Geophys. J. Int. (1998) **135,** 304–306

RESEARCH NOTE

Inversion for multiple parameter classes

B. L. N. Kennett and M. Sambridge
Research School of Earth Sciences, Australian National University, Canberra ACT 0200, *Australia*

Accepted 1998 April 4. Received 1998 March 3; in original form 1997 November 24

SUMMARY

Many geophysical data, such as the frequencies of the free oscillations of the Earth, depend on more than one type of model parameter. For inverse problems depending on multiple parameter classes, an iterative solution procedure is introduced in which each parameter class can be treated in the same way. This approach has considerable advantages where a large number of parameters are employed, but can still be useful for smaller systems.

The iteration by parameter class commences by solving for the direct dependence on a particular parameter class, and at subsequent iterations the cross-dependences between classes are introduced. The update affects only the right-hand side of the equations, and, because the same sets of equations have to be solved at each iteration, an efficient computational implementation can be made. The largest set of equations that has to be solved at a time corresponds to the number of variables in an individual parameter class rather than the full set of parameters, which confers substantial computational benefits for very large problems.

Key words: inversion, numerical techniques.

1 INTRODUCTION

In many geophysical inverse problems, the observed data depend on more than one class of parameter. A significant example is presented by the frequencies of the free oscillations of the Earth, which depend on the radial density, *P*- and *S*-velocity distributions and which may also require the introduction of anisotropic parameters to reconcile spheroidal and toroidal modes (see e.g. Dziewonski & Anderson 1981; Montagner & Kennett 1996). Another case is presented by simultaneous seismic tomography for bulk-sound and shear wavespeed (Su & Dziewonski 1997; Kennett, Widiyantoro & van der Hilst 1998). In many tomographic systems it is desirable to include station corrections and an allowance for the uncertainties in the source parameters, as well as determining velocity structure.

However, most of the algorithms that have been employed in the inversions are based on developments for single-parameter systems, with the balance between different variables achieved by rescaling or through variable damping parameters.

Kennett, Sambridge & Williamson (1987) have demonstrated the value of a subspace projection scheme for multiparameter systems and have demonstrated how these could be linked to conjugate-gradient and higher-dimensional search procedures; they illustrated the approach with seismological problems. The subspace approach has also been exploited and extended by Oldenburg, McGillivray & Ellis (1993) in the context of electromagnetic problems. For problems with just two parameter classes, Mao & Stuart (1996) have demonstrated the effectiveness of separating parameters by using a variant of a partitioned inverse.

For problems with very large data sets and many parameter classes of different types, the concept of problem partitioning employed in the subspace methods can be exploited in a different form in the context of linearized inversion. For an N-parameter-class system, the solution of the normal equations can be constructed with an iterative scheme which requires the use of generalized inverses for only the N diagonal blocks.

The algorithm can be regarded as a partitioned Jacobi iteration and can be employed with a wide range of choices for the objective function that is being minimized. The iteration depends on the solution of N sets of equations with variable right-hand vectors, and, in consequence, is well suited to efficient computational implementation.

Kennett *et al.* (1998) have constructed an algorithm of this class for an inverse problem with two parameter classes. In joint seismic tomography for bulk-sound and shear wavespeeds using traveltimes for *P* and *S* phases, they have demonstrated that a practical procedure can be established for very large problems (more than 500 000 equations). The direct construction procedure used by Kennett *et al.* (1998) depends on explicit partitioning of matrices coupled with a series expansion

© 1998 RAS

of matrix inverses and can only be generalized to higher-order systems with considerable algebraic difficulty.

In this note we give a heuristic derivation of a general algorithm using partitions associated with N subsystems of the normal equations for the linearized inversion.

2 INVERSION WITH N PARAMETER CLASSES

Consider a situation with data $\mathbf{d}$ described in terms of a model vector $\mathbf{m}$ composed of a set of N parameter classes, each associated with their own model vector contribution $\{\mathbf{m}_i\}$. We will use roman indices, for example i, to represent the different parameter classes, for example for a model with M parameters grouped into three parameter classes

$$\mathbf{m} = \{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3\},$$

where $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3$ are vectors of lengths p, q, r such that $M = p + q + r$.

The inverse problem to construct the full model $\mathbf{m}$ can be cast in terms of the minimization of an objective function of the form

$$\mathscr{F}(\mathbf{m}) = \Phi(\mathbf{d}, \mathbf{d}_c(\mathbf{m})) + \Psi(\mathbf{m}). \tag{2.1}$$

Here $\Phi(\mathbf{d}, \mathbf{d}_c(\mathbf{m}))$ is a measure of data fit depending on the observations $\mathbf{d}$ and the theoretical predictions $\mathbf{d}_c(\mathbf{m})$ for a model, and $\Psi(\mathbf{m})$ is a regularization term controlling the behaviour in model space which may, for example, depend on the size of model gradients. Such a form encompasses nearly all commonly employed schemes for linearized inversion (see e.g. Tarantola 1987). An implicit rescaling of the different model classes is normally included in the regularization term through a model covariance matrix C_m, for example

$$\Psi(\mathbf{m}) = (\mathbf{m} - \mathbf{m}_0)^T \mathsf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_0), \tag{2.2}$$

where $\mathbf{m}_0$ is a reference model. In eqs (2.1) and (2.2), the implementation of the data fit $\Phi(\mathbf{d}, \mathbf{d}_c(\mathbf{m}))$, and the regularization term $\Psi(\mathbf{m})$ will involve the full set of parameter classes $\{\mathbf{m}_i\}$, and the inverse model covariance matrix C_m^{-1} will need to be partitioned by parameter class (*cf*. Kennett *et al.* 1988).

We now expand $\mathscr{F}$ about a reference model $\{\mathbf{m}_0\}$. In order to display the dependence on the different parameter classes $\mathbf{m}_i$ we will use the convention of summation over repeated suffices:

$$\mathscr{F} = \mathscr{F}(\mathbf{m}_{0i}) + \nabla_i \mathscr{F}(\mathbf{m}_{0i}) \delta\mathbf{m}_i + \frac{1}{2} \delta\mathbf{m}_i^T \nabla_i \nabla_j \mathscr{F}(\mathbf{m}_{0i}) \delta\mathbf{m}_j + \dots \tag{2.3}$$

$$= \mathscr{F}(\mathbf{m}_{0i}) + \theta_i \delta\mathbf{m}_i + \frac{1}{2} \delta\mathbf{m}_i^T \mathsf{H}_{ij} \delta\mathbf{m}_j + \dots, \tag{2.4}$$

where $\delta\mathbf{m}_j = \mathbf{m}_j - \mathbf{m}_{0j}$ is the deviation from the reference model, and the gradient θ and the Hessian H have been partitioned in terms of the different parameter classes. We seek the set of perturbations $\delta\mathbf{m}_i$ such that $\mathscr{F}$ is minimized. With the quadratic approximation (2.4), this condition yields

$$\mathsf{H}_{ij} \delta\mathbf{m}_j = -\theta_i. \tag{2.5}$$

The system of equations (2.5) has to be solved to determine the set of parameters $\delta\mathbf{m}_j$. Commonly, the number of equations will be very large so that an iterative technique will be employed in the solution.

With multiple parameter classes, the problem can be recast in an iterative form in which generalized inverses are only required for the N diagonal blocks of the matrix H_{ij}. This iterative development has significant advantages for systems with a large number of parameters where direct solution is impractical.

We will illustrate the general approach by an explicit example with three parameter classes. This is sufficient to display the features of the algorithm without being algebraically unwieldy. Consider, then, the set of equations (2.5) when there are three distinct parameter classes whose deviations from the reference value are represented by $\delta\mathbf{m}_1, \delta\mathbf{m}_2, \delta\mathbf{m}_3$. In partitioned form, (2.5) becomes

$$\begin{pmatrix} \mathsf{H}_{11} & \mathsf{H}_{12} & \mathsf{H}_{13} \\ \mathsf{H}_{21} & \mathsf{H}_{22} & \mathsf{H}_{23} \\ \mathsf{H}_{31} & \mathsf{H}_{32} & \mathsf{H}_{33} \end{pmatrix} \begin{pmatrix} \delta\mathbf{m}_1 \\ \delta\mathbf{m}_2 \\ \delta\mathbf{m}_3 \end{pmatrix} = - \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix}. \tag{2.6}$$

If the problem were truly linear and the Hessian matrix H had full rank, then the solutions of (2.6) for the perturbations in the different parameter classes should be equivalent, whatever method might be used. However, for the usual situation for a system of less than full rank, where the constraints on different parameter classes can be quite different, the numerical results can vary significantly because of the 'implicit regularization' (or damping) imposed by the iterative equation solver method being employed (Sambridge 1990).

We introduce a procedure that allows each parameter class to be treated in the same way and which does not require a problem-specific pre-conditioning. We rewrite eqs (2.6) in a form that emphasizes the significance of the diagonal blocks:

$$\begin{aligned} \mathsf{H}_{11} \delta\mathbf{m}_1 &= -\theta_1 - \mathsf{H}_{12} \delta\mathbf{m}_2 - \mathsf{H}_{13} \delta\mathbf{m}_3, \\ \mathsf{H}_{22} \delta\mathbf{m}_2 &= -\theta_2 - \mathsf{H}_{21} \delta\mathbf{m}_1 - \mathsf{H}_{23} \delta\mathbf{m}_3, \\ \mathsf{H}_{33} \delta\mathbf{m}_3 &= -\theta_3 - \mathsf{H}_{31} \delta\mathbf{m}_1 - \mathsf{H}_{32} \delta\mathbf{m}_2. \end{aligned} \tag{2.7}$$

If the off-diagonal blocks are less significant than the diagonal, eqs (2.7) can be solved through an iterative scheme whereby the rth iterate for the $\delta\mathbf{m}_j^{(r)}$ is to be found by solving the system

$$\begin{aligned} \mathsf{H}_{11} \delta\mathbf{m}_1^{(r)} &= -\theta_1 - \mathsf{H}_{12} \delta\mathbf{m}_2^{(r-1)} - \mathsf{H}_{13} \delta\mathbf{m}_3^{(r-1)}, \\ \mathsf{H}_{22} \delta\mathbf{m}_2^{(r)} &= -\theta_2 - \mathsf{H}_{21} \delta\mathbf{m}_1^{(r-1)} - \mathsf{H}_{23} \delta\mathbf{m}_3^{(r-1)}, \\ \mathsf{H}_{33} \delta\mathbf{m}_3^{(r)} &= -\theta_3 - \mathsf{H}_{31} \delta\mathbf{m}_1^{(r-1)} - \mathsf{H}_{32} \delta\mathbf{m}_2^{(r-1)}, \end{aligned} \tag{2.8}$$

where the cross-coupling terms depend on the previous iterates $\delta\mathbf{m}_j^{(r-1)}$. The solution of these equations for the $\delta\mathbf{m}_j^{(r)}$ will require the formal construction of the generalized inverses $\mathsf{H}_{11}^-, \mathsf{H}_{22}^-, \mathsf{H}_{33}^-$, but we note that only the right-hand side of each equation changes with each iteration. Eqs (2.8) can be solved with any convenient linear-equation solver. The procedures used to solve the system (2.8) may themselves be iterative if the diagonal blocks H_{ii} are sparse, but the constant structure of the equation sets means that it is possible to make an efficient computational implementation of the main iteration.

The iterative scheme is initiated by solving independently for each parameter class:

$$\begin{aligned} \mathsf{H}_{11} \delta\mathbf{m}_1^{(0)} &= -\theta_1, \\ \mathsf{H}_{22} \delta\mathbf{m}_2^{(0)} &= -\theta_2, \\ \mathsf{H}_{33} \delta\mathbf{m}_3^{(0)} &= -\theta_3. \end{aligned} \tag{2.9}$$

The first iteration of (2.8) then sees the introduction of the cross-interaction between parameter classes. At the second iteration, the full interaction between the different parameter classes is in place, and subsequent iterations allow refinements of the estimates of $\delta\mathbf{m}_i^{(0)}$ without any need for nested generalized inverses, as would occur in a direct partitioned inverse applied to (2.6). For the case considered by Kennett *et al.* (1998) with two parameter classes, the convergence of the iterative scheme (2.8), (2.9) is rapid, and only five iterations were needed.

The way in which we have introduced the iterative scheme over the parameter classes demonstrates its plausibility, but does not ensure its equivalence to the solution of the original system (2.6). For two parameter classes, Kennett *et al.* (1998, Appendix B) have provided a demonstration of the derivation of the iterative procedure directly from the equivalent equations to (2.6); this demonstration can only be extended to higher-dimensional systems at the cost of considerable algebraic complexity.

For the full problem with N parameter classes, the iterative procedure is initiated by solution of the N sets of equations

$$\mathsf{H}_{ii}\delta\mathbf{m}_i^{(0)} = -\theta_i\,, \tag{2.10}$$

in terms of the block diagonal partitions of the Hessian matrix, which determine the direct dependence on a particular parameter class. The subsequent iteration takes the form

$$\mathsf{H}_{ii}\delta\mathbf{m}_i^{(r)} = -\theta_i - \sum_{i\neq j}^{N} \mathsf{H}_{ij}\delta\mathbf{m}_j^{(r-1)}, \quad r > 1\,, \tag{2.11}$$

reintroducing the cross-dependences between classes.

3 DISCUSSION

The iterative procedure for inversion with N parameter classes, introduced in the previous section, can be interpreted as a multi-stage approach which first treats the different parameter classes as independent and then subsequently develops increasingly refined estimates of the cross-interactions. The algorithm depends on a choice of subsystems such that the diagonal blocks of the Hessian matrix dominate. Where this property does not arise naturally, it may often be achieved using an appropriate formulation of the parameter system employed in the inversion, for example by taking a linear combination of the original variables.

A major advantage of the iterative approach proposed here is that each parameter class is treated in a comparable way. The inversions are only for the diagonal blocks of the partitioned Hessian matrix, with coupling of one parameter class to itself, and the off-diagonal blocks take care of inter-parameter scaling. In consequence, the parameter classes do not require arbitrary rescaling before the inversion can commence. The iterative procedure avoids problems inherent in a direct partitioned inverse (Spencer 1985) or in sequential inversions for different parameter classes (see e.g. Mao & Stuart 1997), which both require nested generalized inverses with a potential for build-up of numerical error.

This iterative approach to inversion for multiple parameter classes is of particular benefit for systems with large numbers of parameters in each class, but can also be useful for smaller systems to implement the equivalent of a partitioned inverse. The algorithm presented in (2.10) and (2.11) is the most direct implementation of an iterative development. A variety of other forms can be also introduced, in which, once the rth iterate has been constructed for the model vector component corresponding to a particular parameter class, it is available for use in the calculation of the rth iterate for the remaining classes. Where there is a clear hierarchy in the significance of the different parameter classes, the equations for the most significant parameter class should be solved first and then the other equations solved in order of importance. As soon as the rth iterate for the model vector for a parameter class is available, it would then be used to construct the cross-correction terms for that class used in the subsequent equations.

REFERENCES

Dziewonski, A.M. & Anderson, D.L., 1981. Preliminary reference Earth model, *Phys. Earth planet. Inter.,* **25,** 297–356.

Kennett, B.L.N., Sambridge, M.S. & Williamson, P.R., 1987. Subspace methods for inverse problems with multiple parameter classes, *Geophys. J. Int.,* **94,** 237–247.

Kennett, B.L.N., Widiyantoro, S. & van der Hilst, R.D., 1998. Joint seismic tomography for bulk-sound and shear wavespeed in the Earth's mantle, *J. geophys. Res.,* **103,** 12 469–12 493.

Mao, W.J. & Stuart, G.W., 1996. Transmission-reflection tomography: application to reverse VSP data, *Geophysics,* **62,** 884–894.

Montagner, J.-P. & Kennett, B.L.N., 1996. How to reconcile body-wave and normal-mode reference Earth models?, *Geophys. J. Int.,* **125,** 229–248.

Oldenburg, D.W., McGillivray, P.R. & Ellis, R.G., 1993. General subspace methods for large-scale inverse problems, *Geophys. J. Int.,* **114,** 12–20.

Sambridge, M.S., 1990. Nonlinear arrival time inversion: constraining velocity anomalies by seeking smooth models, *Geophys. J. Int.,* **102,** 653–677.

Spencer, C., 1985. The use of partitioned matrices in geophysical inversion problems, *Geophys. J. R. astr. Soc.,* **80,** 619–629.

Su, W.-J. & Dziewonski, A.M., 1997. Simultaneous inversion for 3-D variations in shear and bulk velocity in the mantle, *Phys. Earth planet. Inter.,* **100,** 135–156.

Tarantola, A., 1987. Inverse Problem Theory: Methods for Data Fitting and model parameter estimation, Elsevier, Amsterdam.