Inverse Problems 10 (1994) 915–935. Printed in the UK

Subspace linear inverse method

Douglas W Oldenburg and Yaoguo Li

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

Received 2 August 1993

Abstract. This paper presents a robust, flexible and efficient algorithm to solve large scale linear inverse problems. The method is iterative and at each iteration a perturbation in a q-dimensional subspace of an M-dimensional model space is sought. The basis vectors for the subspace are primarily steepest descent vectors obtained from segmenting the data misfit and model objective functions. The efficiency of the algorithm is realized because only a $q \times q$ matrix needs to be inverted at each iteration instead of a matrix of order M. As M becomes large the number of computations per iteration is of order qNM where N is the number of data. An important feature of our algorithm is that positivity can easily be incorporated into the solution. We do this by introducing a two-segment mapping which transforms positive parameters to parameters defined on the real line. The nonlinear mapping requires that a line search involving forward modelling is implemented so that at each iteration we obtain a model which misfits the data to a predetermined level. This obviates the need to carry out additional inversions with trial and error selection of a Lagrange multiplier. In this paper we present the details of the subspace algorithm and explore the effect on convergence of using different strategies for selecting basis vectors and altering adjustable parameters which control the rate of decrease in the misfit and rate of increase in the model norm as a function of iteration number.

1. Introduction

In a typical linear problem we are provided with N data d_j^{obs} and mappings of the form $d_j = \mathcal{L}_j[m]$ which express the relationship between the jth datum d_j and a model m. The data constitute a set of constraints upon the model and in the inverse problem we attempt to find a model which acceptably reproduces these data. The fundamental difficulty is that of non-uniqueness; there are generally infinitely many models which adequately reproduce the observations. Practical inverse problems are therefore formulated by first designing a specific objective functional of the model and then minimizing this functional subject to the data constraints. Generally the objective function is tailored so that the solution from the inverse algorithm is 'close' to a prespecified reference model and also that the constructed model has 'minimum structure' in some sense. A particularly useful objective function for a three dimensional model in a Cartesian coordinate frame is

$$\phi_m(m) = \alpha_s \int_{\text{vol}} w_s(m - m_0)^2 \, \mathrm{d}v + \alpha_x \int_{\text{vol}} w_x \left(\frac{\partial (m - m_0)}{\partial x} \right)^2 \mathrm{d}v + \alpha_y \int_{\text{vol}} w_y \left(\frac{\partial (m - m_0)}{\partial y} \right)^2 \mathrm{d}v + \alpha_z \int_{\text{vol}} w_z \left(\frac{\partial (m - m_0)}{\partial z} \right)^2 \mathrm{d}v. \tag{1}$$

In equation (1) the functions w_s, w_x, w_y, w_z are specified by the user and the constants α_s, α_x, α_y and α_z control the importance of closeness of the constructed model to the reference

0266-5611/94/040915+21$19.50

model m_0 and control the roughness of the model in the three directions. Pragmatic issues generally force us to discretize the model. Here we parametrize the model by a set of M rectangular prisms and require that the model be constant within each cell. With this parametrization the objective function can be written as

$$\phi_m(\mathbf{m}) = (\mathbf{m}-\mathbf{m}_0)^T \left\{\alpha_s \mathbf{W}_s^T\mathbf{W}_s + \alpha_x \mathbf{W}_x^T\mathbf{W}_x + \alpha_y \mathbf{W}_y^T\mathbf{W}_y + \alpha_z \mathbf{W}_z^T\mathbf{W}_z\right\}(\mathbf{m}-\mathbf{m}_0). \tag{2}$$

The elements in the $M \times M$ matrices $\mathbf{W}_s$, $\mathbf{W}_x$, $\mathbf{W}_y$, $\mathbf{W}_z$ are related to the cell dimensions. Equation (2) can be written generically as

$$\phi_m(\mathbf{m}) = (\mathbf{m}-\mathbf{m}_0)^T \mathbf{W}_m^T\mathbf{W}_m(\mathbf{m}-\mathbf{m}_0) = ||\mathbf{W}_m(\mathbf{m}-\mathbf{m}_0)||^2. \tag{3}$$

We will use this form for most derivations even though $\mathbf{W}_m$ may not be explicitly computed. There is no loss of generality because our final computational equations will require only $\mathbf{W}_m^T\mathbf{W}_m$ and this is easily evaluated by using equation (2).

With the same model parametrization the forward mapping $\mathcal{L}[\mathbf{m}] = \mathbf{d}$ takes the form $\mathbf{Gm} = \mathbf{d}$ where $\mathbf{G}$ is an $N \times M$ matrix. The inverse problem is now formulated as the optimization problem:

$$\begin{aligned} &\text{minimize} && \phi_m = ||\mathbf{W}_m(\mathbf{m}-\mathbf{m}_0)||^2 \\ &\text{subject to} && \phi_d = ||\mathbf{W}_d(\mathbf{Gm}-\mathbf{d}^{\text{obs}})||^2 = \phi_d^* \end{aligned} \tag{4}$$

where ϕ_d is our misfit criterion, $\mathbf{W}_d$ is an $N \times N$ data weighting matrix and ϕ_d^* is the target misfit. If the 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 $\mathbf{W}_d = \text{diag}\{1/\sigma_1, ..., 1/\sigma_N\}$. With this assumption, ϕ_d is a 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 optimization in equation (1) can be attacked in a variety of ways. Three approaches are standard.

OPT I. Let $\mathbf{x} = \mathbf{W}_m(\mathbf{m}-\mathbf{m}_0)$ and define $\mathbf{A} = \mathbf{W}_d\mathbf{G}\mathbf{W}_m^{-1}$ and $\mathbf{b} = \mathbf{W}_d(\mathbf{Gm}_0 - \mathbf{d}^{\text{obs}})$. The optimization problem becomes

$$\begin{cases} \text{minimize} & \phi_m = ||\mathbf{x}||^2 \\ \text{subject to} & \phi_d = ||\mathbf{Ax}-\mathbf{b}||^2 = \phi_d^*. \end{cases} \tag{5}$$

The problem is solved to find $\mathbf{x}$ and the model is recovered from $\mathbf{m} = \mathbf{W}_m^{-1}\mathbf{x} + \mathbf{m}_0$.

OPT II. Form the objective function

$$\phi(\mathbf{m}) = \phi_m(\mathbf{m}) + \mu^{-1}(\phi_d(\mathbf{d}) - \phi_d^*) \tag{6}$$

where μ is the Lagrange multiplier. Differentiate with respect to the model parameters to obtain

$$\mathbf{Bm} = \mathbf{b}$$

where

$$\mathbf{B} = \mu\mathbf{W}_m^T\mathbf{W}_m + \mathbf{G}^T\mathbf{W}_d^T\mathbf{W}_d\mathbf{G}$$

and

$$\mathbf{b} = \mu \mathbf{W}_m^T \mathbf{W}_m \mathbf{m}_0 + \mathbf{G}^T \mathbf{W}_d^T \mathbf{W}_d \mathbf{d}^{\text{obs}}.$$

The $M \times M$ matrix system is solved for $\mathbf{m}$.

OPT III. Augment the matrix $\mathbf{G}$ with the equations pertaining to the model objective function, and minimize the misfit of the overdetermined matrix system. That is, minimize

$$\phi_d = ||\mathbf{Cm} - \mathbf{f}||^2 \tag{8}$$

where

$$\mathbf{C} = \begin{bmatrix} \mathbf{W}_d\mathbf{G} \\ \sqrt{\mu\alpha_s}\mathbf{W}_s \\ \sqrt{\mu\alpha_x}\mathbf{W}_x \\ \sqrt{\mu\alpha_y}\mathbf{W}_y \\ \sqrt{\mu\alpha_z}\mathbf{W}_z \end{bmatrix} \quad \text{and} \quad \mathbf{f} = \begin{bmatrix} \mathbf{W}_d\mathbf{d}^{\text{obs}} \\ \sqrt{\mu\alpha_s}\mathbf{W}_s\mathbf{m}_0 \\ \sqrt{\mu\alpha_x}\mathbf{W}_x\mathbf{m}_0 \\ \sqrt{\mu\alpha_y}\mathbf{W}_y\mathbf{m}_0 \\ \sqrt{\mu\alpha_z}\mathbf{W}_z\mathbf{m}_0 \end{bmatrix}. \tag{9}$$

This requires the solution of an $(N + 4M) \times M$ matrix.

The matrix equations in the above problems can be solved using a variety of approaches. Direct decomposition methods (e.g. singular value decomposition) are most appealing but they become computationally expensive if $\mathbf{A}$, $\mathbf{B}$ or $\mathbf{C}$ becomes large. For this reason iterative approaches are generally used. In OPT I the minimum norm solution for $\mathbf{x}$ lies in the row space of the $N \times M$ matrix $\mathbf{A}$. Row action methods (e.g. SIRT) can be used if the equations $\mathbf{Ax} = \mathbf{b}$ are rescaled and if the iterative process is initiated with the zero solution (van der Sluis and Van der Vorst 1987). Alternatively a projection method such as LSQR (Paige and Saunders 1982) can be viable. This is particularly true if the technique, like LSQR, minimizes $||\mathbf{Ax} - \mathbf{b}||^2 + \lambda||\mathbf{x}||^2$ so that solutions for multiple values of λ can be had with little extra effort. The $M \times M$ matrix $\mathbf{B}$ in equation (7) is positive definite and symmetric. A viable, and popular approach to solving this system is to use a projection method like Conjugate Gradient (Hestenes 1980, Golub and Van Loan 1989). Similarly a (bi)conjugate gradient method (Press *et al* 1992) or LSQR can be used to solve the $(N+4M) \times M$ matrix $\mathbf{C}$ in equation (9).

OPT I, OPT II and OPT III methodologies have successfully solved many inverse problems. In our inversions however, they have not always met our needs. OPT I requires the formation of $\mathbf{W}_m$ and its inverse. Because our objective functions are generally complicated, $\mathbf{W}_m^{-1}$ is not available. The strategy of OPT I is therefore prohibited by this task alone. The impediments to using OPT II and OPT III are that the matrices $\mathbf{B}$ and $\mathbf{C}$ can be large, dense and poorly conditioned. In addition, our desire to find a model which misfits the data by a target value ϕ_d^* requires that the correct value of the Lagrange multiplier be specified. Since this is unknown, the solutions via OPT II and OPT III must be carried out a number of times to achieve this result.

In this paper we concentrate upon a fourth strategy which embodies elements of OPT I and OPT II. We formulate the optimization as per OPT II but choose to solve the equations iteratively. At each iteration we search for a model perturbation $\delta\mathbf{m} \in \mathbb{R}^M$ which is a linear combination of vectors $\{\mathbf{v}_i\}$, $i = 1, q$. The model perturbation $\delta\mathbf{m} = \sum_{i=1}^{q} \alpha_i \mathbf{v}_i$ which lies in a q-dimensional subspace of $\mathbb{R}^M$, 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. The success or failure of a subspace approach therefore hinges upon the selection of the spanning vectors for the activated subspace. Our philosophy for selecting vectors is based upon the following ideas. The two objective functions of importance to the inversion are ϕ_m and ϕ_d. Steepest descent vectors associated with each of these quantities are therefore fundamental to the inversion process. Importantly we can partition each of these objective functions and compute steepest descent vectors associated with each segment. Steepest descent vectors formed from segments of ϕ_d are intimately associated with the vectors in OPT I while vectors formed from segments of ϕ_m are associated with vectors from OPT II and thus the subspace approach has elements of both of the standard techniques. In our approach a best estimate for the Lagrange multiplier is found at each iteration and this obviates the need to solve the inverse problem with trial values of μ as required when using OPT II. The caveat to this statement is that in practise the errors on the data, which include additive noise and the effects of discrepancies between the mathematical model and the physical system being emulated, are sometimes unknown. Under such circumstances it may be required to carry out inversions with different target misfits and it can be argued that this is equivalent to solving the inversion with different values of μ. Nevertheless, in our work we have found it advantageous to specify a desired misfit *a priori* and then to alter that misfit depending upon the performance of the inversion.

In the following sections we present the subspace approach and examine the effect of different choices of search vectors and the effects of adjusting parameters related to the selection of the Lagrange multiplier. To illustrate convergence characteristics we invert synthetic induced polarization (IP) data taken over a 2D earth. We next show how the equations can to be altered to enforce positivity of the solution and apply the technique to inverting our IP data. The paper closes with our thoughts about practical utility of the subspace approach in solving linear inverse problems.

2. Subspace algorithm

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 and Bryan (1984) who maximized an entropy norm suitable for image reconstruction, by Kennett and Williamson (1988) and Kennett *et al* (1988) in their inversion of seismic data and by Oldenburg *et al* (1993) in their inversion of DC resistivity data. Those papers all dealt with nonlinear problems. Here we concentrate upon the linear inverse problem.

Consider a model objective function

$$\phi_m(m) = \alpha_s \int_{\text{vol}} w_s(m - m_0)^2 \, \mathrm{d}v + \alpha_x \int_{\text{vol}} w_x \left(\frac{\partial m}{\partial x}\right)^2 \mathrm{d}v + \alpha_y \int_{\text{vol}} w_y \left(\frac{\partial m}{\partial y}\right)^2 \mathrm{d}v + \alpha_z \int_{\text{vol}} w_z \left(\frac{\partial m}{\partial z}\right)^2 \mathrm{d}v. \tag{10}$$

This is the same objective function as in equation (1) with the exception that the reference model has been removed from the derivative terms. The reason for this is practical. In

many inversion problems we are willing to provide an estimate for the value of the model at a specific location but we are less confident about gradients. In such circumstances we do not want to penalize differences in the gradient between the constructed and reference model and we therefore omit m_0 in the gradient terms. Discretization of equation (10) yields

$$\phi_m(\mathbf{m}) = \alpha_s (\mathbf{m}-\mathbf{m_0})^T \mathbf{W}_s^T\mathbf{W}_s(\mathbf{m}-\mathbf{m_0}) + \mathbf{m}^T \left\{\alpha_x \mathbf{W}_x^T\mathbf{W}_x + \alpha_y\mathbf{W}_y^T\mathbf{W}_y + \alpha_z\mathbf{W}_z^T\mathbf{W}_z\right\}\mathbf{m}. \tag{11}$$

This can be written as

$$\begin{aligned}\phi_m(\mathbf{m}) &= (\mathbf{m}-\mathbf{m_0})^T\mathbf{W}_m^T\mathbf{W}_m(\mathbf{m}-\mathbf{m_0}) + 2\mathbf{m}^T\mathbf{W}_v^T\mathbf{W}_v\mathbf{m_0} - \mathbf{m_0}^T\mathbf{W}_v^T\mathbf{W}_v\mathbf{m_0} \\ &= ||\mathbf{W}_m(\mathbf{m}-\mathbf{m_0})||^2 + 2\mathbf{m}^T\mathbf{W}_v^T\mathbf{W}_v\mathbf{m_0} - ||\mathbf{W}_v\mathbf{m_0}||^2 \end{aligned} \tag{12}$$

where $\mathbf{W}_v^T\mathbf{W}_v = \alpha_x\mathbf{W}_x^T\mathbf{W}_x + \alpha_y\mathbf{W}_y^T\mathbf{W}_y + \alpha_z\mathbf{W}_z^T\mathbf{W}_z$. We note that when $\mathbf{m_0}$ is constant, $\mathbf{W}_v\mathbf{m_0} = 0$ and the model objective function reduces to that given in equation (4).

Let $\mathbf{m}^{(n)}$ be the model at the nth iteration and let $\{\mathbf{v}_i\}$, $i = 1, q$ be arbitrary vectors which form a q-dimensional subspace of $\mathbb{R}^M$. We seek a model perturbation of the form $\delta\mathbf{m} = \sum\alpha_i\mathbf{v_i} = \mathbf{V}\alpha$. Substituting into (6) yields

$$\phi(\alpha) = ||\mathbf{W}_m(\mathbf{m}^{(n)} + \mathbf{V}\alpha - \mathbf{m_0})||^2 + \mu^{-1}(||\mathbf{W}_d(\mathbf{G}\mathbf{m}^{(n)} + \mathbf{G}\mathbf{V}\alpha - \mathbf{d}^{\text{obs}})||^2 - \phi_d^*). \tag{13}$$

Setting $\nabla_\alpha\phi(\alpha) = 0$ yields

$$\mathbf{B}\alpha = \mathbf{b} \tag{14}$$

where

$$\mathbf{B} = \mathbf{V}^T(\mathbf{G}^T\mathbf{W}_d^T\mathbf{W}_d\mathbf{G} + \mu\mathbf{W}_m^T\mathbf{W}_m)\mathbf{V}$$

and

$$\mathbf{b} = -\mu\mathbf{V}^T\mathbf{W}_m^T\mathbf{W}_m(\mathbf{m}^{(n)} - \mathbf{m_0}) - \mu\mathbf{V}^T\mathbf{W}_v^T\mathbf{W}_v\mathbf{m_0} - \mathbf{V}^T\mathbf{G}^T\mathbf{W}_d^T\mathbf{W}_d(\mathbf{G}\mathbf{m}^{(n)} - \mathbf{d}^{\text{obs}}).$$

The matrix $\mathbf{B}$ is a $q \times q$ positive definite and symmetric matrix and is easily inverted provided that q is relatively small. We note that this is the same matrix as in OPT II except it has been contracted with the matrix $\mathbf{V}$.

The success of the subspace methodology depends strongly upon the choice of basis vectors. As in Oldenburg *et al* (1993), we partition the misfit and model objective functions. The misfit objective function $\phi_d = (\mathbf{d}-\mathbf{d_0})^T\mathbf{W}_d^T\mathbf{W}_d^k(\mathbf{d}-\mathbf{d_0})$ may be partitioned as

$$\phi_d = \sum\phi_d^k \tag{15}$$

where the kth subset is $\phi_d^k = (\mathbf{d}^k - \mathbf{d}_0^k)^T\mathbf{W}_d^{kT}\mathbf{W}_d^k(\mathbf{d}^k - \mathbf{d}_0^k)$ and the gradient is $\gamma_k = \nabla_m\phi_d^k$. Our norm on model space is controlled by the symmetric positive-definite matrix $\mathbf{W}_m^T\mathbf{W}_m$. A steepest decent† direction can be obtained by multiplying the gradient by $(\mathbf{W}_m^T\mathbf{W}_m)^{-1}$ (Gill *et al* 1981, p 102). We therefore choose vectors

$$\mathbf{v_k} = (\mathbf{W}_m^T\mathbf{W}_m)^{-1}\nabla_m\phi_d^k \tag{16}$$

† Ascent, really, but we are interested only in directions and therefore ignore the minus sign.

as elements in our subspace. The application of $(\mathbf{W}_m^T\mathbf{W}_m)^{-1}$ produces precise steepest descent vectors only when $\mathbf{m}_0$ is a constant for then $\mathbf{W}_v^T\mathbf{W}_v\mathbf{m}_0 = \mathbf{0}$ in the definition of the objective function. If $\mathbf{m}_0$ is not constant then application of $\mathbf{W}_m^T\mathbf{W}_m)^{-1}$ provides only approximate steepest descent vectors. However, it was shown empirically in Oldenburg *et al* (1993) that convergence in the subspace inversion is not substantially altered so long as the vectors are converted to approximate steepest descent vectors and that the operator generating the vectors is applied uniformly to all gradient vectors. This empirical result also permits setting $\alpha_s = 0$ in the model objective function. The approximate steepest descent vector $\mathbf{v}_k$ is then found by using a conjugate gradient algorithm to solve $(\mathbf{W}_m^T\mathbf{W}_m + \epsilon\mathbf{I})\mathbf{v}_k = \nabla_m\phi_d^k$ where ϵ is a small positive quantity and $\mathbf{I}$ is the identity matrix.

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 $\mathbf{W}_d$. The gradient vector for ϕ_d^k is

$$\begin{aligned}\nabla_m\phi_d^k &= 2W_{kk}^2(d_k - d_{0k})\nabla_m d_k \\ &= 2W_{kk}^2(d_k - d_{0k})\mathbf{g}_k\end{aligned} \tag{17}$$

where $\mathbf{g}_k$ is the kth row of the matrix $\mathbf{G}$. 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 $\mathbf{G}$. 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 partitioning of ϕ_d can be effected in a variety of ways. Data groups can be selected arbitrarily, or since physical data are often acquired so that numerous data are associated with a particular source or receiver then data may be grouped on that basis. A more systematic approach is to select data groupings according to misfit. Let $\phi_d^{(n)}$ be the misfit at the nth iteration and let n_d denote the number of data groupings. The data equations are first reordered in terms of misfit with the first equation having the largest misfit. The data groupings are then established by successively including data equations until the cumulative misfit of each group is approximately $\phi_d^{(n)}/n_d$.

The next set of vectors to be included in the subspace should be sensitive to ϕ_m. The steepest descent vector $\mathbf{v} = (\mathbf{W}_m^T\mathbf{W}_m)^{-1}\nabla_m\phi_m$ is always useful but more flexibility to construct a minimum norm model is achieved by subdividing the model objective function. Let ϕ_m^k $(k = 1, 4)$ denote the kth term in the model objective function in equation (10). Steepest descent vectors

$$\mathbf{v_k} = (\mathbf{W}_m^T\mathbf{W}_m)^{-1}\nabla_m\phi_m^k \tag{18}$$

are routinely used in our algorithm. In addition we use the constant vector. We note that the sum of the vectors in equation (18) is equal to the steepest descent vector for ϕ_m. Potentially useful vectors can also be obtained by partitioning this vector. For 2D problems, the cells along individual rows or individual columns can be grouped and the projection of $(\mathbf{W}_m^T\mathbf{W}_m)^{-1}\nabla_m\phi_m$ onto those rows or columns provide good search directions, especially if the earth model has strong lateral or vertical continuity. The analogous situation in 3D is to use horizontal or vertical planes of cells.

We offer no definite way to prescribe an optimum strategy for subdividing $\nabla_m\phi_m$ nor for specifying additional vectors. The process can be dynamic in that it changes with iteration. For instance cells associated with rows could be used at one iteration and cells associated with columns used on the next. The process can also be interactive. Viewing the model can prompt the hypothesis that a certain portion of the model has too much roughness or

has other undesirable characteristics. This area can be subdivided into smaller groups of cells to provide additional flexibility in the inversion. The attractive aspect of the dynamic use of additional search vectors is that in the subspace inversion only a model perturbation is sought. At worst, a poor choice of additional vectors produces little benefit. We note that in the limit of subdividing $\nabla_m \phi_m$ each cell becomes a basis vector. The resulting M vectors completely span model space and no further basis vectors are required. The 'subspace' solution reverts to the traditional solution and the inversion of an $M \times M$ matrix is demanded.

Despite the variety of possible search vectors that can be tried for any problem, we have adopted a simple and automatic strategy which requires no user input other than to specify the number of vectors associated with the data misfit objective function. In all examples we use only the vectors given by equation (16), equation (18) and the constant vector.

3. Numerical considerations and implementation

In addition to selection of search vectors there are other practical issues to be addressed in the subspace methodology. These include guarding against poor conditioning of the matrix $\mathbf{V}$, solving the matrix system of equations and assessing computational requirements. We address these items here.

The subspace formulation demands the inversion of the matrix $\mathbf{V}^T(\mathbf{G}^T\mathbf{W}_d^T\mathbf{W}_d\mathbf{G} + \mu\mathbf{W}_m^T\mathbf{W}m)\mathbf{V}$. This matrix is singular if the column vectors of $\mathbf{V}$ are linearly dependent. We guard against poor conditioning by orthonormalizing the descent vectors prior to using them in the subspace equations. This is done by using a singular-value decomposition algorithm.

The solution of the matrix system in equation (14) is attacked in the following manner. At each iteration the tradeoff between the Lagrange multiplier, the model objective function and the misfit is typified by the curves in figure 1. At the beginning of the $(n+1)$th iteration the data misfit and model objective value are given by $\phi_d^{(n)}$ and $\phi_m^{(n)}$, respectively. Let ϕ_d^a denote the asymptotic data misfit as μ approaches zero. Our goal is to find a perturbation such that the misfit is reduced (if we are not already at the target misfit ϕ_d^*) and which does not allow the model objective function to increase too much. Generally, we are willing to accept a slower convergence path involving smooth models than a route which permits a great deal of roughness to be built up in the model and which subsequently requires a large number of iterations to remove. To achieve this we introduce user supplied constants β_1, β_2, β_3. Define μ^{tar} as that value of μ such that the misfit is equal to $\phi_d^{\text{tar}} = \beta_1\phi_d^{(n)}$, where β_1 sets the rate of misfit reduction. If μ^{tar} does not exist, that is if $\phi_d^{\text{tar}} < \phi_d^a$ then set $\mu^{\text{tar}} = 0$. Define μ^{tr} as that value of μ such that the misfit is equal to $\beta_2\phi_d^a$ and define μ^{mod} as that value of μ such that the model objective function is equal to $\beta_3\phi_m^{(n)}$, where β_3 limits the rate of model norm increase. The selected value of μ is defined by $\mu = \max\{\mu^{\text{tar}}, \mu^{\text{tr}}, \mu^{\text{mod}}\}$.

A primary motivation for developing the subspace computation is to obtain a numerically efficient algorithm to solve linear inverse problems as M becomes large. The major computations are listed in table 1 and the total number of operations per iteration of the subspace inversion is approximately $(1+(1/N)(2q+(r+1)l)+q/M)qNM$. Under usual conditions $M \gg N \gg q > l$ so the majority of the computations are taken up in the matrix multiplications to form $\mathbf{V}^T\mathbf{G}^T\mathbf{W}_d^T\mathbf{W}_d\mathbf{G}\mathbf{V}$ and $\mathbf{V}^T\mathbf{W}_m^T\mathbf{W}_m\mathbf{V}$. The number of iterations required for convergence is problem dependent but is principally controlled by the number of search vectors used and the selection of β_1, β_2, β_3 which control the model perturbation at each iteration.

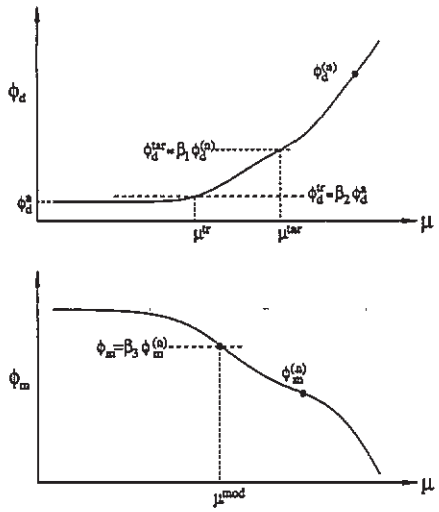

Figure 1. A typical misfit curve $\phi_d(\mu)$ at the $(n+1)$th iteration is given at the top and a typical model norm curve $\phi_m(\mu)$ is shown below. $\phi_d^{(n)}$ and $\phi_m^{(n)}$ are the objective functions at the nth iteration, ϕ_d^{∞} is the asymptotic misfit as μ approaches zero, and β_1, β_2 and β_3 are parameters specified by the user. The curves were obtained from an iteration of the subspace program and all axes are logarithmic. The Lagrangian multiplier is chosen as the maximum of μ^{tr}, mu^{tar}, and μ^{mod}.

Table 1. The major computing operations are listed on the left with the number of operations the right. There are M cells in the model, N data, and q search vectors. l is the number of non-zero elements in each row of $\mathbf{W}_m^T\mathbf{W}_m$ ($l = 7$ for a 3D problem), r is the number of iterations in the conjugate gradient solution required to generate descent vectors (r is typically 20).

Item	Number of iterations
Applying $(\mathbf{W}_m^T\mathbf{W}_m)^{-1}$ to gradient vectors	$rqlM$
Orthogonalization of vectors	q^2M
$(\mathbf{GV})^T(\mathbf{GV})$	$qNM + q^2N$
$\mathbf{V}^T\mathbf{W}_m^T\mathbf{W}_m\mathbf{V}$	$qM(l+q)$

In synthetic modelling, where the data errors are known and correctly specified, the algorithm has performed well. When inverting field data however, we have not always found a model which fits to a desired misfit. We attribute this non-convergence to an incorrect specification of the standard errors of the data. In a typical non-convergent situation, the data misfit changes very little with successive iterations while the model roughness progressively increases. When this occurs, it is necessary to adjust the target misfit to be somewhat above the achievable misfit (or equivalently, to adjust the estimated standard deviations of the data), and restart the inversion.

The algorithm always uses the vectors $(\mathbf{W}_m^T\mathbf{W}_m)^{-1}\nabla\phi_m$ and $(\mathbf{W}_m^T\mathbf{W}_m)^{-1}\nabla\phi_d$. Since these vectors can be combined into a steepest descent vector of the total objective function it would seem that all theorems pertaining to the convergence of iterative solutions for our matrix system by using steepest descent vectors can form a foundation for convergence analysis of the subspace method. Beyond this, there is no quantitative statement that can be made about the general convergence properties of the algorithm. Convergence rates are dictated by the selection of the number and type of search vectors and the three adjustable constants.

4. Subspace inversion of induced polarization data

To illustrate the subspace algorithm we invert computer simulated data from an IP experiment. The interested reader is referred to Sumner (1976) or Fink *et al* (1990) for more information about the IP method and an introduction to the existing literature. The reasons for choosing IP as the example are twofold: (i) IP surveys are commonly used to find subsurface mineralization, and (ii) the nature of the forward mapping characterized by the matrix **G** typifies matrices in other geophysical surveys. The matrix **G** is full and cells close to the sources or receivers exercise great influence upon recorded data while cells at depth have progressively smaller influence. Insight about the algorithm that is gleaned from the IP example will be beneficial in solving the IP inverse problem and also in solving other practical linear inverse problems.

In an IP experiment a DC electric current is input to the ground and the electric potential is measured away from the source. In the field, four electrodes are used. Two are connected to the current generator to provide a closed circuit for the current and two electrodes are needed for measuring a potential difference. However, because of superposition, we can consider a pole–pole experiment in which one current electrode and one potential electrode are moved to 'infinity'. This geometry can be well approximated in field acquisition and is modelled theoretically.

The relationship between intrinsic chargeability **m** and IP data is $\mathbf{Gm} = \mathbf{d}$ where

$$G_{ij} = \frac{\partial \ln \phi_i}{\partial \ln \sigma_j}. \tag{19}$$

In equation (19) ϕ_i is the ith potential in the IP survey and σ_j is the electrical conductivity of the jth cell. The conductivity structure shown in figure 2(a), and recovered from a previous nonlinear inversion of DC potentials, is assumed known. The sensitivities G_{ij} can be calculated once electrode positions have been specified. The chargeability model is shown in figure 2(b). It consists of a chargeable layer with $m = 0.05$ at the surface and a chargeable block of $m = 0.15$ at depth. The centre of the chargeable block has been offset laterally from the position of the conductive high in figure 2(a) and the model has undergone a slight smoothing. Both of these alterations make the synthetic model realistic.

The geophysical survey is carried out by placing surface electrodes every ten metres in the interval $x = (-100, 100)$ metres. Each of the twenty one electrode positions can be activated as a current site and when it is, electric potentials are recorded at the remaining electrodes. The IP potentials for this test are obtained by computing the elements G_{ij} corresponding to the conductivity in figure 2(a) and taking the product **Gm** where **m** is the chargeability in figure 2(b). Each of the 420 data is then contaminated with Gaussian noise having a standard deviation equal to 0.001. This corresponds to about 3% of the average value of the data. The IP data are plotted in pseudosection form in figure 2(c). The surface chargeable layer is visible in the data image but there is little manifestation of the chargeable block which is the target for this inversion.

Our example inverse problem is of intermediate size and consists of 1296 unknowns which are to be determined from 420 data. This is smaller than many of the problems we have been working with but it typifies how the algorithm works and it is small enough that we can run many inversions to study the effects of parameter choices. The inverse problem is attacked by forming an objective function as per equation (10) but neglecting the y-dependence. Moreover, the general background chargeability is zero and correspondingly we set $\mathbf{m}_0 = 0$. The discrete model objective function becomes

$$\phi_m = \mathbf{m}^T \left\{ \alpha_s \mathbf{W}_s^T \mathbf{W}_s + \alpha_x \mathbf{W}_x^T \mathbf{W}x + \alpha_z \mathbf{W}_z^T \mathbf{W}_z \right\} \mathbf{m}. \tag{20}$$

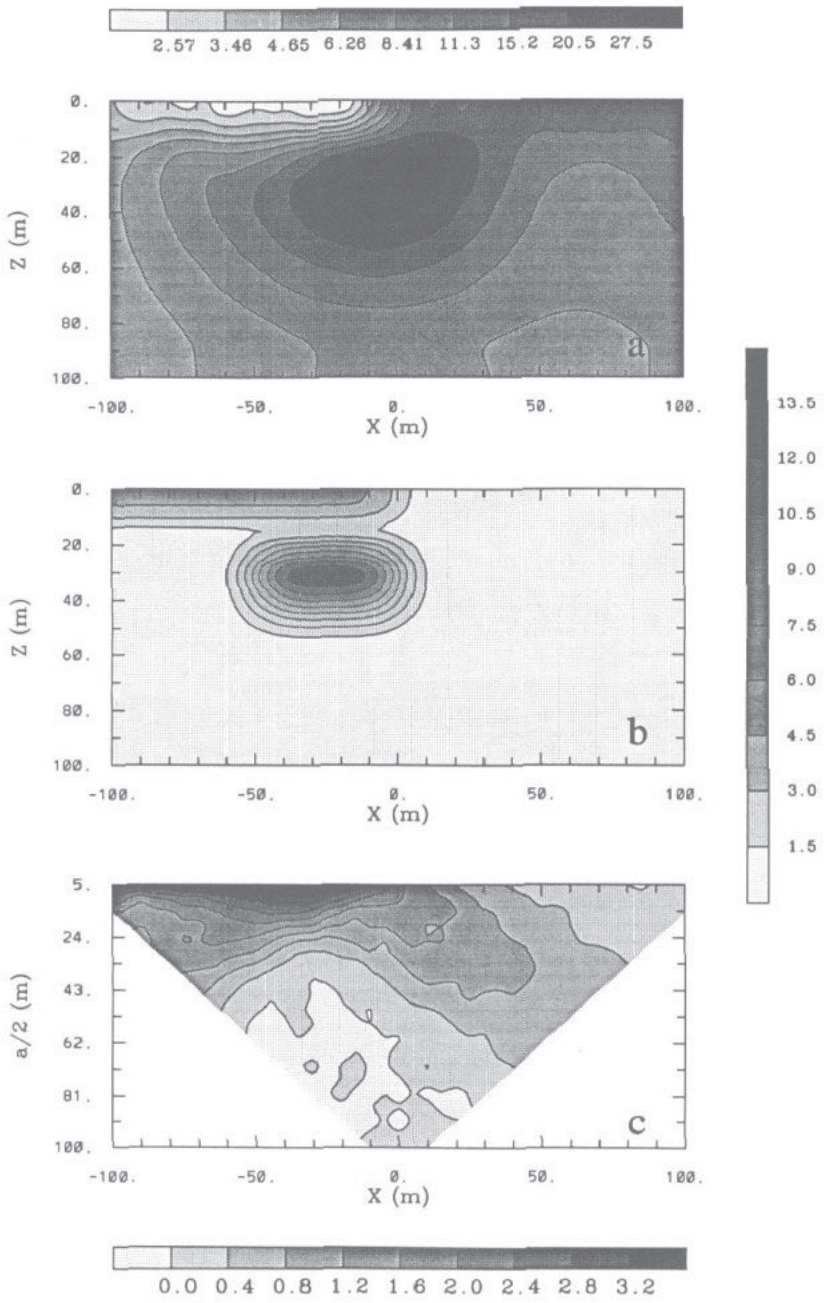

Figure 2. The conductivity model σ from which the matrix **G** is formed is given in (*a*). The grey scale is in mS/m. The chargeability model is in (*b*) and the noise contaminated IP data are displayed in (*c*). Each datum is plotted midway between the current and potential electrode and at a depth $a/2$ where a is the separation between the electrodes. True chargeability is confined to the range [0,1). The values displayed in (*b*) and (*c*) are scaled by 100 and the respective grey scales lie to the right and beneath the images.

In equation (20) $\mathbf{W}_s$ is a diagonal matrix with elements Δz where Δx is the length of the cell and Δz is its thickness, $\mathbf{W}_x$ has elements $\pm\sqrt{\Delta z/\delta x}$ where δx is the distance between the centres of horizontally adjacent cells, and $\mathbf{W}_z$ has elements $\pm\sqrt{\Delta x/\delta z}$ where δz is the distance between the centres of vertically adjacent cells. We have chosen $\alpha_s = 10^{-4}$, $\alpha_x = 1.0$, $\alpha_z = 1.0$.

The inversion was carried out using 25 vectors consisting of 21 groupings of the data (one group for each current electrode source), three vectors from the model objective function (as per equation (18) and the constant vector. The recovered model, shown in figure 3, can be compared with the true model in figure 2. Both the surface chargeability and buried chargeable block are well delineated. The amplitude of the chargeable block is somewhat less than that of the true model and it extends to greater depth. This is characteristic of our model objective function which discriminates against variation in the horizontal and vertical directions. We also notice the large regions of small negative values on the left- and right-hand sides of the model. The convergence plots shown in figure 3 are typical for subspace inversions. The misfit reduction in the early iterations is limited by the choice of β_1 which is 0.5 for this example. As the misfit is reduced there is a region where it is not possible to achieve the desired target without increasing the model norm by a factor greater than β_3 which has been set to 2.0 for this example. The final target misfit is achieved at iteration 12 and is maintained with further iterations. At early iterations ϕ_m is small and there is no restriction on its relative change between iterations. When ϕ_d approaches within a factor of 10 of the final target misfit, the restriction $\beta_3 = 2.0$ is invoked and this is manifested as the straight line between iterations 4 and 8. Once the final target misfit is achieved the inversion continues and generally produces a decrease in the model norm. The algorithm is terminated at iteration 17 when the convergence criterion $||\delta\mathbf{m}||/||\mathbf{m}|| < 0.01$ was met.

5. Exploration of parameters

To obtain some insight about how the convergence and model are affected by using different search vectors and by altering the parameters β_1 and β_3 ($\beta_2 = 1.0001$ for all inversions carried out here) we present the following results. Unless explicitly stated, each inversion is carried out with the same choice of search vectors and parameters as the base example in figure 3.

We first investigate the effects of using various groupings of the data equations. Convergence curves for $n_d = 1, 2, 4, 10, 21, 42, 63$ are shown in figure 4. The curves have the following characteristics. Irrespective of the number of data groupings the algorithm is able to produce the desired reduction in misfit of a factor of 2 in the early iterations. However, as the misfit decreases, the limitations of having very few vectors prevent the algorithm from achieving a reduction of this magnitude. More basis vectors are advantageous. At some limit, however, additional vectors provide imperceptible benefit and the convergence is limited by β_3. The model objective function curves all have the same characteristics. They achieve a maximum at the iteration at which the desired target misfit is achieved and then decrease and asymptote to a common value. The convergence curves corresponding to 42 and 63 data groupings are indistinguishable and show that the benefit of having extra vectors is limited.

One of our goals in designing a workable algorithm is to minimize the user input. In figure 4 we explored the effects of changing the number of data groupings but did not investigate the effects of vector selection in making up the groups. We therefore consider

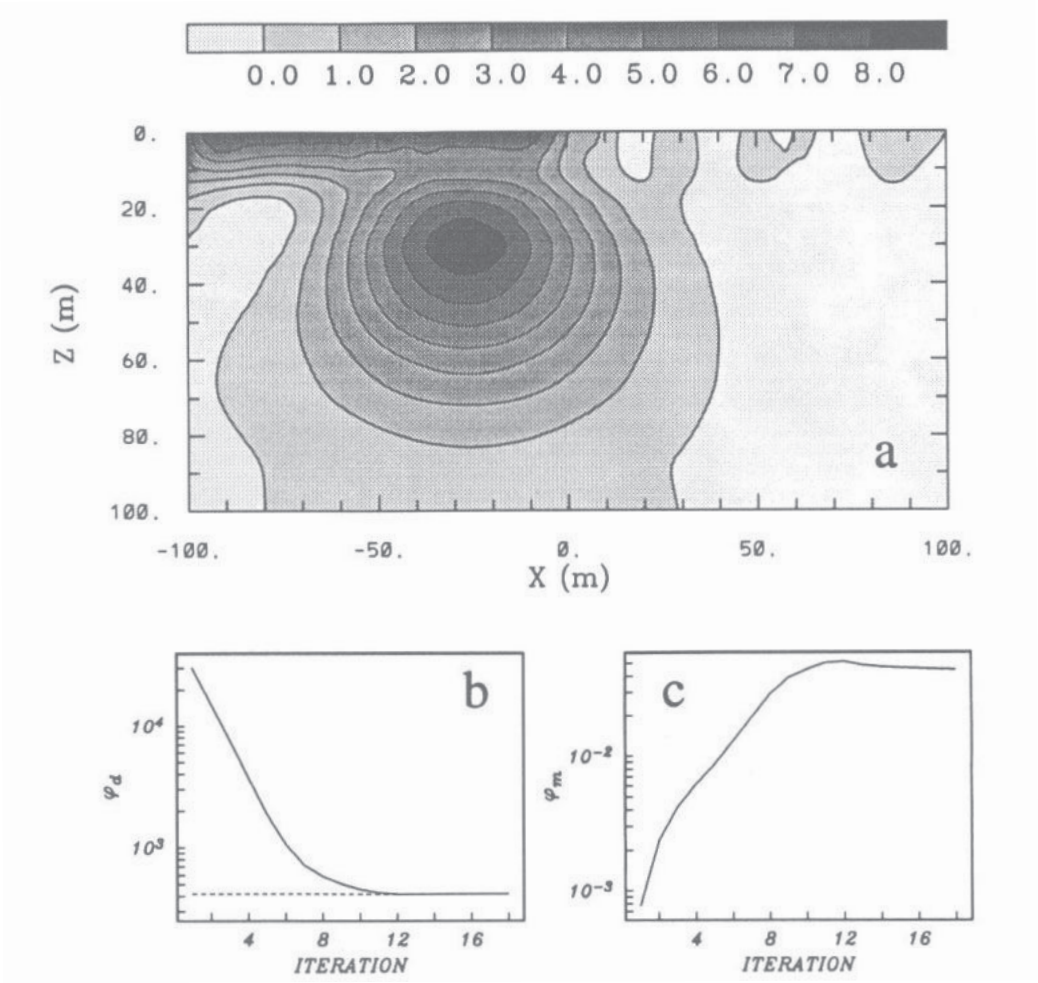

Figure 3. Inversion result obtained by grouping the data equation into 21 fixed groups. The recovered chargeability in panels (*a*) can be compared with the true chargeability in figure 2(*b*). Convergence curves are shown below. Panel (*b*) is the data misfit and (*c*) is the model objective function. The dashed line in (*b*) indicates the target misfit for the inversion.

the base example which has $n_d = 21$ but at each iteration we order the misfit equations according to misfit and group the data so that each grouping contributes equally to the total misfit. The results of the inversion are presented in figure 5 and can be compared with those in figure 3. The differences are not dramatic but some improvement is noticeable; the desired misfit is achieved one iteration earlier, the maximum in ϕ_m is slightly lower, and final convergence is achieved in one less iteration.

The effects of altering the choices of β_1 and β_3 are displayed in figure 6. The contour plots, generated from 100 inversion runs, indicate the number of iterations required before the target misfit of $\phi_d = 420$ is achieved and the number of iterations required before the final convergence criterion is satisfied. In order to make the comparisons meaningful the convergence criterion was the achievement of $\phi_m < 0.042$. This was the value of ϕ_m achieved in the base example at iteration 17. The results in figure 6 are generally as expected. The number of iterations required to achieve the desired misfit decreases as

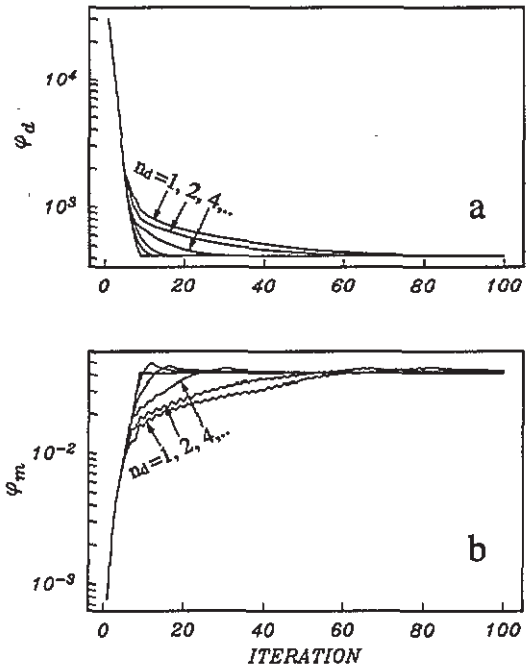

Figure 4. Convergence curves obtained by using $n_d = 1, 2, 4, 10, 21, 42, 63$ groupings of the data equations, respectively. The data misfit and model objective function are shown in panels (a) and (b) respectively. All the convergence curves have the same characteristics. The model objective function peaks at the iteration at which the desired misfit is achieved, and asymptotes to a common value in the subsequent iterations. 82 iterations are required before the inversion with $n_d = 1$ achieves the target misfit of 420. The curves for $n_d = 42$ and $n_d = 63$ are indistinguishable.

β_3 increases provided that β_1 is sufficiently small. In the region $3.5 < \beta_3 < 4.5$ and $0.01 < \beta_1 < 0.1$ only 6 to 8 iterations are required to reach the desired misfit. Those parameters are not optimum for overall convergence however, since about 24 iterations are required. This is a reflection that too rapid achievement of the desired misfit may produce unwanted structure which requires additional iterations to eliminate. There is fine scale structure in figure 6(b) but the general conclusion is that good choices for the parameters are $\beta_1 < 0.6$ and $1.5 < \beta_3 < 3.0$. Our values of $\beta_1 = 0.5$ and $\beta_3 = 2.0$ were reasonable, if not slightly conservative choices.

A practical difficulty encountered in field data inversions is that we do not know the 'error' in the data. Even if the data errors did satisfy the criteria of being Gaussian and independent (which is unlikely) there is still the difficulty in estimating a standard deviation for each datum. If the assigned standard deviation is too small then the algorithm will attempt to fit the noise and the constructed model will have artifacts. To illustrate the performance characteristics of the subspace algorithm in the event of asking for an unrealistically small data misfit (or equivalently an assignment of standard deviations which are too small) we redo the base inversion but ask for a target misfit of 10 instead of 420. The result in shown in figure 7. The misfit is 420 at iteration 12 and continues to decrease, but slowly. It decreases to 380 by iteration 100 and to 359 by iteration 500. The model norm is characterized by a general increase as a function of iteration. By iteration 500 it has reached 0.108 and is still increasing. The additional structure can be evidenced by

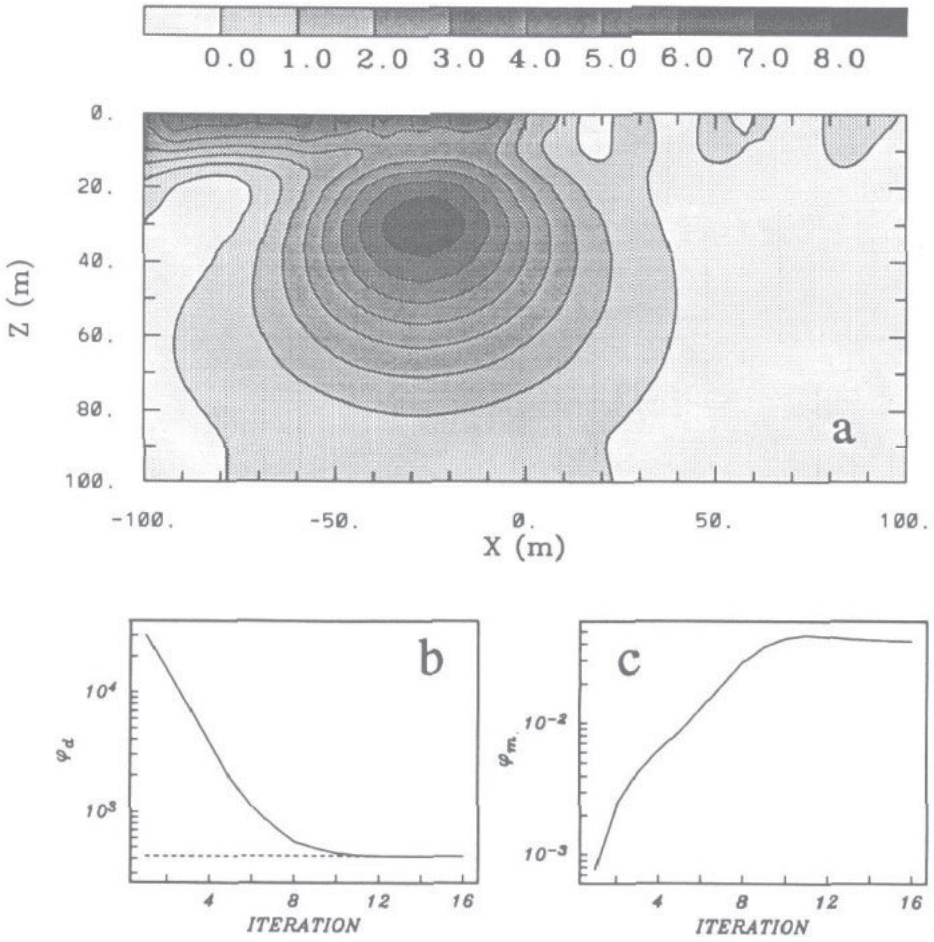

Figure 5. Inversion result obtained by grouping equations according to data misfit. $n_d = 21$ and each group contributes equal amounts ($\phi_d^{(n)} / n_d$) to the total misfit. The recovered model in (a) is virtually identical to that in figure 3(a) which was obtained using fixed data groupings. The data misfit and model objective functions are shown in panels (b) and (c), respectively.

comparing with the result in figure 3. Overall, the results in figure 7 are encouraging. The true misfit between the accurate and inaccurate data is 442. In asking for a misfit which is substantially lower than this value, the algorithm enters a mode where it is unable to significantly reduce the misfit and yet the model norm continues to increase. Setting a target misfit somewhat above the asymptotic misfit value and restarting the inversion would seem to be a viable methodology.

6. Subspace solutions with a positivity constraint

In many inverse problems the unknown variable is positive. Intrinsic chargeability is in this category and is confined to the region [0,1). However, without explicit imposition of a positivity constraint the constructed model can be expected to have regions of negative

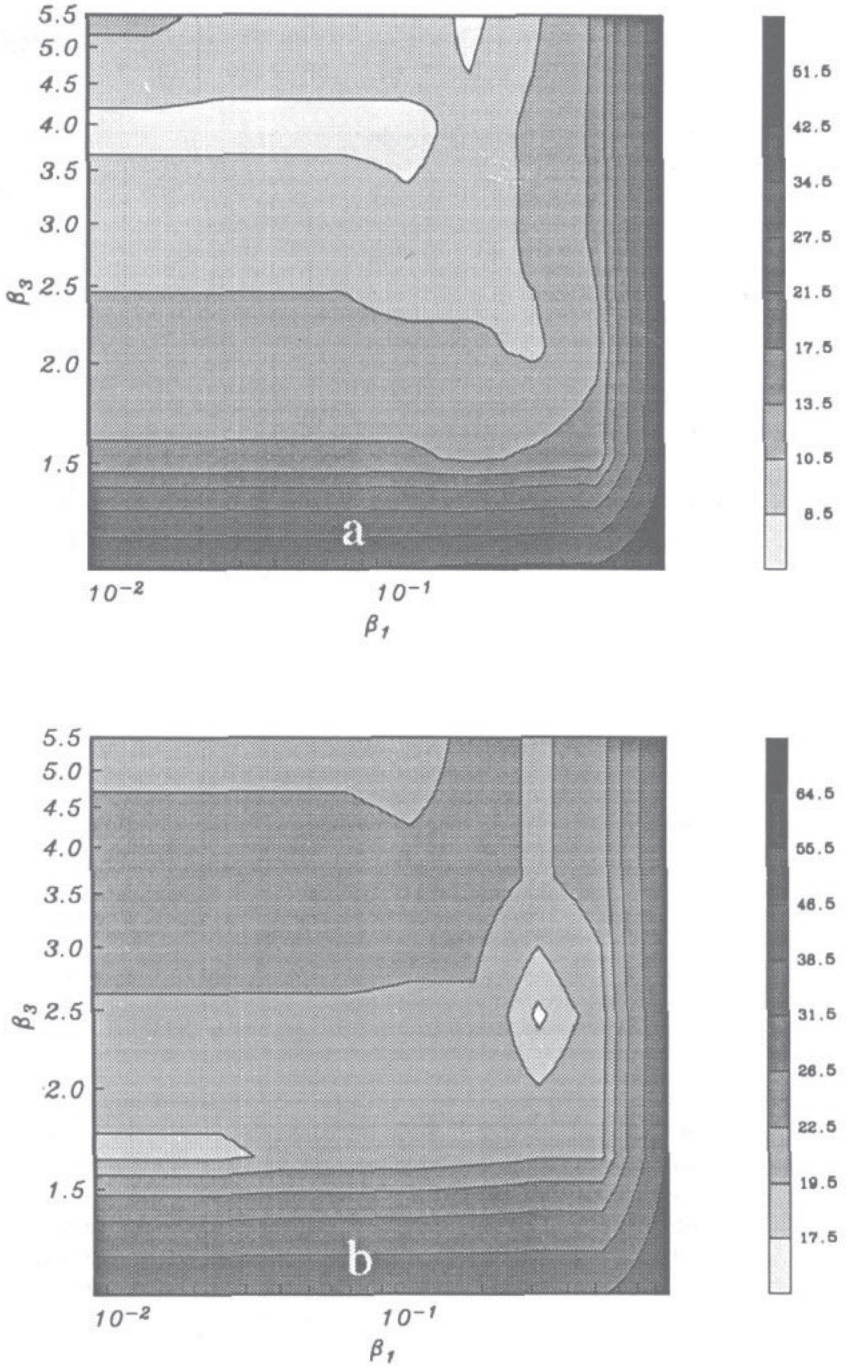

Figure 6. The effect on the convergence by varying β_1 and β_3 which determine the selection of the Lagrange multiplier at each iteration. The number of iterations required to find a model which misfits the data to a final target value of 420 is presented in panel (*a*). The number of iterations required to achieve convergence defined by having $\phi_d = 420$ and $\phi_m < 0.042$ is in panel (*b*). All inversions start with a common initial model.

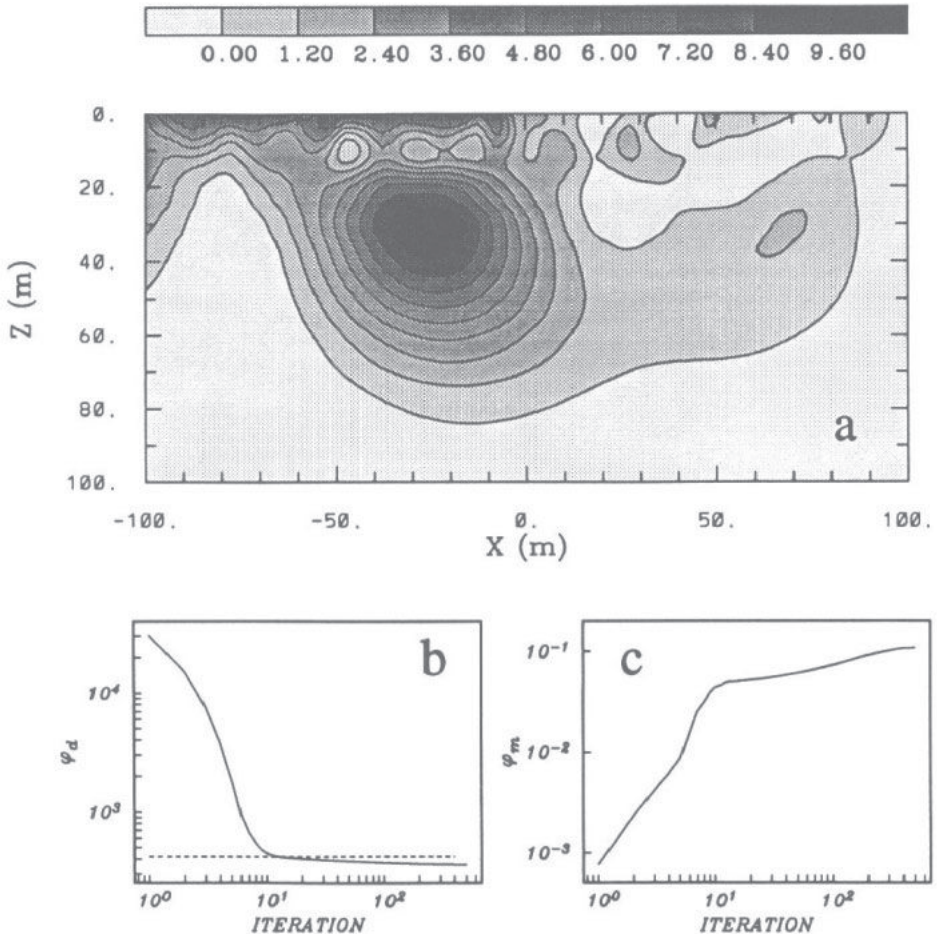

Figure 7. The model and convergence curves obtained when an attempt is made to overfit the data. The target misfit is 10. Five hundred iterations are performed. The final misfit was 359. Note the additional structure in the model compared to that in figure 3.

values. This is illustrated by the solution in figure 3(a). In that example the amplitude of the negative numbers is small enough that the interpretation is not greatly hindered. However, that is not always the case for IP inversion, nor is it the case when inverting other data types when the model is strictly positive.

Our goal therefore is to introduce a modification to the subspace optimization so that the recovered model remains positive. The subspace algorithm intrinsically works with functions which have positive and negative values and so to invoke positivity we introduce a mapping between the positive model and a function $p \in (-\infty, \infty)$. Let $m = f(p)$ and $p = f^{-1}(m)$ be mappings linking the two functions. Let $\mathbf{p}^{(n)}$ denote the model at the nth iteration and $\delta\mathbf{p}$ denote the sought perturbation. Performing a Taylor expansion of the perturbed model objective function about the point $\mathbf{p}^{(n)}$ yields

$$\phi_m(\mathbf{p}^{(n)} + \delta\mathbf{p}) = ||\mathbf{W}_m\mathbf{F}\delta\mathbf{p} + \mathbf{W}_m(\mathbf{m}^{(n)} - \mathbf{m}_0)||^2 + 2(\mathbf{m}^{(n)} + \mathbf{F}\delta\mathbf{p})^T\mathbf{W}_v^T\mathbf{W}_v\mathbf{m}_0 - ||\mathbf{W}_v\mathbf{m}_0||^2 \tag{21}$$

where **F** is a diagonal matrix with the elements

$$F_{ii} = \left.\frac{\partial f_i}{\partial p_i}\right|_{p^{(n)}} = \left.\frac{\partial m_i}{\partial p_i}\right|_{p^{(n)}}. \tag{22}$$

A similar Taylor expansion applied to the misfit objective functional $\phi_d(\mathbf{p}^{(n)} + \delta\mathbf{p})$ yields

$$\phi_d = ||\mathbf{W}_d\mathbf{J}\mathbf{F}\delta\mathbf{p} + \mathbf{W}_d(\mathbf{d}(\mathbf{p}^{(n)}) - \mathbf{d}^{\text{obs}})||^2. \tag{23}$$

The minimization of equation (21) subject to $\phi_d = \phi_d^*$ is solved in precisely the same manner as in section 2. The perturbation $\delta\mathbf{p}$ is represented as $\mathbf{V}\alpha$ and the objective function $\phi = \phi_m + \mu^{-1}(\phi_d - \phi_d^*)$ is solved by minimizing with respect to the coefficients α. The final equations are

$$\begin{aligned}
&\mathbf{B}\alpha = \mathbf{b} \\
&\mathbf{B} = \mathbf{V}^T\mathbf{F}^T(\mu\mathbf{G}^T\mathbf{W}_d^T\mathbf{W}_d\mathbf{G} + \mathbf{W}_m^T\mathbf{W}_m)\mathbf{F}\mathbf{V} \\
&\mathbf{b} = -\mu\mathbf{V}^T\mathbf{F}^T\mathbf{G}^T\mathbf{W}_d^T\mathbf{W}_d(\mathbf{d} - \mathbf{d}^{\text{obs}}) - \mathbf{V}^T\mathbf{F}^T\mathbf{W}_m^T\mathbf{W}_m(\mathbf{m}^{(n)} - \mathbf{m}_0) - \mu\mathbf{V}^T\mathbf{F}^T\mathbf{W}_v^T\mathbf{W}_v\mathbf{m}_0.
\end{aligned} \tag{24}$$

These are the same equations as in equation (14) with the exception that the matrix **V** has been replaced by the matrix **FV**. The only computational difference in the solution with positivity from that presented earlier is that the line search to find the appropriate Lagrange multiplier must now invoke a forward modelling. This involves an extra xNM operations where x is typically four to six.

The principle concern is the choice of mapping. The goal is to find a function which keeps a linear relationship between m and p in the range of m which is characteristic of the model values, and which maps a large range of p, corresponding to 'small' values of m, into zero. This has prompted us to use a two-segment mapping composed of an exponential and a straight line. The mapping is defined by

$$m = \begin{cases} 0 & p < p_b \\ e^p - e^{p_b} & p_b \leqslant p \leqslant p_1 \\ (p - p_1 + 1)e^{p_1} - e^{p_b} & p > p_1 \end{cases}. \tag{25}$$

The mapping and the first derivative are continuous at the transition point $p = p_1$. Values for the parameters p_b and p_1 can be assigned using the following reasoning. Let m_c denote a characteristic value of the model and let m_b be a threshold such that values of m below this are not sensibly distinguished from zero. For example, setting m_c is reasonable and thus $p_b = \ln m_b$ is defined. The specification of p_1 requires two considerations. In the region near m_c we desire a linear relationship between p and m and accordingly $p_1 = \ln m_c$ is a reasonable choice. However, the ith row of **V** is multiplied by F_{ii} and if this value is too small, the ith row of **V** is essentially annihilated and correspondingly there will be no possibility of adjusting the value of the ith cell. Therefore the values of F_{ii} should not be too disparate (factors of 10–100 might be reasonable limits) and this in turn limits the relative sizes of p_1 and p_b. Also p_1 and p_b should not be too close because in the limit $p_1 \to p_b$, the mapping in (25) degenerates to a linear truncation and the algorithm may not converge.

The inversion of the IP data with a positivity constraint and choice of $p_b = -6.9$ ($m_b = 0.001$) and $p_1 = -3.0$ ($m_1 = 0.05$) is shown in figure 8. The model is an improved

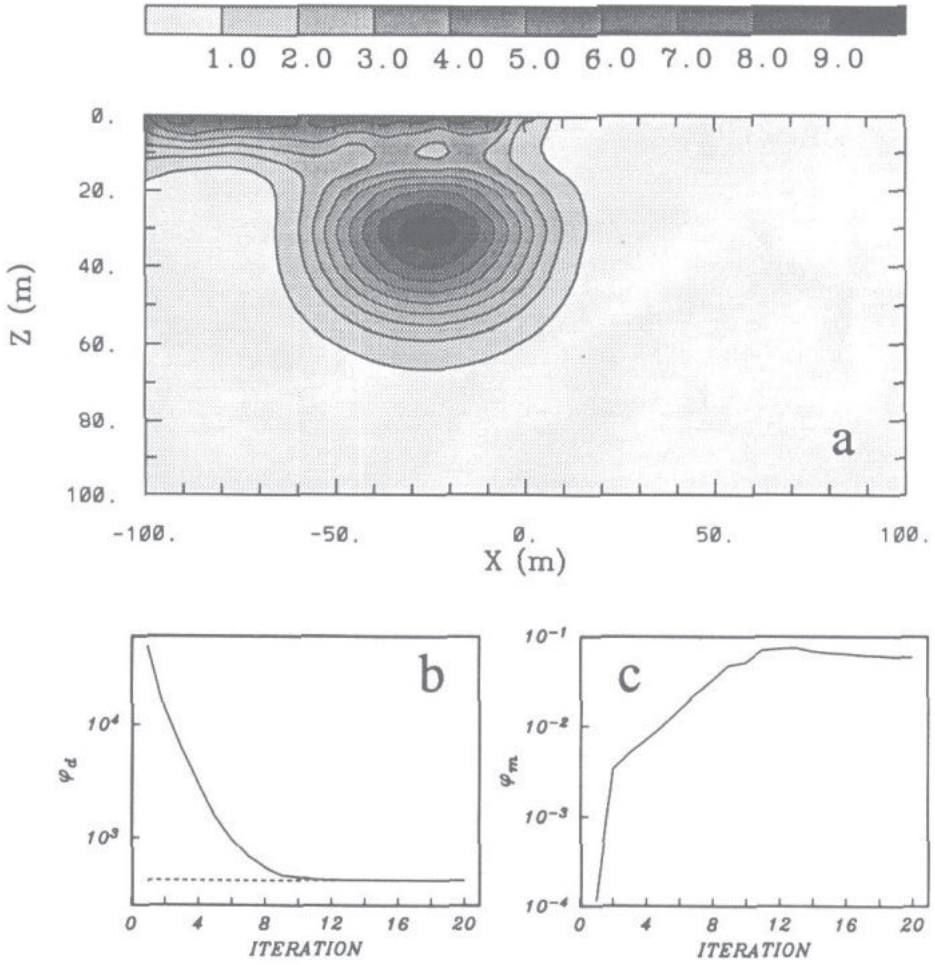

Figure 8. The constructed model when positivity is invoked using the two-stage mapping with $p_h = -6.9$ and $p_l = -3.0$ is shown in (*a*). The model has fewer artifacts and the chargeable body at depth is better delineated than in the inversion without positivity shown in figure 3.

result from that obtained in figure 3. The structure outside the regions of chargeability is reduced and the amplitude and confinement region of the chargeable block are in better agreement with the true model in figure 2. The convergence curves are somewhat different from their counterparts in figure 3 but the overall rate of convergence and number of computations are comparable to that obtained in the inversion without positivity.

The convergence characteristics of the subspace inversion are dependent upon the mapping parameters p_l and p_h. Eighty four inversions were carried out with different values of the parameters and the results are presented in figure 9. The contour plots indicate a substantial corridor for choices of p_h and p_l where good convergence is achieved. Our choice of $p_h = -6.9$ and $p_l = -3.0$ was satisfactory but not optimum.

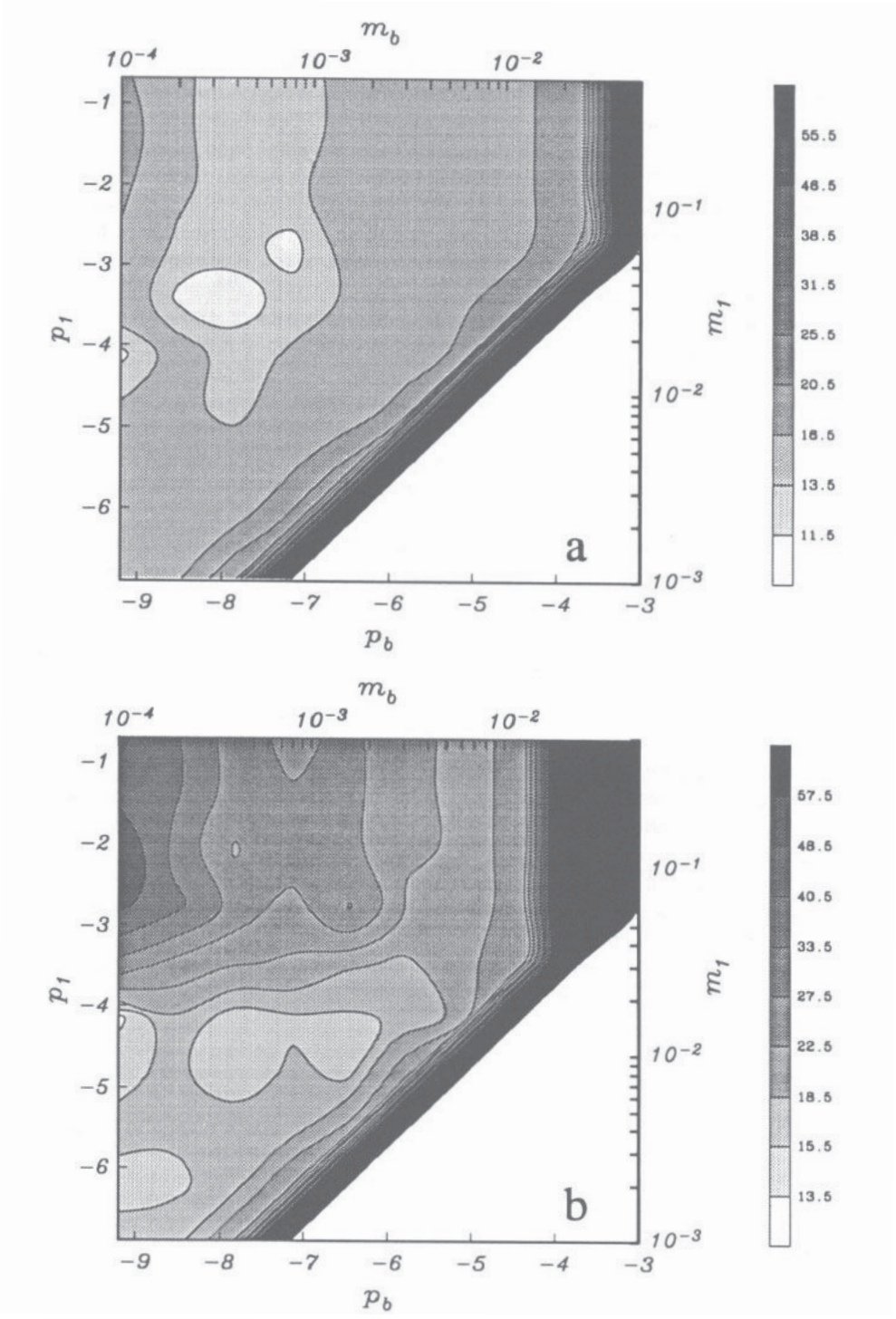

Figure 9. The effect of choice of parameters for the two-segment mapping. The number of iterations required to find a model which misfits the data to a final target value of 420 is presented in panel (*a*). The number of iterations required to achieve the asymptotic value of the model objective function is in panel (*b*). The axes on the bottom and left indicate the value of parameters p_b and p_1, respectively, while the axes on the top and right indicate the

7. Discussion

The subspace algorithm is efficient, flexible and robust. The efficiency is evaluated by the computational requirements per iteration which is approximately qNM as M becomes large. This is equivalent to q forward modellings. The final computation count depends upon the number of iterations required for convergence, and in our applications this is typically 10–25. The flexibility of the algorithm lies in the ease with which arbitrary model objective functions can be minimized, the ability for the user to incorporate positivity in the solution, the ability for the user to incorporate special search vectors in the inversion, and the ability to restart the inversion from any output result. The robustness of the algorithm is primarily attributable to forming search vectors from segmented data misfit and model objective functions. Our success in using the subspace method to solve linear inverse problems is perhaps related to the inherent non-uniqueness of the solution, the fact that the kernels (rows of $\mathbf{G}$) are generally smooth, and our desire to generate a 'simple' distribution of the physical parameter so that the result is interpretable. Basis functions which are formed from steepest descent vectors of the data misfit and model objective function are reasonably smooth and therefore may be particularly good choices for generating our desired models. Despite the endless possibilities for specifying the search vectors, the basic strategy outlined in this paper seems to work well. We use n_d vectors grouped according to misfit, one vector for each component of ϕ_m and a constant vector. The final number of iterations to achieve convergence is not greatly altered by how the data equations are grouped nor by their number once q is sufficiently large. Although the choice for Lagrange multiplier depends upon user constants β_1, β_2, β_3 there is a considerable range for which the algorithm works sensibly. The same is true with the selection of the constants p_h and p_l for the two-stage mapping to invoke positivity. A fundamental aspect of our formulation of the inverse problem is that we desire to construct a model that misfits the data to a predetermined amount. Since the appropriate Lagrange multiplier is found as the inversion continues, this obviates the need the carry out additional trial and error inversions. In the event that an unreasonably small misfit is sought, the algorithm has the characteristic of plateauing toward a minimum misfit. This situation can be recognized in practise, the target misfit can be readjusted and the inversion restarted.

The numerical examples presented here were relegated to a problem of moderate size having 1296 cells and 420 data but in a separate study of inverting magnetic data to recover a 3D distribution of magnetic susceptibility we have routinely used 1000 data to recover estimates of 40 000 model parameters. The algorithm was equally successful and this provides optimism for its utility in solving large-scale inverse problems.

References

Fink J B, McAlister E O, Sternberg B K, Wieduwilt W G and Ward S H (eds) 1990 Induced polarization: applications and case histories. *Investigations in Geophysics No. 4 for Soc. Expl. Geophys.*

Gill P E, Murray W and Wright M H 1981 *Practical Optimization* (New York: Academic) 401 pp

Golub G H and Van Loan C F 1989 *Matrix Computations* 2nd edn (Baltimore, MD: Johns Hopkins University Press)

Hestenes M R 1980 *Conjugate Direction Methods in Optimization* (New York: Springer)

Kennett B L N and Williamson P R 1988 Subspace methods for large-scale nonlinear inversion *Mathematical Geophysics: a Survey of Recent Developments in Seismology and Geodynamics* ed N J Vlaar, G Nolet, M J R Wortel and S A P L Cloetingth (Dordrecht: Reidel)

Kennett B L N, Sambridge M S and Williamson P R 1988 Subspace methods for large inverse problems with multiple parameter classes *Geophys. J.* **94** 237–47

Oldenburg D W, McGillivray P R and Ellis R G 1992 Generalized subspace methods for large scale inverse problems *Geophys. J. Int.* **114** 12–20

Paige C C and Saunders M A 1982 LSQR: An algorithm for sparse linear equations and sparse least squares *ACM Trans. Math. Software* **8** 43–71

Press W H, Teukolsky S A, Vetterling W T and Flannery B P 1992 *Numerical Recipes in Fortran* 2nd edn (Cambridge: Cambridge University Press)

Skilling J and Bryan R K 1984 Maximum entropy image reconstruction: general algorithm *Mon. Not. R. Astron. Soc.* **211** 111–24

Sumner J S 1976 *Principles of Induced Polarization for Geophysical Exploration* (Amsterdam: Elsevier)

van der Sluis A and van der Vorst H A 1987 Large sparse systems *Seismic Tomography with Applications in Global Seismology and Exploration Geophysics* ed G Nolet (Dordrecht: Reidel)