. Experimental uncertainties are assumed to be conveniently described using a covariance matrix C_{d_0}.

Let $m_0(\mathbf{r})$ be the *a priori* model, i.e. a model that has been defined without using the observed values of the data set, and which should be preferred in the absence of any effective measurement of the data set. Let $C_{m_0}(\mathbf{r}, \mathbf{r}')$ be the covariance function describing our confidence in the *a priori* model $m_0(\mathbf{r})$. We recall here that a covariance function is a straightforward generalization of a covariance matrix: the value $C_{m_0}(\mathbf{r}, \mathbf{r})$ represents the variance at the point $\mathbf{r}$ and the value $C_{m_0}(\mathbf{r}, \mathbf{r}')$ represents the covariance between points $\mathbf{r}$ and $\mathbf{r}'$. For instance, the simplest analytical forms that can be used for $C_{m_0}(\mathbf{r}, \mathbf{r}')$ are:

$$C_{m_0}(\mathbf{r}, \mathbf{r}') = \sigma^2 \exp[-\tfrac{1}{2}(\mathbf{r}-\mathbf{r}')^2/L^2]$$

$$C_{m_0}(\mathbf{r}, \mathbf{r}') = \sigma^2 \exp(-|\mathbf{r}-\mathbf{r}'|/L)$$

$$C_{m_0}(\mathbf{r}, \mathbf{r}') = \begin{cases} \sigma^2 \text{ if } |\mathbf{r}-\mathbf{r}'| < L \\ 0 \text{ otherwise} \end{cases}$$

where σ represents the *a priori* uncertainty at each point, and where L represents the correlation length of uncertainties. Intuitively such choices correspond to the *a priori* assumption that we can accept, as solutions to the inverse problem, functions whose deviations from $m_0(\mathbf{r})$ are, at each point, of the order of σ, and such that if at a given point there is a deviation from $m_0(\mathbf{r})$ of given sign and magnitude, we want in a neighbourhood of $\mathbf{r}$ deviations of the same sign and similar magnitude, i.e. we want the deviation from $m_0(\mathbf{r})$ to be *smooth*, with smoothness length equal to L.

Of course, more complicated forms may be used for the *a priori* covariance function, in particular, forms depending effectively on $\mathbf{r}$ *and* $\mathbf{r}'$, and not only on the difference $\mathbf{r}-\mathbf{r}'$.

As we will see later, the use of that kind of *a priori* information will allow us to obtain in a very intuitive manner results related with the so-called 'regularization' techniques (Tikhonov 1963; Mikhlin 1970).

The matrix C_{d_0} and the function $C_{m_0}(\mathbf{r}, \mathbf{r}')$ can be considered as the kernels of linear operators acting respectively in the data space and in the model space. The corresponding operators are named *covariance operators*, and will be noted here respectively C_{d_0} and C_{m_0}. It is well known that covariance operators are, by definition, positive semi-definite (see for instance Pugachev 1965). If we assume that they do not contain null or infinite variances, or perfect correlations, they will be positive definite, i.e. regular operators. Their inverses, $C_{d_0}^{-1}$ and $C_{m_0}^{-1}$, can then be uniquely defined.

The generalized non-linear least-squares problem can now be stated as the problem of finding the pair $\mathbf{d}$ and $\mathbf{m}$, which is, among all the pairs obeying the forward equation $\mathbf{d} = \mathbf{g}(\mathbf{m})$, the closest to the pair $\mathbf{d}_0$ and $\mathbf{m}_0$ in the least squares sense. More precisely, $\mathbf{d}$ and $\mathbf{m}$ are defined by the condition

$$(\mathbf{d}_0-\mathbf{d})^* \, C_{d_0}^{-1} \, (\mathbf{d}_0-\mathbf{d}) + (\mathbf{m}_0-\mathbf{m})^* \, C_{m_0}^{-1} \, (\mathbf{m}_0-\mathbf{m}) \text{ minimum} \tag{4}$$

under the constraint

$$\mathbf{d} = \mathbf{g}(\mathbf{m}). \tag{5}$$

As shown in Paper I, the solution of the problem satisfies

$$\mathbf{m} = \mathbf{m}_0 + C_{m_0} G^* (C_{d_0} + G \, C_{m_0} G^*)^{-1} \{\mathbf{d}_0 - \mathbf{g}(\mathbf{m}) + G(\mathbf{m}-\mathbf{m}_0)\} \tag{6}$$

where the linear operator G is the (Fréchet) derivative at the point $\mathbf{m}$ of the non-linear operator $\mathbf{g}$, and where G^* is the adjoint of G.

To solve equation (6) we may use a fixed point algorithm:

$$\mathbf{m}_{k+1} = \mathbf{m}_0 + C_{m_0} G_k^* (C_{d_0} + G_k C_{m_0} G_k^*)^{-1} \{\mathbf{d}_0 - \mathbf{g}(\mathbf{m}_k) + G_k(\mathbf{m}_k - \mathbf{m}_0)\} , \tag{7}$$

where G_k is the derivative of $\mathbf{g}$ at the point $\mathbf{m}_k$.

Our experience in using this algorithm for solving non-linear problems has shown that the convergence is good and that a correct solution is obtained in general after a few iterations (the number of which ranges, say, between 1 and 10), unless the problem is strongly non-linear in which case the algorithm may not converge.

To make effective use of equation (7) we must rewrite it in terms of the kernels of the operators. Let us note such kernels as $G^i(\mathbf{r})$. We have

$$m_{k+1}(\mathbf{r}) = m_0(\mathbf{r}) + \sum_i W_k^i \int d\mathbf{r}' \, C_{m_0}(\mathbf{r}, \mathbf{r}') \, G_k^i(\mathbf{r}') \tag{8a}$$

where

$$W_k^i = \sum_j (S_k^{-1})^{ij} V_k^j, \tag{8b}$$

$$S_k^{ij} = (C_{d_0})^{ij} + \int d\mathbf{r} \int d\mathbf{r}'\, G_k^i(\mathbf{r})\, C_{m_0}(\mathbf{r}, \mathbf{r}')\, G_k^j(\mathbf{r}') \tag{8c}$$

and

$$V_k^j = (\mathbf{d}_0)^j - g^j(\mathbf{m}_k) + \int d\mathbf{r}\, G^j(\mathbf{r})\, \{m_k(\mathbf{r}) - m_0(\mathbf{r})\}. \tag{8d}$$

The integrals in equations (8) should be evaluated using traditional techniques of numerical integration (i.e. by defining a grid in the space dense enough to warrant sufficient accuracy). For particular problems, more astute techniques can be imagined for computing these sums more economically (see for instance Section 3).

The solution for a linear problem can readily be obtained from equation (6). That a problem is linear means that the forward problem takes the form

$$\mathbf{d} = \mathbf{g}(\mathbf{m}) = G\mathbf{m}$$

where the linear operator G is independent of $\mathbf{m}$ and coincides with the derivative of $\mathbf{g}$. We obtain

$$\mathbf{m} = \mathbf{m}_0 + C_{m_0} G^* (C_{d_0} + G C_{m_0} G^*)^{-1} (\mathbf{d}_0 - G\mathbf{m}_0), \tag{9}$$

and we see that the solution is explicit (no iterations are needed). The solution for this linear case was first obtained in a different (although equivalent) form by Franklin (1970).

In the linear case the *a posteriori* covariance operator in the model space can easily be computed (see Paper I):

$$C_m = (I - A)\, C_{m_0} \tag{10}$$

where I is the identity operator and where A is defined by

$$A = C_{m_0} G^* (C_{d_0} + G C_{m_0} G^*)^{-1} G. \tag{11}$$

The explicit form of equation (10) is

$$C_m(\mathbf{r}, \mathbf{r}'') = \int d\mathbf{r}'\, \{\delta(\mathbf{r} - \mathbf{r}') - A(\mathbf{r}, \mathbf{r}')\}\, C_{m0}(\mathbf{r}', \mathbf{r}''). \tag{12}$$

If we were able to perform a set of measurements (i.e. to use sets of functionals g^i) such that we had

$$A(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}'),$$

then the *a posteriori* covariance operator C_m would vanish, i.e. the model would be perfectly resolved. We see then that the function $A(\mathbf{r}, \mathbf{r}')$ defined by (11) generalizes the definition of the 'resolving kernel' of Backus & Gilbert, integrating naturally the estimated error distribution in the data set and the *a priori* information on the model. To take the comparison further, let us introduce the 'true model' $\mathbf{m}_T$ and assume that, by chance, we have obtained observed values for the data set which are *not* contaminated by errors, i.e. such that they are identical to the values predicted by the forward problem for the true model:

$$\mathbf{d}_0 = G\mathbf{m}_T.$$

Using (9) and (11) we obtain

$$\mathbf{m} - \mathbf{m}_0 = A(\mathbf{m}_T - \mathbf{m}_0),$$

i.e. the computed correction to the *a priori* model is a filtered version of the true correction, with the filter A. The solution $\mathbf{m}$ obtained by Backus & Gilbert [when they use as a criterion of closeness between $A(\mathbf{r}, \mathbf{r}')$ and $\delta(\mathbf{r} - \mathbf{r}')$ a least squares criterion] is obtained as a special case of (9) by assuming that experimental errors are negligible ($C_{d0} = 0$) and that the *a priori* information is non-existent (C_{m_0} = constant I).

Although we have shown that the concept of resolving the kernel can be introduced in our theory, it is clear that all the information about resolution that can be extracted from the knowledge of $A(\mathbf{r}, \mathbf{r}')$ can also be extracted from the knowledge of $C_m(\mathbf{r}, \mathbf{r}')$, with the advantage that the covariance operator also gives the absolute error (variance), and allows an easy interpretation of the results in terms of probability.

We have been able to give a closed form for the *a posteriori* covariance operator only for linear problems. For a rigorous computation in the non-linear case, we would need to compute not only the derivatives of the functionals g^i, but also higher order derivatives, and the corresponding formulae become too complicated to be useful. Nevertheless, when the algorithm (7) has converged to the true solution of the non-linear problem and if the non-linearity of the functionals g^i is not too strong in the vicinity of the solution, then a reasonable approximation to the *a posteriori* operator C_m can be obtained using formulae (10–11) where the derivative G is taken at the point where we decide to stop the iterations.

Traditionally, to solve a non-linear inverse problem, the method is to 'linearize' the forward problem around some 'starting point', and to use a linear inverse method for obtaining a 'corrected point'. This corrected point is used as a starting point for the next computation, and this procedure is iterated until convergence. As the problem we are dealing with here is essentially under-determined, the use of this traditional approach should ask for each corrected point to be close to the previous point. This method is not the same as we follow in this paper, where it is asked for the *final* model to be close to the *a priori* model. Although the results between the two approaches do not differ dramatically, our experience shows that the second approach converges faster (besides the fact that we believe it to be fundamentally more coherent). (See Paper I for a more detailed discussion.)

Our final remark will be that our algorithm needs the inversion of the matrix

$$S = C_{d_0} + G C_{m_0} G^*.$$

It is easy to prove that if C_{d_0} is regular, S will also be regular. Our experience shows that for all reasonable choices of C_{d_0} and C_{m_0}, the matrix is never ill-conditioned, whatever the derivative G can be.

3 3-D seismic velocity inverse problem using arrival time data

The pioneering work in that domain was made, using a block technique, by Aki, Christofferson & Husebye (1977) for distant earthquakes, and by Aki & Lee (1976) for local earthquakes. More recently, some papers have applied the Backus & Gilbert approach to this problem (Chou & Booker 1979; Yanovskaya 1980; Thomson & Gubbins 1982).

An array of stations is assumed at the surface of the Earth. These stations record signals from distant or local earthquakes. The problem consists of using the arrival time data at the array to infer the 3-D model of seismic velocities under the array. The two main physical assumptions to be made are: (1) the travel time is obtained by integration along the geometric ray, and (2) the phase from which we use the arrival time is unambiguously determined (which is not so easy for a true 3-D heterogeneous medium). Furthermore, we assume in this paper that the position of each one of the sources is exactly known; if it is not the case, these positions have to be introduced as unknowns into the inverse problem. Spencer & Gubbins (1980) show the best manner of doing that.

Let $n(\mathbf{r})$ denote the slowness of the medium (or $\mathbf{n}$ in compact notation), and t^i the travel time for the ith ray path. The solution of the forward problem can be written

$$t^i = g^i(\mathbf{n}) = \int_{R^i(\mathbf{n})} ds^i \, n(\mathbf{r}^i) \tag{13}$$

where

$$\int_{R^i(\mathbf{n})} ds^i$$

denotes integration along the ith ray path, and where the symbol $\mathbf{n}$ in $R^i(\mathbf{n})$ recalls that the ray path itself depends on $\mathbf{n}$. It is then clear that equation (13) is non-linear.

Let us note by G^i the derivative of the non-linear operator g^i. By definition of the derivative, we must have, $\delta\mathbf{n}$ being any 'small' slowness function,

$$g^i(\mathbf{n}+\delta\mathbf{n}) - g^i(\mathbf{n}) = G^i \delta\mathbf{n} + o(\delta\mathbf{n}^2). \tag{14}$$

We have

$$g^i(\mathbf{n}+\delta\mathbf{n}) - g^i(\mathbf{n}) = \int_{R^i(\mathbf{n}+\delta\mathbf{n})} ds^i \{n(\mathbf{r}^i) + \delta n(\mathbf{r}^i)\} - \int_{R^i(\mathbf{n})} ds^i n(\mathbf{r}^i).$$

Fermat's principle states that the travel time is stationary along the true ray path; in particular, this means that

$$\int_{R^i(\mathbf{n}+\delta\mathbf{n})} ds^i \{n(\mathbf{r}^i) + \delta n(\mathbf{r}^i)\} = \int_{R^i(\mathbf{n})} ds^i \{n(\mathbf{r}^i) + \delta n(\mathbf{r}^i)\} + o(\delta\mathbf{n}^2).$$

We then have

$$g^i(\mathbf{n}+\delta\mathbf{n}) - g^i(\mathbf{n}) = \int_{R^i(\mathbf{n})} ds^i \delta n(\mathbf{r}^i) + o(\delta\mathbf{n}^2),$$

and by comparison with the definition (14), we see that the derivative of the non-linear operator g^i at the point $\mathbf{n}$ is the linear operator G^i that to any function $n'(\mathbf{r})$ associates the number

$$\int_{R^i(\mathbf{n})} ds^i n'(\mathbf{r}^i).$$

If we wish to introduce the kernel of G^i, we must write

$$G^i\mathbf{n}' = \int d\mathbf{r}\, G^i(\mathbf{n};\mathbf{r})\, n'(\mathbf{r}) = \int_{R^i(\mathbf{n})} ds^i n'(\mathbf{r}^i). \tag{15}$$

Where we denote by $G^i(\mathbf{n};\mathbf{r})$ [rather than by $G^i(\mathbf{r})$] the kernel to recall explicitly that it depends on $\mathbf{n}$. We see that the 'function' $G^i(\mathbf{n};\mathbf{r})$ is a delta-like function, null everywhere in the space except along the corresponding ray path.

Let $n_0(\mathbf{r})$ be the *a priori* slowness model, $C_{n_0}(\mathbf{r},\mathbf{r}')$ the *a priori* covariance function, t_0^i the observed arrival times, and $(C_{t_0})^{ij}$ the covariance matrix describing experimental uncertainties. The best model is then obtained using equations (8a–d), (13) and (15). We readily obtain

$$n_{k+1}(\mathbf{r}) = n_0(\mathbf{r}) + \sum_i W_k^i \int_{R^i(\mathbf{n}_k)} ds^i C_{n_0}(\mathbf{r},\mathbf{r}^i) \tag{16a}$$

$$W_k^i = \sum_j (S_k^{-1})^{ij} V_k^j \tag{16b}$$

$$S_k^{ij} = (C_{t_0})^{ij} + \int_{R^i(\mathbf{n}_k)} ds^i \int_{R^j(\mathbf{n}_k)} ds^j C_{n_0}(\mathbf{r}^i, \mathbf{r}^j) \tag{16c}$$

$$V_k^j = t_0^j - \int_{R^j(\mathbf{n}_k)} ds^j n_0(\mathbf{r}^j). \tag{16d}$$

When this iterative process has converged to the final model $\mathbf{n} = \mathbf{n}_\infty$, uncertainties on this *a posteriori* model can be computed approximately using equations (10–11). Using explicit notation, we have:

$$C_{\mathbf{n}}(\mathbf{r}, \mathbf{r}') = C_{n_0}(\mathbf{r}, \mathbf{r}') - \sum_i \sum_j \int_{R^i(\mathbf{n}_\infty)} ds^i \int_{R^j(\mathbf{n}_\infty)} ds^j C_{n_0}(\mathbf{r}, \mathbf{r}^i) (S_\infty^{-1})^{ij} C_{n_0}(\mathbf{r}^j, \mathbf{r}'). \tag{16e}$$

Let us make some comments on these results. The first is that, at each iteration, the rays corresponding to the current model $\mathbf{n}_k$ must be known. To do that, a useful strategy is to compute $n_k(\mathbf{r})$, at each iteration, on the points of a grid which is well adapted to the needs of the ray-tracing routine. It should be noted that, without extra computing cost, we can obtain at each iteration the *gradient* of the slowness model: from equation (16a) we have

$$\nabla n_{k+1}(\mathbf{r}) = \nabla n_0(\mathbf{r}) + \sum_i W_k^i \int_{R^i(\mathbf{n}_k)} ds^i \nabla C_{n_0}(\mathbf{r}, \mathbf{r}^i).$$

This gradient should be useful for most ray-tracing routines.

The second remark is that if the corrections to the *a priori* model are small, the first iteration can give an accurate enough solution. This means in particular that no iterations should be needed, and that if the *a priori* model is simple (homogeneous or layered model), no ray-tracing routines should be needed.

In the block approach to the problem, the main computational task is to divide each ray into the blocks, to form the matrix containing the time spent by each ray in each block, and to multiply this matrix by its transpose. We see here that in the present approach no such matrix exists, and that the main computational task is to evaluate the simple and double integrals along ray paths. Our personal experience shows that both approaches need a computing time of the same order.

As the solution is densely defined (i.e. is defined for each point of the space), we suggest that the most convenient output for the results can be obtained by defining a series of planes representing cross-sections in both horizontal and vertical directions, by representing each one of these planes on a cathodic colour screen terminal, by computing the value of $n_\infty(\mathbf{r})$ at each point of the screen and by plotting the corresponding colour for each point. The absolute *a posteriori* uncertainty, $\sqrt{C_{\mathbf{n}_\infty}(\mathbf{r}, \mathbf{r})}$, should be plotted in the same manner, and if we wish to have an idea of the resolving power attained with the data set, the *a posteriori* covariance function itself, $C_{\mathbf{n}_\infty}(\mathbf{r}_0, \mathbf{r})$, should be plotted, as a function of $\mathbf{r}$, for some chosen points $\mathbf{r}_0$, around which we wish to study the resolution.

In the following paper (Nercessian, Hirn & Tarantola 1984), an example of the results obtained with the present method is shown.

4 Conclusion

The main characteristics of the approach presented here are: (1) it does not need any *a priori* partition of the space into blocks; (2) it is directly defined for non-linear problems and it practically proves to be superior to the approaches based on a 'linearization' of the forward

problem; (3) in the linear case, and if the *a priori* information on the model is neglected, it gives the same results as the Backus & Gilbert approach. We have shown, as a particular application, the formulae obtained for the 3-D seismic inversion problem using arrival time data; probably the most convincing argument in favour of these formulae can be obtained by a look at the results of the accompanying paper (Nercessian *et al.* 1984).

Acknowledgments

We thank our colleagues G. Jobert and B. Valette for helpful discussions and suggestions. This work has partially been supported by the Recherche Coopérative sur Programme 264, Etude Interdisciplinaire des Problèmes Inverses.

References

Aki, K., Christofferson, A. & Husebye, E. S., 1977. Determination of the three-dimensional seismic structure of the lithosphere, *J. geophys. Res.*, **82**, 277–296.

Aki, K. & Lee, W. H. K., 1976. Determination of three-dimensional velocity anomalies under a seismic array using first *P* arrival times from local earthquakes. 1. A homogeneous initial model, *J. geophys. Res.*, **81**, 4381–4399.

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

Chou, W. C. & Booker, R. J., 1979. A Backus–Gilbert approach to inversion of travel-time data for three-dimensional velocity structure, *Geophys. J. R. astr. Soc.*, **59**, 325–344.

Franklin, J. N., 1970. Well-posed stochastic extensions of ill-posed linear problems, *J. math. Analysis Applic.*, **31**, 682–716.

Mikhlin, S. G., 1970. *Mathematical Physics, an Advanced Course,* North-Holland, Amsterdam.

Nercessian, A., Hirn, A. & Tarantola, A., 1984. Three-dimensional seismic transmission prospecting of the Mont Dore volcano, France, *Geophys. J. R. astr. Soc.*, **76**, 307–315.

Pugachev, V. S., 1965. *Theory of Random Functions,* Pergamon Press, Oxford.

Spencer, C. & Gubbins, D., 1980. Travel-time inversion for simultaneous earthquake location and velocity structure determination in laterally varying media, *Geophys. J. R. astr. Soc.*, **63**, 95–116.

Tarantola, A. & Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys. Space Phys.*, **20**, (2) 219–232.

Thomson, C. J. & Gubbins, D., 1982. Three-dimensional lithospheric modelling at NORSAR: linearity of the method and amplitude variations from the anomalies, *Geophys. J. R. astr. Soc.*, **71**, 1–36.

Tikhonov, A. N., 1963. On the solution of improperly posed problems and the method of regularization, *Dokl. Akad. Nauk SSSR*, **151**, 3, 501 (in Russian).

Yanovskaya, T. B., 1980. Solution of the inverse problem in geometrical seismics for a laterally-inhomogeneous medium, in *Methods and Algorithms of Interpretation of Seismological Data,* Nauka, Moscow (Computational seismology, 13) (in Russian).