Geophys. J. Int. (1993) **114,** 12–20

Generalized subspace methods for large-scale inverse problems

D. W. Oldenburg, P. R. McGillivray and R. G. Ellis

UBC-Geophysical Inversion Facility, Department of Geophysics and Astronomy, University of British Columbia, Vancouver, BC V6T 1Z4, *Canada*

Accepted 1992 October 24. Received 1992 April 10

SUMMARY

Numerical efficiency and efficacy of subspace methods for solving large-scale geophysical inverse problems are investigated. The primary advantage of subspace techniques over traditional Gauss–Newton algorithms lies in the need to invert only a matrix equal to the dimension of the subspace. The efficacy of the method lies in a judicious choice of basis vectors. Vectors associated with gradients of the data misfit or gradients of the model component of the objective function are of great utility, but substantial improvement in convergence rates can be obtained by using basis vectors associated with gradients of a segmented objective function. To quantify these benefits we invert data acquired in a synthetic dc resistivity experiment. 420 electric potentials obtained at the surface of a 2-D earth are inverted to recover estimates of the electrical conductivity of 1296 cells. The number of basis vectors range from two to 95 and convergence rates, model norms and final models are compared. In an effort to reduce the computations we investigate the possibility of using only linear information in the data-misfit objective function. This is shown to be effective at early iterations and is computationally efficient since it obviates the need to calculate curvature information in the data misfit and because it can also be implemented without a line search. The effects of using gradient vectors versus steepest descent vectors in the inversion are examined. Accordingly we introduce two methods by which approximate descent vectors can be fabricated from gradient vectors. They show that even simple preconditioning of gradient vectors can dramatically improve convergence rates provided that all vectors are preconditioned in the same manner.

Key words: dc resistivity, inverse theory, optimization, subspace.

1 INTRODUCTION

In a typical geophysical inverse problem we are supplied with observations $\mathbf{d}_0 = (d_{01}, d_{02}, \ldots, d_{0N})$ and their estimated uncertainties, and we desire to find the function m which gave rise to those observations. We refer to m as the 'model' and the relationship between the jth datum and the model is $d_{0j} = g_j(m)$. The functionals g_j are assumed to be known. In order to carry out forward modelling to generate theoretical responses, and also to attack the inverse problem, we are generally forced to parameterize m as

$$m = \sum_{k=1}^{M} m_k \psi_k \tag{1}$$

where ψ_k are basis functions for the functional model space and m_k are coefficients. The parameter vector $\mathbf{m} = \{m_1, m_2, \ldots, m_M\}$ is also referred to as the 'model' and the inverse problem is reduced to finding a vector $\mathbf{m} \in R^M$ which adequately reproduces the observations.

Realistic data are inaccurate and the constructed model should not reproduce the observations to greater or lesser fidelity than is justified by the data uncertainties. Over fitting the data generally manifests itself as additional structure on the constructed model. Such structure is not physically interpretable; it is merely an artefact of the noise. Alternatively, by under fitting the data, information about the model that is contained in the data may be lost. Numerous possibilities for characterizing the misfit exist but here we choose

$$\phi_d(\mathbf{d}, \mathbf{d}_0) = \|\mathbf{W_d}(\mathbf{d} - \mathbf{d}_0)\|^2 \tag{2}$$

where $\mathbf{W_d}$ is an $N \times N$ matrix. If noise contaminating the jth observation is an uncorrelated Gaussian random variable having zero mean and standard deviation σ_j then an

appropriate form for $\mathbf{W_d}$ is $\mathbf{W_d} = \text{diag}\{1/\sigma_1, \ldots, 1/\sigma_N\}$. With this assumption, ϕ_d is the random variable distributed as chi squared with N degrees of freedom. The expected value of ϕ_d is therefore approximately equal to N and accordingly, the model sought from the inversion algorithm should reproduce the observations to about this value.

The inverse problem is non-unique in that if there is one model which adequately reproduces the observations, then there are (generally) infinitely many other models which will be equally acceptable. To deal with this non-uniqueness we design our inversion algorithm to compute a specific model; of all acceptable models, we desire to find that one which minimizes a global objective function of the model. Here we minimize

$$\phi_m(\mathbf{m}, \mathbf{m}_0) = \|\mathbf{W_m}(\mathbf{m} - \mathbf{m}_0)\|^2 . \tag{3}$$

In eq. (3) $\mathbf{m}_0$ is a base model and $\mathbf{W_m}$ is a general weighting matrix which is designed so that a model with specific characteristics is produced. The minimization of ϕ_m yields a model that is close to $\mathbf{m}_0$ with the metric defined by $\mathbf{W_m}$ and so the characteristics of the recovered model are directly controlled by these two quantities.

The inverse problem becomes:

$$\begin{aligned} &\text{minimize} \quad \phi_m(\mathbf{m}, \mathbf{m}_0) \\ &\text{subject to} \quad \phi_d(\mathbf{d}, \mathbf{d}_0) = \phi_d^* \end{aligned} \tag{4}$$

where ϕ_d^* is a desired target misfit. The appropriate objective function to be minimized is

$$\phi(\mathbf{m}) = \phi_m(\mathbf{m}, \mathbf{m}_0) + \mu(\phi_d(\mathbf{d}, \mathbf{d}_0) - \phi_d^*) \tag{5}$$

where μ is a Lagrange multipler. The inverse problem is non-linear and is attacked by linearizing (5) about the current model $\mathbf{m}^{(n)}$ and iterating. If $\boldsymbol{\delta}\mathbf{m}$ is a model perturbation, then a Taylor expansion which has terms only up to second order is

$$\begin{aligned} \phi[\mathbf{m}^{(n)} + \boldsymbol{\delta}\mathbf{m}] = \phi_m + \boldsymbol{\gamma}_\mathbf{m}^T \boldsymbol{\delta}\mathbf{m} + \tfrac{1}{2}\boldsymbol{\delta}\mathbf{m}^T \mathbf{H_m}\, \boldsymbol{\delta}\mathbf{m} \\ + \mu(\phi_d + \boldsymbol{\gamma}_\mathbf{d}^T \boldsymbol{\delta}\mathbf{m} + \tfrac{1}{2}\boldsymbol{\delta}\mathbf{m}^T \mathbf{H_d}\, \boldsymbol{\delta}\mathbf{m} - \phi_d^*) \end{aligned} \tag{6}$$

where $\boldsymbol{\gamma}_\mathbf{m} = \nabla_\mathbf{m}\phi_\mathbf{m}$ and $\boldsymbol{\gamma}_d = \nabla_\mathbf{m}\phi_d$ are gradient vectors, $\mathbf{H_m} = \nabla_\mathbf{m}\nabla_\mathbf{m}\phi_m$ and $\mathbf{H_d} = \nabla_\mathbf{m}\nabla_\mathbf{m}\phi_d$ are Hessian matrices and $\nabla_\mathbf{m}$ is the operator $(\partial/\partial m_1, \partial/\partial m_2, \ldots, \partial/\partial m_M)^T$. In eq. (6) ϕ_m is understood to be $\phi_m[\mathbf{m}^{(n)}, \mathbf{m}_0]$ and ϕ_d is $\phi_d[\mathbf{d}^{(n)}, \mathbf{d}_0]$.

The general solution proceeds by differentiating eq. (6) with respect to $\boldsymbol{\delta}\mathbf{m}$ and μ to obtain an $M \times M$ system of equations to be solved for $\boldsymbol{\delta}m$ and a constraint equation used to evaluate the misfit and hence adjust the value of μ. The computational difficulties lie in the solution of the matrix system when M becomes large. There are two general avenues of attack. Iterative techniques (e.g. using conjugate gradient directions, Hestenes 1980) may produce approximate solutions to the full system of equations after a number of iterations which is considerably less than M. This can be computationally feasible. Alternatively, explicit subspace approaches may be used to generate precise solutions to a restricted problem. In a subspace approach, the 'model' perturbation $\boldsymbol{\delta}\mathbf{m} \in R^M$ is restricted to lie in a q-dimensional subspace of R^M which is spanned by the vectors $\{\mathbf{v}_i\}$ $i = 1, q$. The model perturbation can be written as

$$\boldsymbol{\delta}\mathbf{m} = \sum_{i=1}^{q} \alpha_i \mathbf{v}_i \equiv \mathbf{V}\boldsymbol{\alpha} \tag{7}$$

and is therefore specified once the parameters α_i are determined. The principal advantage of this approach is that only a $q \times q$ matrix needs to be inverted. An immediate disadvantage is that in restricting the activated portion of model space, it may be that vectors which are important in finding the global minimum of the desired objective function are not available and an inferior solution is obtained. Despite this potential drawback, the desire to solve inverse problems without having to invert large matrices is so appealing that we focus upon this method. Some of our ideas are not new. The basic philosophy of using subspace inversion is common place in the mathematical literature. In fact many of the earliest inversions resorted to representing the model as a linear combination of chosen basis elements (often sinusoids) and finding coefficients of these vectors which produced acceptable agreement to the observations. More recently the computational efficiency of the subspace method and its application to large-scale inverse problems has been presented by Skilling & Bryan (1984) who maximized an entropy norm suitable for image reconstruction, and by Kennett & Williamson (1988) in their inversion of seismic data. Our approach builds upon ideas in those papers but we make improvements in a number of areas.

The success or failure of a subspace approach hinges upon the selection of the spanning vectors for the activated subspace. Much of our attention is focused upon questions of 'which type' and 'how many'? At each iteration in an inversion we attempt to select those vectors which are most intimately related to the changes in the objective function. Specifically, we investigate the effectiveness of dividing the misfit and model components of the objective function into parts and using steepest-descent vectors associated with each component. We show via example the benefits of using steepest-descent vectors as compared to gradient vectors and then present two methods by which approximate steepest-descent vectors can be generated.

Successful implementation of a subspace method also requires that the computations be carried out in reasonable time. The numerical difficulties encountered in solving an inverse problem may manifest themselves in solving the forward problem, computing sensitivities or computing the Hessian matrix, or in solving a matrix equation. When solving the linearized equations we need to estimate a value of μ (in eq. 5) which provides a desired reduction in the data misfit or keeps the misfit at a predetermined value. This requires a line search which can be computationally expensive. For other inverse problems the computational difficulties lies in obtaining sensitivities or Hessian information. There are some problems in which the sensitivity is available as a byproduct of the forward modelling (e.g. seismic traveltime tomography problems) but usually considerable effort is required. When sensitivity or Hessian information is the computational impediment, we want to abandon (if possible) the need to have such detailed information about the curvature in ϕ_d. We show that this may be feasible at early iterations.

To make quantitative comparisons for convergence rates and achieved models using different basis vectors and to examine the neglect of curvature information we invert a single synthetic data set simulating a dc resistivity survey over a 2-D conductive structure. Although some of the conclusions will be dependent upon the explicits of this problem, it is hoped that the insight achieved regarding the

subspace approach will be applicable to other geophysical inversions.

2 SUBSPACE EQUATIONS

The equations for the subspace formulation are generated by substituting (7) into (6) to yield

$$\phi[\mathbf{m}^{(n)} + \mathbf{V}\boldsymbol{\alpha}] = \phi_m + \boldsymbol{\gamma}_\mathbf{m}^T\mathbf{V}\boldsymbol{\alpha} + \tfrac{1}{2}\boldsymbol{\alpha}^T\mathbf{V}^T\mathbf{H}_\mathbf{m}\mathbf{V}\boldsymbol{\alpha} + \mu(\phi_d + \boldsymbol{\gamma}_\mathbf{d}^T\mathbf{V}\boldsymbol{\alpha} + \tfrac{1}{2}\boldsymbol{\alpha}^T\mathbf{V}^T\mathbf{H}_\mathbf{d}\mathbf{V}\boldsymbol{\alpha} - \phi_d^*). \quad (8)$$

This is a quadratic objective function to be solved for the parameter vector α. Differentiating (8) with respect to α and μ and setting the resultant equations equal to zero yields

$$\mathbf{V}^T(\mathbf{H}_\mathbf{m} + \mu\mathbf{H}_\mathbf{d})\mathbf{V}\boldsymbol{\alpha} = -\mu\mathbf{V}^T\boldsymbol{\gamma}_\mathbf{d} - \mathbf{V}^T\boldsymbol{\gamma}_\mathbf{m} \quad (9a)$$

$$\boldsymbol{\gamma}_\mathbf{d}^T\mathbf{V}^T\boldsymbol{\alpha} + \tfrac{1}{2}\boldsymbol{\alpha}^T\mathbf{V}^T\mathbf{H}_\mathbf{d}\mathbf{V}\boldsymbol{\alpha} = \phi_d^* - \phi_d. \quad (9b)$$

The solution of these equations requires that $\mathbf{V}^T(\mathbf{H}_\mathbf{m} + \mu\mathbf{H}_\mathbf{d})\mathbf{V}$ be inverted and the numerical efficiency of the inversion is therefore realized since this is a $q \times q$ matrix. Computationally we note $\mathbf{H}_\mathbf{m} = 2\mathbf{W}_\mathbf{m}^T\mathbf{W}_\mathbf{m}$ is specified at the outset and is available. However, $\mathbf{H}_\mathbf{d} = 2\mathbf{J}^T\mathbf{W}_\mathbf{d}^T\mathbf{W}_\mathbf{d}\mathbf{J} + \mathbf{Q}$ where $J_{ij} = \partial g_i[m^{(n)}]/\partial m_j$ is the $N \times M$ Jacobian or sensitivity matrix and

$$Q_{ij} = \sum_l (d_l - d_{0l}) \frac{\partial^2 d_l}{\partial m_i\, \partial m_j}. \quad (10)$$

The matrix $\mathbf{Q}$ involves second derivatives of the individual data functionals with respect to the model parameters but, because each element is multiplied by a data misfit, its size diminishes as convergence to the inverse problem is achieved. Neglecting $\mathbf{Q}$, which we do here, yields the Gauss–Newton solution. Then $\mathbf{V}^T\mathbf{H}_\mathbf{d}\mathbf{V} = 2\mathbf{V}^T\mathbf{J}^T\mathbf{W}_\mathbf{d}^T\mathbf{W}_\mathbf{d}\mathbf{J}\mathbf{V} = 2(\mathbf{W}_\mathbf{d}\mathbf{J}\mathbf{V})^T(\mathbf{W}_\mathbf{d}\mathbf{J}\mathbf{V})$ can be computed by calculating $\mathbf{J}$ or, more efficiently, computing individually the q vectors in the matrix product $\mathbf{W}_\mathbf{d}\mathbf{J}\mathbf{V}$. The latter computation is very efficient if only a few vectors are needed.

In much of this paper we will demand that a desired target misfit is achieved at each iteration. As a consequence, the linearized misfit eq. 9(b) is not adequate and a line search involving forward modelling will be carried out to find the value of μ. This involves an initial guess for μ, solving eq. 9(a) by SVD for the vector $\boldsymbol{\alpha}$, computing the perturbation $\boldsymbol{\delta}_\mathbf{m}$, carrying out a forward modelling to evaluate the true responses and misfit, and then adjusting μ.

3 CHOICE OF BASIS FUNCTIONS

At each iteration we desire a model perturbation which minimizes ϕ_m and alters ϕ_d so that it achieves a specific decrease or stays at the target value ϕ_d^*. Vectors that are intimately related to changes in ϕ_m or ϕ_d are potentially valuable. Accordingly, our first choices for basis vectors are obtained by partitioning these objective function components and forming gradients.

The misfit objective function $\phi_d = (\mathbf{d} - \mathbf{d}_0)^T\mathbf{W}_\mathbf{d}^T\mathbf{W}_\mathbf{d}(\mathbf{d} - \mathbf{d}_0)$ may be partitioned as

$$\phi_d = \sum \phi_d^k \quad (11)$$

where the kth subset is $\phi_d^k = (\mathbf{d}^k - \mathbf{d}_0^k)^T\mathbf{W}_\mathbf{d}^T\mathbf{W}_\mathbf{d}(\mathbf{d}^k - \mathbf{d}_0^k)$. The gradient $\boldsymbol{\gamma}_k = \nabla_\mathbf{m}\phi_d^k$ can be calculated. Our goal is to minimize ϕ_m and according to (3) our norm on model space is controlled by the symmetric positive-definite matrix $\mathbf{W}_\mathbf{m}^T\mathbf{W}_\mathbf{m}$. A steepest ascent direction can be obtained by multiplying the gradient by $(\mathbf{W}_\mathbf{m}^T\mathbf{W}_\mathbf{m})^{-1}$ (Gill, Murray & Wright 1981, p 102). We therefore choose vectors

$$\mathbf{v}_k = (\mathbf{W}_\mathbf{m}^T\mathbf{W}_\mathbf{m})^{-1}\nabla_\mathbf{m}\phi^k \quad (12)$$

as elements in our subspace.

The method of partitioning ϕ_d is no doubt problem dependent. For many geophysical problems the data to be inverted are a composite of results from many physical sources. A typical seismic problem requires a simultaneous inversion of data acquired from individual shots; in an E-SCAN dc resistivity survey, the potentials arising from different current sources are to be inverted. Given that there can be particular difficulties or characteristics associated with each source, an objective function corresponding to data arising from each source function is a logical choice. This has a practical side benefit in that the gradient vector can generally be computed by a single solution of the adjoint equation driven by a distribution of sources located at the measuring sites and scaled by the data misfit. The data misfit objective function can also be partitioned on a physical basis. For instance, geophysical data can often be grouped in terms of distance between the source and receiver. Near-offset data are often most influenced by near-surface structure while far-offset data sample deeper portions of the earth. When this physical reasoning is valid, a partitioning based upon grouping data which are primarily sensitive to certain portions of the earth may be useful.

The partitioning of the data-misfit objective function can be carried to the extreme so that a gradient vector is obtained for each datum. In this case $\phi_d^k = W_{kk}^2(d_k - d_{0k})^2$ where W_{kk} is the kth diagonal element of the matrix $\mathbf{W}_\mathbf{d}$. The gradient vector for ϕ_d^k is $\nabla_\mathbf{m}\phi_d^k = 2W_{kk}^2(d_k - d_{0k})\nabla_\mathbf{m}d_k$. This is simply a scaled version of the sensitivity vector $\nabla_\mathbf{m}d_k$ for the kth datum. If only these vectors are used in the inversion then the activated portion of model space is limited to that spanned by the rows of the sensitivity matrix. This is an implementation that is often used when faced with solving an underdetermined inverse problem and it requires the solution of an $N \times N$ matrix.

The remaining vectors to be chosen should be sensitive to ϕ_m. A vector which generally should be included is

$$\mathbf{v} = (\mathbf{W}_\mathbf{m}^T\mathbf{W}_\mathbf{m})^{-1}\boldsymbol{\gamma}_\mathbf{m} = [\mathbf{m}^{(n)} - \mathbf{m}_0]. \quad (13)$$

This vector can be used as is or it can be partitioned so that basis vectors which concentrate only upon certain portions of the model can be explicitly introduced. This allows the user to focus upon those areas of the model where specific improvements are desired. Additional vectors obtained by smoothing or roughening $\boldsymbol{\gamma}_\mathbf{m}$, perturbation vectors from previous iterations, and vectors such as constants or linear ramps can also be useful choices.

For many geophysical problems the model is cellularized so that the model value is constant within each cell. Effectively the initial basis function ψ_k in eq. (1) is a pulse of unit amplitude on the kth cell and zero everywhere else; the parameter m_k is then equal to the value of the model on the kth cell. The partitioning of the model vector in eq. (13) can

also be carried to the extreme so that each descent vector refers to a single cell. Thus

$$\mathbf{v}_k = [\mathbf{m}_k^{(n)} - \mathbf{m}_{0k}]. \tag{14}$$

In this limit the M vectors $\mathbf{v}_k$ completely span model space and no further basis vectors are required. This 'subspace' solution reverts to the traditional Gauss–Newton solution and the inversion of an $M \times M$ matrix is demanded.

Having selected the basis vectors, there is one further modification that is required. The subspace formulation demands the inversion of the matrix $\mathbf{V}^T(\mathbf{H_m} + \mu\mathbf{H_d})\mathbf{V}$. This matrix is singular if the column vectors of $\mathbf{V}$ are linearly dependent. As a last step then, we guard against poor conditioning by orthonormalizing the descent vectors prior to using them in the subspace equations.

4 SYNTHETIC MODEL EXAMPLE

To illustrate the effects of choice of basis vectors we invert surface electric potentials obtained from a pole–pole dc resistivity survey carried out over a 2-D conductivity structure shown in Fig. 1(a). Surface electrodes are located every 10 m in the interval $x = (-100, 100)$ m. Each of the 21 electrode positions can be activated as a current site and when it is, electric potentials are recorded at the remaining electrodes. The observed data set consists of 420 potential values each of which has been contaminated by Gaussian noise having a standard deviation equal to 5 per cent of the true potential. The data are generated using a finite difference code (McGillivray 1992) and the mesh, used both for forward modelling and for the inversion, is given in Fig. 1(b). There are 1296 mesh elements. The apparent conductivity pseudo-section, shown in Fig. 1(c), provides a direct indication of the lateral change in surface conductivity but only vaguely indicates the presence of a conducting body in the centre of the model. The conductive tail extending to the left at the surface in Fig. 1(c) is due primarily to the change in surface conductivity.

To evaluate the effect that selection of different basis vectors has on the inversion, we have adopted the following strategy. Each inversion begins with a half-space of conductivity 0.005 Sm^{-1}. At every iteration we ask for a 50 per cent decrease in the misfit objective function until the target misfit $\phi_d = N$ is achieved. A line search using forward modelling ensures that this is achieved, or, in cases where it is not achievable, the line search is used to find that value of μ which provides the greatest decrease in the misfit. Once the target misfit has been obtained, the line search ensures that the misfit remains at the target value and hence subsequent iterations alter only ϕ_m.

Our choice for ϕ_m is guided by a desire to find a model which has minimum structure in the vertical and horizontal directions and at the same time is close to a base model m_0. To accomplish this we minimize a discretized approximation to

$$\phi_m(m, m_0) = \alpha_s \int\int (m - m_0)^2\, dx\, dz + \int\int \left\{ \alpha_x \left[\frac{\partial(m - m_0)}{\partial x} \right]^2 + \alpha_z \left[\frac{\partial(m - m_0)}{\partial z} \right]^2 \right\} dx\, dz \tag{15}$$

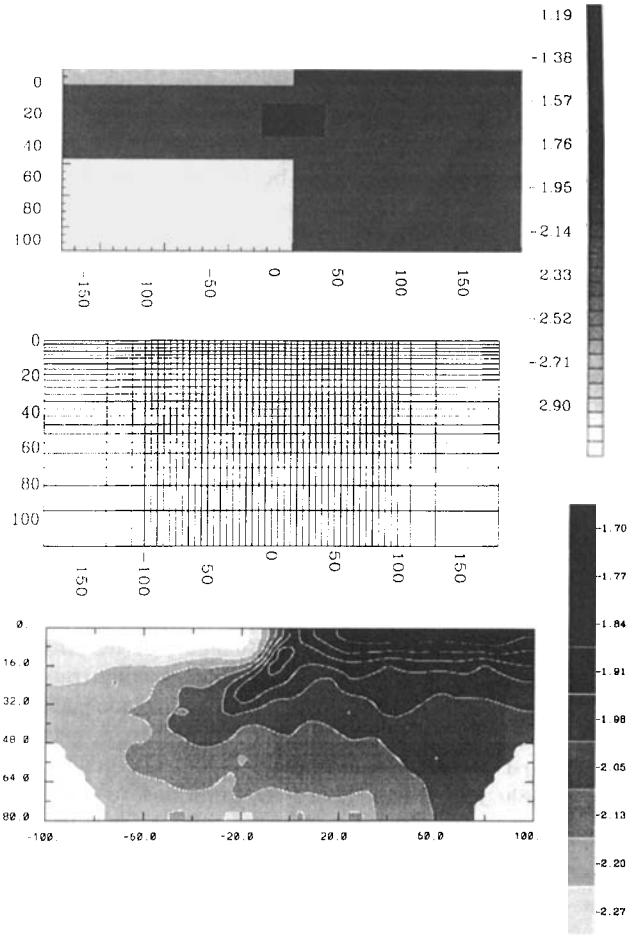

Figure 1. The conductivity model is shown at the top. Surface electrodes are spaced 10 m apart over the interval $x = (-100, 100)$ m. Current is input at each electrode site in turn, and potentials are observed at the remaining 20 sites. The finite difference grid for forward modelling and for inversion is shown in the middle. An additional set of border cells (not shown) completes the grid. The pole–pole potentials V are converted to apparent conductivity $\sigma_a = 1/(2\pi rV)$, where r denotes the distance between current and potential electrodes; these are plotted at the bottom. The apparent conductivity value is plotted midway between the current and potential electrodes at a (pseudo) depth of $z = 0.86r$.

where α_s, α_x, and α_z are adjustable constants. The discrete form of (15) is

$$\begin{aligned}\phi_m &\equiv \phi_s + \phi_x + \phi_z \\ &= \alpha_z \|\mathbf{W}_z(\mathbf{m} - \mathbf{m}_0)\|^2 + \alpha_x \|\mathbf{W}_x(\mathbf{m} - \mathbf{m}_0)\|^2 \\ &\quad + \alpha_z \|\mathbf{W}_z(\mathbf{m} - \mathbf{m}_0)\|^2 \\ &= (\mathbf{m} - \mathbf{m}_0)^T \\ &\quad \times \{\alpha_s \mathbf{W}_s^T\mathbf{W}_s + \alpha_x \mathbf{W}_x^T\mathbf{W}_x + \alpha_z \mathbf{W}_z^T\mathbf{W}_z\}(\mathbf{m} - \mathbf{m}_0) \\ &\equiv (\mathbf{m} - \mathbf{m}_0)^T \mathbf{W}_\mathbf{m}^T \mathbf{W}_\mathbf{m}(\mathbf{m} - \mathbf{m}_0).\end{aligned} \tag{16}$$

In (16) $\mathbf{W}_s$ is a diagonal matrix with elements $\sqrt{\Delta x\, \Delta z}$ where Δx is the length of the cell and Δz is its thickness, W_x has elements $\pm\sqrt{\Delta z/dx}$ where dx is the distance between the centres of horizontally adjacent cells, and W_z has elements $\pm\sqrt{\Delta x/dz}$ where dz is the distance between the centres of

vertically adjacent cells. We have chosen $\alpha_s = 0.0002$, $\alpha_x = 1.0$, $\alpha_z = 1.0$ and $m_0 = 0.005\, Sm^{-1}$. When the descretized matrices are applied to the synthetic model, $\phi_m = 1154$ (with $\phi_s = 51$, $\phi_x = 311$, and $\phi_z = 792$).

5 INVERSIONS I1–I4

The dc resistivity data will now be inverted four times in inversions labelled I1–I4. The four inversions will successively use 2, 3, 23 and 95 basis vectors. The progression of increased basis vectors is such that subsequent inversions contain the basis vectors used in previous inversions. This aids in quantifying the benefits of incorporating additional basis vectors.

In I1 we use the minimum number of basis vectors possible for our formulation. The two vectors selected are descent vectors associated with the gradient of the total misfit and model norm components of the objective function. That is,

$$\begin{aligned} \mathbf{v}_1 &= (\mathbf{W}_\mathbf{m}^T \mathbf{W}_\mathbf{m})^{-1} \nabla_\mathbf{m} \phi_d, \\ \mathbf{v}_2 &= (\mathbf{W}_\mathbf{m}^T \mathbf{W}_\mathbf{m})^{-1} \nabla_\mathbf{m} \phi_m. \end{aligned} \tag{17}$$

The misfit and model norm as a function of iteration are presented in Figs 2(a) and (b). The starting misfit is 1.08×10^5. The two vectors in eq. (17) provide sufficient flexibility for ϕ_d to be successively reduced by 50 per cent for the first six iterations. Beyond that convergence slows considerably and it has taken 89 iterations to reach the target value of 420. At iteration 100 $\phi_m = 41.9$ (with $\phi_s = 6.9$, $\phi_x = 9.9$, and $\phi_z = 25.0$). This is considerably less than ϕ_m for the true model. The model at iteration 100 is shown in Fig. 3(a). The discontinuous surface layer is well reproduced and the conductive prism is very visible even though its conductivity is lower than that of the true model. There is no indication of the resistive ledge at the bottom left of the image.

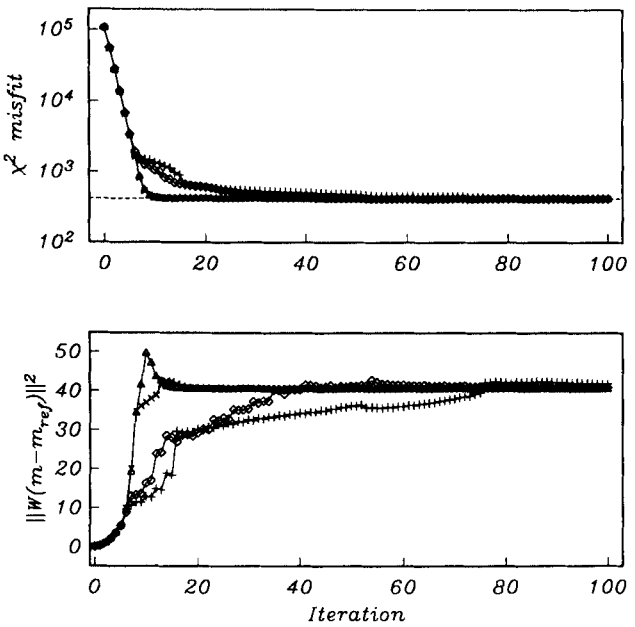

Figure 2. Convergence curves for inversions I1–I4 are denoted respectively by $(+, \diamond, \times, \triangle)$. In I1 steepest descent vectors associated with $\nabla_m \phi_m$ and $\nabla_m \phi_d$ are used; in I2 a third, constant, vector is incorporated; in I3 the misfit objective function is subdivided so that 23 basis vectors are used; and in I4 an additional 72 block-basis vectors are added so that 95 basis vectors used. The target misfit of 420 was reached at iteration 89, 55, 13 and 10 for the respective inversions and the model norm achieved at iteration 100 was respectively 41.9, 40.9, 40.6 and 40.5.

Inversion I2 is identical to I1 except that an additional, constant, basis vector has been incorporated. The convergence plots are presented in Fig. 2. A comparison with those for I1 shows that addition of this single basis function has been beneficial. Convergence is smoother and somewhat more rapid. The target misfit was achieved at iteration 55 and the model at iteration 100 has $\phi_m = 40.9$ (with $\phi_s = 6.6$, $\phi_x = 9.7$, and $\phi_z = 24.5$). The model is virtually indistinguishable from that in Fig. 3(a) and so is not presented.

In I3 we subdivide ϕ_d into data sets associated with each current electrode. The basis vectors are

$$\begin{aligned} \mathbf{v}_k &= (\mathbf{W}_\mathbf{m}^T \mathbf{W}_\mathbf{m})^{-1} \nabla_\mathbf{m} \phi_d^k \qquad k = 1, 21 \\ \mathbf{v}_{22} &= (\mathbf{W}_\mathbf{m}^T \mathbf{W}_\mathbf{m})^{-1} \nabla_\mathbf{m} \phi_m \\ \mathbf{v}_{23} &= \text{constant}. \end{aligned} \tag{18}$$

The increase in convergence rate is dramatic. Fig. 2 shows that the desired misfit is achieved by iteration 13. Further iterations keep the misfit at that level and work to reduce the model norm. At iteration 20, $\phi_m = 40.6$ (with $\phi_s = 6.5$, $\phi_x = 9.7$ and $\phi_z = 24.2$). The model at this iteration is given in Fig. 3(b). The features of this model are not substantially different from those in Fig. 3(a) although the conductive prism is slightly more compact and lateral variation in the surface layer is slightly decreased.

In I4 we subdivide the model norm objective function. Effectively, this amounts to introducing extra basis vectors associated with groups of cells. The additional cells, 72 in total, are delineated in Fig. 4. The inversion was performed with $\mathbf{v}_1$ to $\mathbf{v}_{23}$ as in I3 and the extra 72 basis vectors. The convergence path, delineated in Fig. 2, is not discernably different from that in I3 but the convergence was slightly faster; the desired misfit was achieved at iteration 10. After 20 iterations, the final model norm $\phi_m = 40.5$ (with $\phi_s = 6.5$, $\phi_x = 9.9$ and $\phi_z = 24.1$). This is nearly identical to that in I3 and so was the final model. It would seem that with imposed

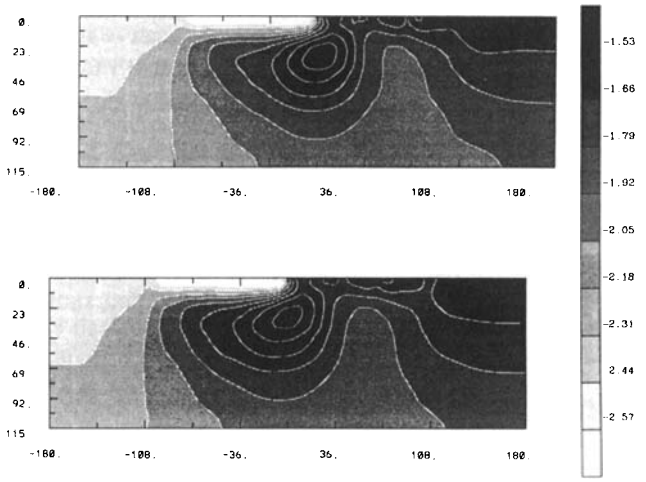

Figure 3. The constructed conductivity model obtained after 100 iterations for inversion I1 is shown in (a). The recovered model from I3 obtained after 20 iterations is shown in (b).

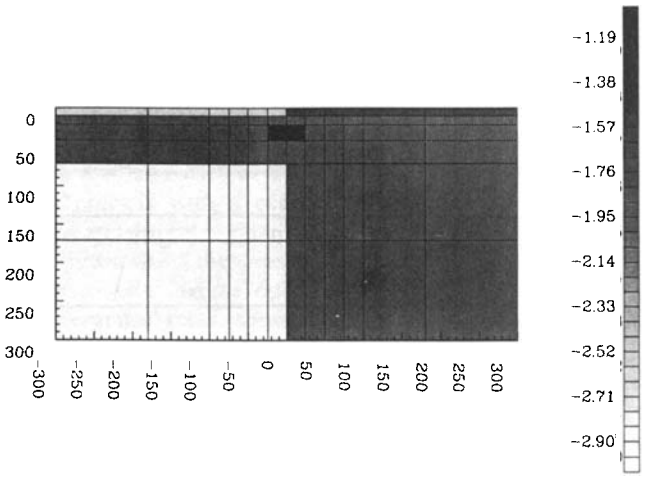

Figure 4. Each of the 72 cells is used as an additional basis vector in inversion I4.

constraints of reducing the misfit by a factor of two at each iteration there is little to be gained by increasing the number of basis vectors for this example. Moreover, the lack of difference between the models produced by using 23 and 95 basis vectors suggests that carrying out a full Gauss–Newton solution where each cell is a basis vector for the inversion would not have produced a substantially different model. Yet the computations to do this, which would have demanded the inversion of many matrices of size 1296×1296, is vastly greater than that carried out here.

6 NEGLECT OF CURVATURE INFORMATION IN ϕ_d AND INVERSIONS I5–I6

Because of the quadratic term in eq. (9) it is not possible to solve for μ directly. An iterative procedure which requires a number of matrix inversions must be used. This is not difficult if q is relatively small. Although estimates of μ based upon a linearized analysis are often adequate at initial iterations, and at final iterations if the model perturbations become sufficiently small, a line search is usually demanded at intermediate iterations. To carry out a line search, eqs (9) are solved with trial values of μ, and ϕ_d is evaluated after performing a forward modelling. A value of μ is selected which yields the desired misfit or results in the maximum possible decrease in the misfit-objective function.

The difficulties in computing $\mathbf{H_d}$ and μ, and also the fact that in the initial stages of an inversion the inclusion of curvature information may not be essential, motivates the development of a solution which omits the quadratic term in the misfit objective function. Neglect of $\mathbf{H_d}$ produces the resultant system of equations

$$\begin{aligned} \mathbf{V}^T\mathbf{H_m}\mathbf{V}\boldsymbol{\alpha} &= -\mu\mathbf{V}^T\boldsymbol{\gamma}_\mathbf{d} - \mathbf{V}^T\boldsymbol{\gamma}_\mathbf{m} \\ \boldsymbol{\gamma}_\mathbf{d}^T\mathbf{V}\boldsymbol{\alpha} &= \phi_d^* - \phi_d \end{aligned} \tag{19}$$

which has a closed form solution

$$\begin{aligned} \boldsymbol{\alpha} &= -(\mathbf{V}^T\mathbf{H_m}\mathbf{V})^{-1}\{\mu\mathbf{V}^T\boldsymbol{\gamma}_\mathbf{d} + \mathbf{V}^T\boldsymbol{\gamma}_\mathbf{m}\} \\ \mu &= \frac{\phi_d - \phi_d^* - \boldsymbol{\gamma}_d\mathbf{V}(\mathbf{V}^T\mathbf{H_m}\mathbf{V})^{-1}\mathbf{V}^T\boldsymbol{\gamma}_\mathbf{m}}{\boldsymbol{\gamma}_\mathbf{d}^T\mathbf{V}(\mathbf{V}^T\mathbf{H_m}\mathbf{V})^{-1}\mathbf{V}^T\boldsymbol{\gamma}_\mathbf{d}}. \end{aligned} \tag{20}$$

Equation (19) can be used to compute a model perturbation under the assumption that curvature information in the misfit-objective function can be neglected. This greatly reduces the computations but the perturbation can be quite incorrect when this assumption is no longer valid. To see this more explicitly, it is convenient to write an arbitrary perturbation as $\boldsymbol{\delta}\mathbf{m} = \boldsymbol{\delta}\mathbf{m}^{\parallel} + \boldsymbol{\delta}\mathbf{m}^{\perp}$ where parallel and perpendicular refer to projections of $\boldsymbol{\delta}\mathbf{m}$ onto the gradient of the misfit objective function $\boldsymbol{\gamma}_\mathbf{d}$. We note that any amount of $\boldsymbol{\delta}\mathbf{m}^{\perp}$ can be added without affecting the linear misfit in eq. (19). The basis vectors for inversions I1–I3 were of the form

$$\boldsymbol{\delta}\mathbf{m} = \sum \alpha_i\mathbf{v}_i + \beta[\mathbf{m}^{(n)} - \mathbf{m}_0] \equiv \boldsymbol{\delta}\mathbf{m}_1 + \boldsymbol{\delta}\mathbf{m}_2 \tag{21}$$

where we have chosen to separate out the basis vector associated with the model-norm component. Taking the inner product with $\boldsymbol{\gamma}_\mathbf{d}$ yields

$$\boldsymbol{\gamma}_\mathbf{d}^T\boldsymbol{\delta}\mathbf{m} = \sum \alpha_i\boldsymbol{\gamma}_\mathbf{d}^T\mathbf{v}_i + \beta\boldsymbol{\gamma}_\mathbf{d}^T[\mathbf{m}^{(n)} - \mathbf{m}_0]. \tag{22}$$

At each iteration the α_i and β are to be found such that $\boldsymbol{\gamma}_\mathbf{d}^T\boldsymbol{\delta}\mathbf{m} = \phi_d^* - \phi_d$. If $\boldsymbol{\gamma}_\mathbf{d}^T[\mathbf{m}^{(n)} - \mathbf{m}_0]$ is near zero then the vector $\mathbf{m}^{(n)} - \mathbf{m}_0$ does not contribute significantly to the linearized change in misfit. Its use is relegated to reducing the model norm and this is efficiently achieved by having β be approximately -1. The model norm becomes

$$\begin{aligned} &\|\mathbf{W_m}[\mathbf{m}^{(n)} - \mathbf{m}_0 + \boldsymbol{\delta}\mathbf{m}]\|^2 \\ &\quad = \|\mathbf{W_m}\{\mathbf{m}^{(n)} - \mathbf{m}_0 + \boldsymbol{\delta}\mathbf{m}_1 - \beta[\mathbf{m}^{(n)} - \mathbf{m}_0]\}\|^2 \\ &\quad \approx \|\mathbf{W_m}\,\boldsymbol{\delta}\mathbf{m}_1\| \end{aligned} \tag{23}$$

and the updated model is $\mathbf{m}^{(n+1)} \approx \mathbf{m}_0 + \boldsymbol{\delta}\mathbf{m}_1$. Thus the updated model takes a leap towards the initial reference model. We have found that this generally occurs. Nevertheless eqs (19) can always be tried and validity established by a single forward modelling. If the perturbation is desirable, it is kept; if it is not desirable, then curvature information needs to be estimated and the inversion methodology can revert to that used in inversions I1–I4. To illustrate the efficacy of this approach we present inversions I5 and I6 which use the 23 basis vectors as in I3. In I5 eqs (19) are adopted directly and used until the achieved misfit (evaluated by forward modelling) is no longer within 50 per cent of the target value predicted from solving the linear equations. At that iteration the inversion reverts to using full curvature information and line searches. The convergence plots are shown in Fig. 5. We see that the linear equations have been effective for the first eight iterations and ϕ_d is reduced to 2443. At iteration 8, $\phi_m = 7.3$ (with $\phi_s = 4.5$, $\phi_x = 1.5$ and $\phi_z = 1.3$). These values are nearly identical to those at iteration 5 of I3. However, the computational savings to this point are substantial since only gradient information has been computed and only two forward modellings are carried out per iteration. At iteration 20 $\phi_m = 41.3$ (with $\phi_s = 6.6$, $\phi_x = 10.0$ and $\phi_z = 24.7$).

In inversion I6 we use eqs (19) to evaluate the coefficients α but implement a line search to find a best value of μ. The convergence plots are presented in Fig. 5. In this mode, five iterations were valid before the program required the use of curvature information. At iteration 5 $\phi_d = 3548$ and $\phi_m = 5.9$ (with $\phi_s = 4.1$, $\phi_x = 0.8$ and $\phi_z = 1.0$). At iteration

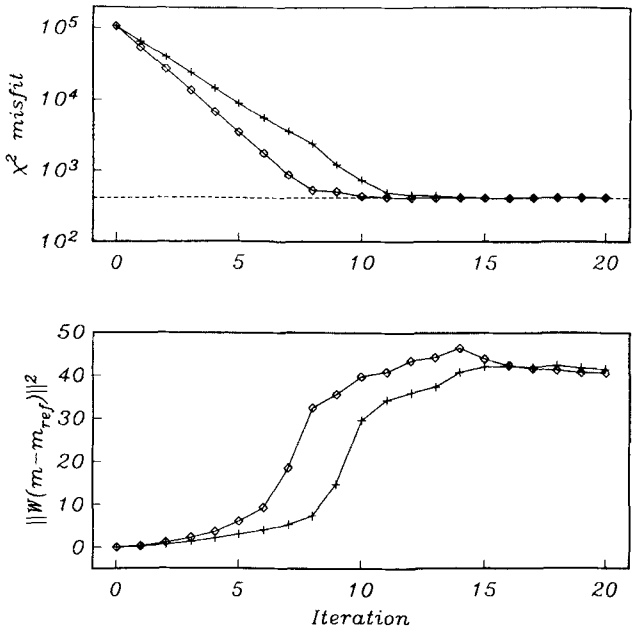

Figure 5. Convergence curves for inversions I5 (+) and I6 (◇). The inversions neglect misfit curvature information in early iterations and 23 basis vectors are used. In I5 the analytic expression of μ was used for the first eight iterations. Later iterations used curvature information and line searches. I6 neglects curvature information but incorporates a line search for the first five iterations. Subsequent iterations use full curvature information and line searches.

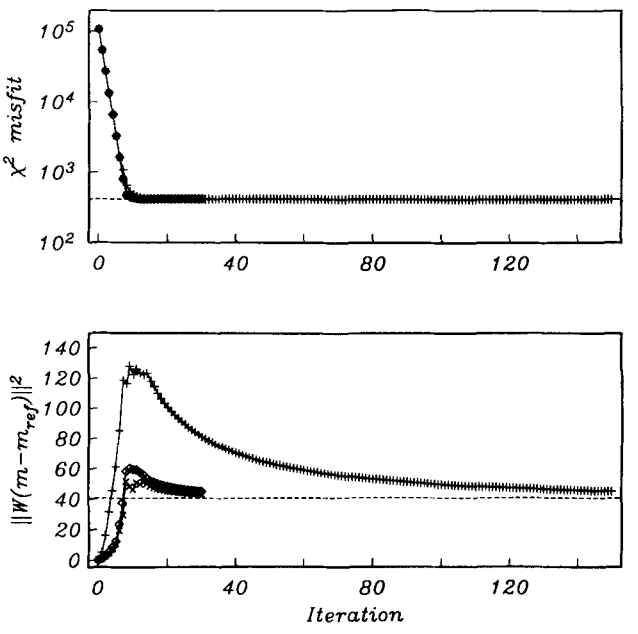

Figure 6. Convergence curves for inversions I7–I9 are represented by (+, ◇, ×) respectively. The effect of using gradient vectors rather than steepest descent vectors is illustrated by I7. The improvement obtained by preconditioning these vectors with a smoothing opertor in eq. (26) is shown by I8. In I9 the smoothing operator is applied twice.

20 $\phi_m = 40.6$ (with $\phi_s = 6.5$, $\phi_x = 9.9$ and $\phi_z = 24.1$). These values are identical to those achieved in inversion I3 at iteration 20.

7 APPROXIMATIONS TO $(\mathbf{W}_m^T\mathbf{W}_m)^{-1}$ AND INVERSIONS I7–I9

Preconditioning gradient vectors with $(\mathbf{W}_m^T\mathbf{W}_m)^{-1}$ is very important when working with only a few vectors in a subspace solution. To quantify the advantages of using steepest descent vectors we first carry out the inversion, referred to as I7, using only gradient vectors

$$\begin{aligned} \mathbf{v}_i &= \nabla_m \phi_d^i \qquad i = 1, 21 \\ \mathbf{v}_{22} &= \nabla_m \phi_m \\ \mathbf{v}_{23} &= \text{constant.} \end{aligned} \tag{24}$$

as basis vectors. Convergence plots are shown in Fig. 6. The desired misfit is achieved at iteration 14 but at that iteration $\phi_m = 123$. (with $\phi_s = 7.3$, $\phi_x = 37.2$ and $\phi_z = 78.7$). This model norm is more than three times as large as that achieved with descent vectors. The model at this stage is exceedingly rough, particularly in the near-surface blocks. Successive iterations reduce this roughness but even at iteration 100 the model norm $\phi_m = 49.6$ (with $\phi_s = 6.6$, $\phi_x = 13.1$ and $\phi_z = 29.9$). It does appear that with continued iterations, the gradient vectors may eventually produce the quality of solution obtained with descent vectors. However, the slowness with which this goal is achieved and the unattractive model produced at early iterations shows this route to be far less satisfactory than using descent vectors.

The matrix $\mathbf{W}_m^T\mathbf{W}_m$ is generally sparse, positive definite and symmetric. Since its inverse is to be applied to many vectors throughout the course of the inversion, it is reasonable to form $(\mathbf{W}_m^T\mathbf{W}_m)^{-1}$, store the result, and apply it directly. The efficiency of this approach however, decreases with the size of $\mathbf{W}_m^T\mathbf{W}_m$. Not only does the computation effort required to generate the inverse matrix increase, but the storage requirements and the application of the inverse can be burdensome. Note that even though $\mathbf{W}_m^T\mathbf{W}_m$ is sparse, $(\mathbf{W}_m^T\mathbf{W}_m)^{-1}$ is a full matrix.

There is a great difference in character between gradient and descent vectors. Yet the inversion results using these vectors are similar provided that sufficient iterations are allowed. This leads us to investigate whether an operator which approximates the effect of $(\mathbf{W}_m^T\mathbf{W}_m)^{-1}$ could be applied to gradient vectors and yield convergence rates which are near those achieved from steepest descent vectors. In our example $\mathbf{W}_m^T\mathbf{W}_m$ is a roughening matrix. Neglecting details about cell dimensions, the operator $\mathbf{W}_m^T\mathbf{W}_m$ is approximated by the 2-D stencil

$$\begin{pmatrix} 0 & -1 & 0 \\ -1 & 4+\epsilon & -1 \\ 0 & -1 & 0 \end{pmatrix}. \tag{25}$$

This structure arises because we have used first-order differences in the vertical and horizontal directions for the $\mathbf{W}_x$ and $\mathbf{W}_z$ matrices in eq. (16); the ϵ accounts for $\mathbf{W}_s$. As the operator in eq. (25) is moved over a 2-D grid, its effect is to roughen the vector it is operating upon. Conversely, the inverse operator must be a smoothing operator. Many approximate choices are available. We look for a nine-element 2-D operator which when convolved with (25)

produces an output that is approximately an impulse. The operator provided by the stencil

$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix} \tag{26}$$

achieves this goal. The 2-D convolution of (25) and (26) is a 5×5 matrix with a central element surrounded by eight zeros and an outer ring of small truncation values. Inversion I8 illustrates the results achieved by applying the approximate inverse operator to all gradient vectors. The convergence rates, shown in Fig. 6, are greatly improved compared to those achieved with gradient vectors. The target misfit was attained at iteration 12. At iteration 20 $\phi_m = 48.4$ (with $\phi_s = 7.5$, $\phi_x = 9.6$ and $\phi_z = 31.3$). These results are nearly as good as those obtained by using an exact $(\mathbf{W}_\mathbf{m}^T\mathbf{W}_\mathbf{m})^{-1}$ matrix. In fact, allowing the inversion to continue for another 10 iterations produced $\phi_m = 44.9$ (with $\phi_s = 7.1$, $\phi_x = 9.5$ and $\phi_z = 28.4$).

The convergence rates achieved with the operator in (26) are fairly insensitive to minor alterations in the coefficients. Interestingly, most changes designed to achieve better agreement between a unit impulse and the output of a 2-D convolution produce a larger value for the central element. This in turn reduces the smoothing characteristics of the operator and the convergence rates decrease somewhat. Applying the smoothing twice as in inversion I9, yielded an additional improvement. Convergence plots are given in Fig. 6. At iteration 20 $\phi_m = 45.3$ (with $\phi_s = 7.4$, $\phi_x = 9.1$ and $\phi_z = 28.8$). Why a double smoothing leads to an improvement is not known. The reasons may lie in the fact that stencils (25) and (26) are constructed with the assumption of uniform cells whereas the true aspect of ratios of cells can be large. Such details aside, these results indicate that approximate descent vectors can be almost as efficient as exact descent vectors and that any reasonable approximation to $(\mathbf{W}_\mathbf{m}^T\mathbf{W}_\mathbf{m})^{-1}$ might produce near-optimal results.

In a second approach to generating an approximate $(\mathbf{W}_\mathbf{m}^T\mathbf{W}_\mathbf{m})^{-1}$ we solve a system of equations directly. Our goal is to compute $\mathbf{v}_k = (\mathbf{W}_\mathbf{m}^T\mathbf{W}_\mathbf{m})^{-1}\gamma_k$ where γ_k denotes a typical gradient vector. Rewriting this as

$$\mathbf{W}_\mathbf{m}^T\mathbf{W}_\mathbf{m}\mathbf{v}_k = \gamma_k \tag{27}$$

produces the desired system of equations which we solve using a conjugate gradient technique (D'Azeveda, Kightley & Forsyth, 1991). This is quite efficient since $\mathbf{W}_\mathbf{m}^T\mathbf{W}_\mathbf{m}$ is sparse. In fact for the size of problem treated here, it is more efficient to apply $(\mathbf{W}_\mathbf{m}^T\mathbf{W}_\mathbf{m})^{-1}$ exactly by using a conjugate gradient method than it is to store the matrix and read it. This computational advantage is further enhanced if one does not require that the equations be solved to high accuracy. Let

$$\xi = \frac{\|\gamma_k - \mathbf{W}_\mathbf{m}^T\mathbf{W}_\mathbf{m}\mathbf{v}_k\|}{\|\gamma_k\|} \tag{28}$$

denote a maximum relative misfit error used as a convergence criterion in the conjugate gradient solver. Inversions I10–I14 have been carried out with values of $\xi = 10^{-6}$, 0.1, 0.50, 0.75 and 0.90. The convergence curves are given in Fig. 7. Significant deterioration in the rate of convergence is not

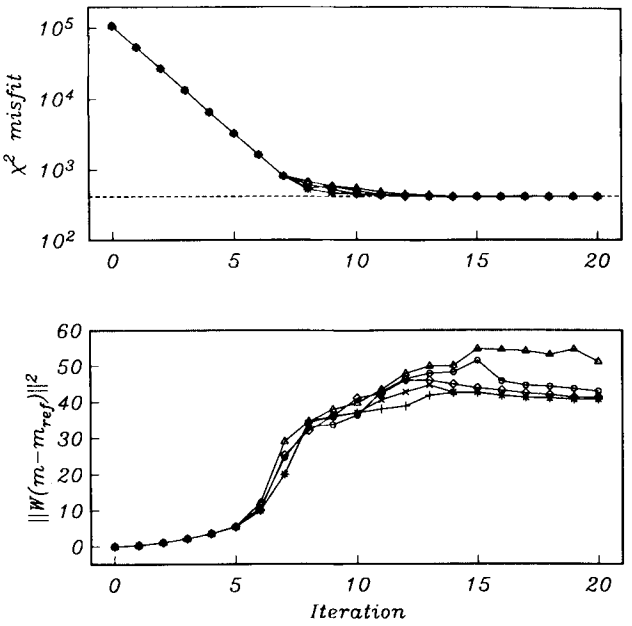

Figure 7. Approximate steepest descent vectors are obtained by soving a system of equations using a conjugate gradient solver. The convergence curves for relative rms misfits for $\xi = 10^{-6}$, 0.1, 0.25, 0.5, 0.9 are shown respectively by the symbols (+, ×, ◇, ○, △).

observed until $\xi = 0.50$. This further substantiates our hypothesis that even somewhat poor approximations to $(\mathbf{W}_\mathbf{m}^T\mathbf{W}_\mathbf{m})^{-1}$ can provide excellent preconditioners for gradient vectors.

8 CONSISTENCY OF APPLICATION OF $(\mathbf{W}_\mathbf{m}^T\mathbf{W}_\mathbf{m})^{-1}$ AND INVERSIONS I15–I16

The previous sections have emphasized the importance of using descent vectors in the inversion. The gradient of the model norm component is

$$\nabla_\mathbf{m}\phi_m = \mathbf{W}_\mathbf{m}^T\mathbf{W}_\mathbf{m}(\mathbf{m} - \mathbf{m}_0) \tag{29}$$

and so a steepest descent vector $\mathbf{m} - \mathbf{m}_0$ is available without additional computation. It might be thought that this vector should be used straight away and that the approximate inverse operator be applied only to the data gradients. This appears not to be so. Our work has shown that application of a preconditioning should be carried out consistently on all gradient vectors. To illustrate this we present two final inversions. In I15 the basis vectors are

$$\begin{aligned} \mathbf{v}_k &= \nabla_\mathbf{m}\phi_d^k \quad k = 1, 21 \\ \mathbf{v}_{22} &= (\mathbf{W}_\mathbf{m}^T\mathbf{W}_\mathbf{m})^{-1}\nabla_\mathbf{m}\nabla_\mathbf{m}\phi_m = (\mathbf{m} - \mathbf{m}_0) \\ \mathbf{v}_{23} &= \text{constant.} \end{aligned} \tag{30}$$

that is, $(\mathbf{W}_\mathbf{m}^T\mathbf{W}_\mathbf{m})^{-1}$ is not consistently applied to all gradient vectors. The convergence results are shown in Fig. 8. The target misfit is achieved after 11 iterations but the model norm is $\phi_m = 118$. Importantly, there is little recovery from this state and the final model remains rough so long as the subspace vectors are mixtures of gradient and descent vectors. At iteration 20, $\phi_m = 116$. (with $\phi_s = 7.8$, $\phi_x = 28.8$ and $\phi_z = 79.4$). We note that better results were obtained by working purely with gradient vectors. (compare with results for I7, Fig. 6).

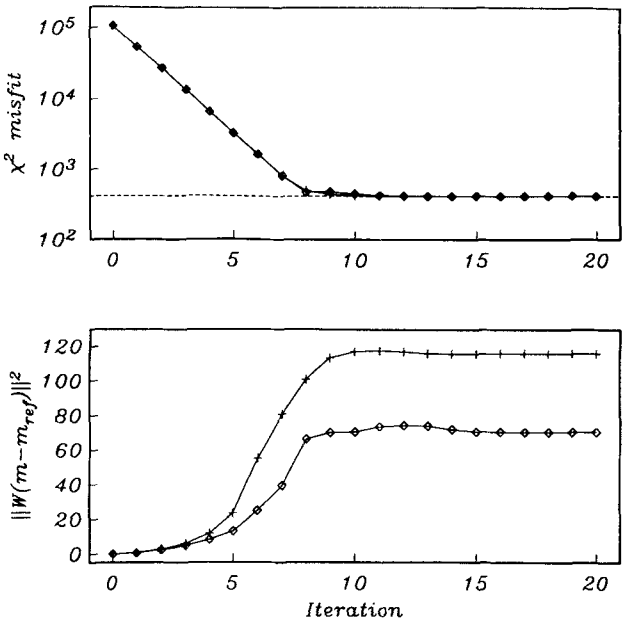

Figure 8. Inversions I15 and I16 illustrate the need for being consistent in the use of (approximate) steepest descent vectors and gradient vectors. I15 (+) is an inversion which uses gradient vectors for the misfit functional but a steepest descent vector for the model norm. I16 (◇) is an improved result obtained by preconditioning the misfit gradient vectors with the smoothing operator in eq. (26) but the results are still inferior to those in Fig. 6 where the same operator was applied to all gradient vectors.

In the next example we reduce this inconsistency somewhat by working with the vectors

$$\mathbf{v}_k = (\widetilde{\mathbf{W}_m^T \mathbf{W}_m})^{-1} \nabla_m \phi_d^k \quad k = 1, 21$$
$$\mathbf{v}_{22} = (\mathbf{W}_m^T \mathbf{W}_m)^{-1} \nabla_m \phi_m = (\mathbf{m} - \mathbf{m}_0) \tag{31}$$
$$\mathbf{v}_{23} = \text{constant}.$$

where $(\widetilde{\mathbf{W}_m^T \mathbf{W}_m})^{-1}$ is the operator given in (26). The convergence results for I16 are presented in Fig. 8. These are improved results but still inferior to those achieved in I8 with $(\widetilde{\mathbf{W}_m^T \mathbf{W}_m})^{-1}$ applied uniformly. The model norm now plateaus at $\phi_m = 70.4$ (with $\phi_s = 8.7$, $\phi_x = 10.3$ and $\phi_z = 51.5$).

In a general sense the need for consistency in using gradient or descent vectors is not surprising. Consider the use of gradient vectors alone in making up the model. Both $\nabla_m \phi_d$ and $\nabla_m \phi_m = \mathbf{W}_m^T \mathbf{W}_m (\mathbf{m} - \mathbf{m}_0)$ are 'rough' vectors. It is possible, and is noted in practise, that these two 'rough' vectors can be combined to generate a smooth perturbation. Alternatively, by using a 'smooth' model vector $(\mathbf{W}_m^T \mathbf{W}_m)^{-1} \nabla_m \phi_m = (\mathbf{m} - \mathbf{m}_0)$ and 'rough' data gradient vectors, the flexibility to remove roughness in the model is greatly reduced. Hence the model norm plateaus at significantly high values in these circumstances.

9 DISCUSSION

As we increase the size of our inverse problems to handle the true complexity of earth structure we inevitably reach a limit dictated by our computing power. The three main computational roadblocks are forward modelling, calculation of sensitivities and solving a large system of equations. As such, an ideal inversion methodology requires a minimum number of forward modellings, a minimum amount of sensitivity information, and the inversion of only a small matrix. In this paper we investigate some of the benefits of subspace methods in reducing the computational burden. The efficacy of a subspace approach is dependent upon the number and character of the selected spanning vectors. Steepest descent vectors associated with (subdivided) data misfit and model norm components of the objective function are always valuable and we have focused upon these. The examples presented here suggest that at initial iterations it may be possible to neglect misfit curvature information and a full iteration can be carried out with a single forward modelling or a few forward modellings if a line search is instigated. Eventually, the curvature information must be included but excellent convergence rates can be obtained with relatively few basis vectors. Our examples illustrate the improved convergence rates achievable with steepest descent vectors compared to gradient vectors. Moreover, we show that good convergence rates can also be obtained by using only approximate steepest descent vectors. Such vectors can be generated by applying a simple smoothing operator to the gradient vectors or by using an iterative equation solver like the conjugate gradient method to solve the system to a predetermined accuracy. Irrespective of the details of the preconditioner, it is important that the same preconditioning be applied to all gradient vectors. Failure to do so can produce an inferior solution.

ACKNOWLEDGMENTS

This work was supported by NSERC grant 5-84270 and 5-80141. The authors are grateful for that support.

REFERENCES

D'Azevedo, E. F., Kightley, J. R. & Forsyth, P. A., 1991. *MATI Iterative Sparse Matrix Solver: User's guide*, Department of Computer Science, University of Waterloo, Orlando, Florida.

Gill, P. E., Murray, W. & Wright, M. H., 1981. *Practical Optimization*, Academic Press, New York.

Hestenes, M. R., 1980. *Conjugate Direction Methods in Optimization*, Springer-Verlag, Berlin.

Kennett, B. L. N. & Williamson, P. R., 1988. Subspace methods for large-scale nonlinear inversion, in *Mathematical Geophysics: a survey of recent developments in seismology and geodynamics*, pp. 139–154, eds Vlaar, N. J., Nolet, G., Wortel, M. J. R. & Cloetingh, S. A. P. L., D. Reidel, Dordrecht.

McGillivray, P. R., 1992. Forward modeling and inversion of dc resistivity and MMR data, *PhD thesis*, Department of Geophysics and Astronomy, University of British Columbia, Vancouver, Canada.

Skilling, J. & Bryan, R. K., 1984. Maximum entropy image reconstruction: general algorithm, *Mon. Not. R. astr. Soc.* **211**, 111–124.