Geophys. J. R. astr. Soc. (1975) **42**, 957–976

Stable Iterative Methods for the Inversion of Geophysical Data

D. L. B. Jupp and K. Vozoff

(Received 1974 September 26)

Summary

Interpretation of earth electrical measurements can often be assisted by inversion, which is a non-linear model-fitting problem in these cases. Iterative methods are normally used, and the solution is defined by ' best fit ' in the sense of generalized least-squares.

The inverse problems we describe are ill-posed. That is, small changes in the data can lead to large changes in both the solution and in the iterative process that finds the solution. Through an analysis of the problem, based on local linearization, we define a class of methods that stabilize the iteration, and provide a robust solution. These methods are seen as generalizations of the well-known Singular Value Truncation and Marquardt Methods of iterative inversion.

Here, and in a companion paper, we give examples illustrating the successful application of the method to ill-posed problems relating to the resistivity of the Earth.

1. Introduction

In this paper we present an analysis of the solution to a number of geophysical inverse problems. We also provide a reference for the companion paper (Joint Inversion of Geophysical Data, Vozoff & Jupp 1975), where the results are applied to some specific examples.

Solutions to geophysical inverse problems are generally non-unique (Backus & Gilbert 1967, 1968, 1970), and it is usual to reduce the non-uniqueness by restricting the complexity of the Earth models. The mathematical problem that arises is commonly ill-posed (unstable) in the sense that small changes in the data lead to large changes in the solution. The solution methods must take careful account of this inherent problem.

In the companion paper, and the example given in Section 3 we have data in the form of apparent resistivity measurements for both magnetotelluric (MT), and Direct Current (DC) survey methods. The restricted class of earth models consists of horizontally layered, isotropic media, with constant resistivity in each layer. The simplified inverse problem is, in this case, to find the layer resistivities and thicknesses that best fit the observed data.

The analysis of the problem is not, however, restricted to layered models, but applies to any geophysical inverse problem in which the partial derivatives of the (predicted) data with respect to the (unknown) model parameters can be calculated.

Extending the ideas of Madden (1972), we classify the model parameters as Important, Unimportant, and Irrelevant. The classification is based on the Singular Value Decomposition (Lanczos 1958) of the Jacobian matrix, which also provides an analysis of sensitivity of the solution, and stability of the solution method.

The sensitivity and stability are analysed directly, without a statistical error hypothesis, as the structure of the error is rarely known in practice. For example, model inadequacy is normally a significant, and unknown component of the error, which is not statistical in nature. If error statistics are known, then they can be easily incorporated as extra information in our analysis. The statistical approach to these problems is well developed in Jackson (1972), and its relation to the direct perturbation approach that we adopt, is well described in Faddeev & Faddeeva (1969).

2. An analysis of the inverse problem, and its solution

2.1 *General notation*

The M data values $d_1, d_2, \ldots, d_M$, corresponding to M sample points, or instrument readings, are written as the vector

$$\mathbf{d} = (d_1, d_2, \ldots, d_M)^T.$$

In our examples, and in the companion paper, for the DC data d_i is the apparent resistivity at the i'th array spacing; for the MT data d_i is the apparent resistivity or phase at the i'th period.

The restricted earth models are determined by N free parameters, which we write as the vector,

$$\mathbf{x} = (x_1, x_2, \ldots, x_N)^T.$$

In our examples, these are logarithms of the layer thicknesses and resistivities for reasons given in Section (3.1).

The forward problem generates a set of model data for each setting of $\mathbf{x}$. This is denoted as a vector function by

$$\mathbf{g}(\mathbf{x}) = (g_1(\mathbf{x}), g_2(\mathbf{x}), \ldots, g_M(\mathbf{x}))^T.$$

Here, $g_i(\mathbf{x})$ is the value predicted by the model, and corresponds to the observation d_i.

The inverse problem determines values of $\mathbf{x}$ such that $\mathbf{g}(\mathbf{x})$ matches $\mathbf{d}$ in some sense, which in this paper is the minimum of the familiar Root Mean Squared relative (RMS) error between model and data,

$$F(\mathbf{x}) = \left\{ \frac{1}{M} \sum_{i=1}^{M} \frac{(d_i - g_i)^2}{d_i^2} \right\}^{\frac{1}{2}}. \tag{2.1.1}$$

A wide range of possible error measures can be encompassed by the notation

$$F(\mathbf{x}) = \|\boldsymbol{\varepsilon}\|_W = (\boldsymbol{\varepsilon}^T W \boldsymbol{\varepsilon})^{\frac{1}{2}} \tag{2.1.2}$$

where

$$\boldsymbol{\varepsilon} = \mathbf{d} - \mathbf{g}$$

and W is a positive (semi) definite matrix. W is usually diagonal, as it is for the RMS error, with

$$w_{ii} = (1/Md_i^2) \quad \text{for} \quad i = \overline{1, M}$$

2.2 *Inversion by iteration*

Iterative methods are common tools for practical inversion of geophysical data.

Some special problems may be solved directly (for example, McMechan & Wiggins 1972) but they are the exception. The iterative method successively improves a current model until the error measure is small and the parameters are stable with respect to reasonable changes in the model.

When the partial derivatives of the model data with respect to their parameters can be obtained accurately and relatively easily, the Gauss method (see Kowalik & Osborne 1968) is an attractive iterative inverse method. To derive it we expand $\mathbf{g}(\mathbf{x})$ about $\mathbf{x}$ in a Taylor expansion

$$\mathbf{g}(\mathbf{x}+\delta\mathbf{x}) = \mathbf{g}(\mathbf{x}) + J\delta\mathbf{x} + \mathbf{R}(\mathbf{g}, \delta\mathbf{x}) \tag{2.2.1}$$

where

$$J = \left| \frac{\partial g_i}{\partial x_j} \right|_{\substack{i=\overline{1,M} \\ j=\overline{1,N}}}$$

is the Jacobian matrix of the vector function $\mathbf{g}(\mathbf{x})$. The remainder term $\mathbf{R}$ depends on $\mathbf{g}$, and we assume the reasonable condition (on $\mathbf{g}$) that

$$\|\mathbf{R}\| = 0(\|\delta\mathbf{x}\|^2).$$

If $\mathbf{g}(\mathbf{x})$ is a linear function, then $\mathbf{R}(\mathbf{g}, \mathbf{x})$ is zero, and

$$\mathbf{g}(\mathbf{x}+\delta\mathbf{x}) = \mathbf{g}(\mathbf{x}) + J\delta\mathbf{x}, \text{ exactly.}$$

The original Gauss method iteratively changes the current $\mathbf{x}$ by an amount $\delta\mathbf{x}$ calculated by solving the linear least-squares problem

$$\textit{minimize } \|\boldsymbol{\varepsilon} - J\delta\mathbf{x}\|_W,$$

where

$$\boldsymbol{\varepsilon} = \mathbf{d} - \mathbf{g}.$$

When $J^T W J$ is non-singular we find that

$$\delta\mathbf{x} = (J^T W J)^{-1} J W \boldsymbol{\varepsilon}.$$

For convenience, we have dropped the W matrix in the following sections. That this causes no loss of generality has been demonstrated by Jackson (1972, page 103).

2.3. *The Jacobian and irrelevant parameters*

In model fitting problems, J is called the ' sensitivity matrix ' for the model with respect to the parameters, since an element of J

$$J_{ij} = \frac{\partial g_i}{\partial x_j} \quad (1 \leqslant i \leqslant M,\ 1 \leqslant j \leqslant N)$$

measures the variation in the i'th model value with respect to variations in x_j. However it is important to determine the amount of *independent* information J contains.

Example 1. If the j'th column of J is exactly zero, then the corresponding parameter x_j would have no influence on the model values at any of the observation points. In this case we call x_j an Irrelevant parameter.

Example 2. If we have 100 data values which are replicates at a single site, then there is only one piece of information, even though its value is precisely known. In this case most of the 100 data values are Redundant.

Both of these deficiencies are measured by P, the rank of J. P is the maximum

number of linearly independent rows or columns, and is always $\leqslant$ minimum (M, N). When $P < \min(M, N)$, the original Gauss method can be easily extended (Wiggins 1972) by using the ' generalized inverse ' of J. Introduced by Moore (1920), and Penrose (1955), the generalized inverse of a matrix J is the unique matrix J^+ satisfying the four conditions

(i) $J J^+ J = J$

(ii) $J^+ J J^+ = J^+$

(iii) $(J^+ J)^T = J^+ J$

(iv) $(J J^+)^T = J J^+$.

A more revealing, and equivalent, operational definition of J^+ is its action of producing $\delta\mathbf{x}$ from $\boldsymbol{\varepsilon}$.

This unique $\delta\mathbf{x}$ satisfies the two minimum conditions

(i) $\delta\mathbf{x}$ minimizes $\|\boldsymbol{\varepsilon} - J\delta\mathbf{x}\|$

(ii) $\delta\mathbf{x}$ minimizes $\|\delta\mathbf{x}\|$ among solutions to (i).

The operation can be written in the form of matrix multiplication as $\delta\mathbf{x} = J^+ \boldsymbol{\varepsilon}$, where the matrix J^+ is a ' sensitivity matrix ' for the parameters corresponding to (small) changes in the data.

The generalized Gauss method estimates the correction step as

$$\delta\mathbf{x} = J^+ \boldsymbol{\varepsilon}.$$

If J has rank N, (all parameters are relevant) then we can show that $J^+ = (J^T J)^{-1} J^T$ and the original and generalized forms of the Gauss iteration are the same.

The rank of J is difficult to determine precisely when it is $< \min(M, N)$. For example, if we look for a zero pivot in Gauss elimination of $J^T J$, this must be determined by ' machine precision ' zero. In practice we do not need to determine the exact rank of J if we use the extended methods described in Section (2.7).

2.4. *The singular values of J, and the Unimportant parameters*

Lanczos (1958) introduced a decomposition which provides a spectral expansion for an arbitrary matrix.

Let J be any $M \times N$ matrix with rank $P \leqslant \min(M, N)$ then

$$J = USV^T = \sum_{i=1}^{N} s_i \mathbf{u}_i \mathbf{v}_i^T$$

where U is $M \times N$, S is $N \times N$, V is $N \times N$, with

$$U^T U = I_M, \quad \text{and} \quad V^T V = I_N.$$

In the summation, $\mathbf{u}_i$ is a column of U, $\mathbf{v}_i$ is a column of V, and s_i is an element of the (diagonal) matrix S. Each s_i is the non-negative square root of an eigenvalue of $J^T J$, and we may assume that

$$s_1 \geqslant s_2 \geqslant \ldots \geqslant s_N \geqslant 0.$$

The s_i have been called the ' singular values ' of J, (Forsythe & Moler 1967) and if J has rank P, then

$$s_P > 0 \quad \text{and} \quad s_{P+1} = s_{P+2} = \ldots = s_N = 0.$$

The generalized inverse of J may be simply defined in terms of the Singular Value Decomposition (' SVD ') (Lanczos 1958) as

$$J^+ = V S^+ U^T$$

where S^+ is the diagonal matrix with

$$s_i{}^+ = \begin{cases} \dfrac{1}{s_i} \text{ if } s_i > 0 \\ 0 \text{ if } s_i = 0. \end{cases}$$

By direct substitution, J^+ satisfies the four ' Penrose ' conditions ((i)–(iv)) of Section (2.3).

Since $s_i = 0$ for $i > P$, (J has rank P)

$$J = \sum_{i=1}^{N} s_i \mathbf{u}_i \mathbf{v}_i{}^T = \sum_{i=1}^{P} s_i \mathbf{u}_i \mathbf{v}_i{}^T,$$

and

$$J^+ = \sum_{i=1}^{P} \frac{1}{s_i} \mathbf{v}_i \mathbf{u}_i{}^T.$$

For this reason, the columns of U and V corresponding to zero singular values are often omitted from the definition, and correspond to the Irrelevant parameters, and Redundant information, of Section 2.3.

To be precise, from the Taylor expansion (2.2.1) of Section 2.2,

$$\delta \mathbf{g} \simeq J \delta \mathbf{x}$$

represents the variation in the model data with respect to the (small) change $\delta \mathbf{x}$ in parameters.

Let $k_i = s_i/s_1$

$$K = \text{diag}\,(k_1, k_2, \ldots, k_N)$$

and

$$\delta \mathbf{p} = s_1 V^T \delta \mathbf{x},$$

then,

$$\begin{aligned} \delta \mathbf{g} \simeq J \delta \mathbf{x} &= U S V^T \delta \mathbf{x} \\ &= U K \delta \mathbf{p} \\ &= \tilde{J} \delta \mathbf{p}. \end{aligned}$$

The change from $\delta \mathbf{x}$ to $\delta \mathbf{p}$ is a rotation of axes in the parameter space, and provides the basis for the classification of parameters. $\tilde{J}$ is the Jacobian, or sensitivity, matrix for the transformed parameters, and

$$\delta \mathbf{g} \simeq \tilde{J} \delta \mathbf{p} = \sum_{i=1}^{N} k_i \delta p_i \mathbf{u}_i = \sum_{i=1}^{P} k_i \delta p_i \mathbf{u}_i. \tag{2.4.1}$$

The size of the variation is

$$\|\delta \mathbf{g}\|^2 \simeq \sum_{i=1}^{N} k_i{}^2 |\delta p_i|^2 = \sum_{i=1}^{P} k_i{}^2 |\delta p_i|^2. \tag{2.4.2}$$

The sum involves only the δp_i corresponding to non-zero singular values of J, and $\delta p_{P+1}, \delta p_{P+2} \ldots \delta p_N$ (which correspond to zero columns of $\tilde{J}$) will be called Irrelevant.

Parameters corresponding to small (relative to s_1) singular values will be classed

as Unimportant, since variations in the corresponding δp_i will have little effect on the model. For these parameters, the contribution to $\delta \mathbf{g}$ is basically $k_i \delta p_i \mathbf{u}_i$, which is small, and the contribution to $\|\delta \mathbf{g}\|^2$ is basically $k_i^2 |\delta p_i|^2$.

The point where k_i becomes ' too small ' is called the threshold, and divides the parameters into the Important and Unimportant classes. A general idea of the value of the threshold can be obtained from the equations (2.4.1) and (2.4.2). We retain the threshold as a variable since its value is used to stabilize the iteration and to mollify the effects of gross errors. Parameters with k_i near the threshold are called Threshold parameters, and will often be the most interesting class. In practice, while the threshold can be well defined, the Irrelevant parameters will correspond to k_i whose values are zero, within machine precision.

2.5. *Ill-posed problems, and unstable iterations*

The inverse problems in geophysics exert their ill-posedness through Irrelevant parameters (zero singular values of J), and Unimportant parameters (small singular values of J). There is generally a limit to the number of well defined, or important parameters that may be resolved from the data.

Suppose that we are at a solution of the original problem. That is, we are at a (strict) minimum of $F(\mathbf{x})$. The necessary condition for a minimum is that the gradient of $F(\mathbf{x})$ (as defined by (2.1.2) with W suppressed as described in Section 2.2).

$$\nabla F(\mathbf{x}) = -\frac{1}{\|\boldsymbol{\varepsilon}\|} J^T \boldsymbol{\varepsilon} = \mathbf{0}.$$

That is, either $\|\boldsymbol{\varepsilon}\| = 0$, or else $J^T \boldsymbol{\varepsilon} = \mathbf{0}$
Since,

$$J^T \boldsymbol{\varepsilon} = V S U^T \boldsymbol{\varepsilon} = V S \mathbf{r}$$

(where $\mathbf{r} = U^T \boldsymbol{\varepsilon}$), and V is $N \times N$ and orthogonal, the condition for a minimum is

$$S\mathbf{r} = \mathbf{0} \quad \text{or} \quad s_i r_i = 0 \text{ for } i = \overline{1, N}.$$

Because

$$s_i = 0 \quad \text{for} \quad i > P,$$

the minimum condition reduces to

$$\mathbf{u}_i^T \boldsymbol{\varepsilon} = r_i = 0 \quad \text{for} \quad i = \overline{1, P}. \tag{2.5.1}$$

That is, the error, $\boldsymbol{\varepsilon}$, is orthogonal to the directions $\mathbf{u}_i$, $(i = 1, P)$, which are the directions in which the relevant parameters may change the model.

If the original data are perturbed by a small amount $\delta \mathbf{d}$, then we can estimate the perturbation in $\mathbf{x}$ by the Gauss correction step as

$$\delta \mathbf{x} = J^+ \delta \mathbf{d}.$$

This estimate assumes the correction step is valid near to a solution, and may be used to provide linearized ' error bounds ' or parameter sensitivities with respect to known perturbations in the data.

Suppose that $\delta \mathbf{d}$ is a perturbation of the data satisfying

$$\|\delta \mathbf{d}\| \leqslant q.$$

Then since the columns of U are orthonormal, we know that

$$\|U^T \delta \mathbf{d}\| \leqslant \|\delta \mathbf{d}\| \leqslant q.$$

If we put $\delta \mathbf{r} = U^T \delta \mathbf{d}$, and use the SVD and our previous definitions of K and $\delta \mathbf{p}$, we find

$$\delta \mathbf{p} = K^+ \delta \mathbf{r}$$

or

$$\delta p_i = \frac{\delta r_i}{k_i} \text{ for } i = \overline{1, P}.$$

Since

$$|\delta r_i| \leqslant q$$

we obtain

$$|\delta p_i| \leqslant \frac{q}{k_i} \text{ for } i = \overline{1, P}. \tag{2.5.2}$$

The inequalities (2.5.2) provide estimates for the linear variation of the transformed parameters with respect to the data perturbation $\delta\mathbf{d}$. It is clear from these estimates that the Unimportant parameters, which contribute little to the variation of the model, can undergo very large changes for small changes in the data. Precisely, they contain the ill-posed nature of the problem.

In the same way as they express indeterminacy in the solution, the small singular values of J can cause instability in the iterative method. At any current model we may regard $\mathbf{x}$ as the exact solution for some data from which the actual data are perturbed. If k_i is small (but not zero) the generalized Gauss correction step

$$\delta p_i = \frac{r_i}{k_i}$$

is large. This may distort the model, and cause the correction step to exceed the limits in which the linear approximation is valid. For example an unseen shallow layer could become thin and highly conductive and prevent the deeper, more Important layers from properly interpreting the data.

2.6. *Two current inversion methods*

The more successful current methods of the Gauss type damp or eliminate variations in the Unimportant parameters and use some form of threshold control. One such method is the Singular Value Truncation technique (Madden 1972; Osborne 1972) which has a specified threshold μ. (Truncation of small singular values has been successfully applied to linear inverse problems by Hanson (1971) and Varah (1973), whose results apply locally in the non-linear case.)

If $k_l \geqslant \mu$ and $k_{l+1} < \mu$, then there are l important parameters, and only variations in these are accepted.

That is,

$$\delta\mathbf{x} = (1/s_1)\, V\delta\mathbf{p}^+$$

where

$$\begin{aligned} \delta p_i{}^+ &= \delta p_i \quad \text{if} \quad k_i \geqslant \mu \\ &= 0 \quad \text{if} \quad k_i < \mu. \end{aligned}$$

At a ' solution ' of the problem found by the Truncation strategy, we find that, instead of the full minimum condition (2.5.1) that

$$\mathbf{u}_i{}^T \boldsymbol{\varepsilon} = 0 \quad \text{for} \quad i = \overline{1, l}, \tag{2.6.1}$$

but $\mathbf{u}_i{}^T \boldsymbol{\varepsilon}$ may be non-zero for $i > l$. This solution is not an exact minimum of $F(\mathbf{x})$, but it has the property that only changes in Unimportant parameters will cause further reduction in the error.

For this method, the class of models reachable from the initial point is restricted by the elimination of search directions. The restricted class depends on the starting point, which must be well chosen for good results.

Another popular method for ill-posed inverse problems is the Marquardt method (Marquardt 1963; Morrison 1960).

The Marquardt correction step is usually defined operationally as the solution of the ' damped least squares ' problem (Jennings & Osborne 1970):

$$\textit{minimize } \|\boldsymbol{\varepsilon} - J\delta\mathbf{x}\|^2 + \nu^2\|\delta\mathbf{x}\|^2.$$

Here, ν^2 is the ' Marquardt Parameter ', which is used to control the size and direction of the step $\delta\mathbf{x}$ during the iteration (Marquardt 1963). Without further modification, the Marquardt method finds a full minimum of $F(\mathbf{x})$ in a stable, or damped, fashion. In our previous terms, the method changes Unimportant parameters towards the end of its iteration so as to reduce the residual error to a strict minimum.

2.7. *Damped approximate inverses*

The two methods of Section 2.6 are particular cases of a class of methods, based on modifications of J^+, which control the variations in the Unimportant parameters.

Let $B^+ = VTS^+U^T$ where T is an $N \times N$ diagonal matrix with entries

$$0 \leqslant t_i \leqslant 1, \quad \text{for} \quad i = \overline{1, P}, \quad \text{and} \quad t_i = 0 \quad \text{for} \quad i > P.$$

The t_i are called ' damping factors '. Various choices of T define known algorithms, and suggest interesting extensions based on the correction step

$$\delta\mathbf{x} = B^+\boldsymbol{\varepsilon} = \sum_{i=1}^{P} t_i \frac{r_i}{s_i}\mathbf{v}_i \tag{2.7.1}$$

or

$$\delta p_i = \begin{cases} \dfrac{t_i r_i}{k_i} & i = \overline{1, P} \\ 0 & i > P. \end{cases}$$

Both the truncation method, and the Marquardt method are defined by an appropriate choice of T. With some algebra, the equivalent choice of T for the Marquardt method is

$$t_i = \begin{cases} \dfrac{s_i^2}{s_i^2 + \nu^2} & \text{for} \quad i = \overline{1, P} \\ 0 & i > P \end{cases}$$

or

$$t_i = \begin{cases} \dfrac{k_i^2}{k_i^2 + \mu^2} & \text{for} \quad i = \overline{1, P} \\ 0 & i > P \end{cases}$$

where $\mu = \nu/s_1$ is a relative threshold.

An interesting class of methods, which includes both the Marquardt and Truncation estimates for B^+, is defined by choosing $T = T^{(N)}$ with entries,

$$t_i^{(N)} = \begin{cases} \dfrac{k_i^{2N}}{k_i^{2N} + \mu^{2N}} & \text{for} \quad i = \overline{1, P} \\ 0 & i > P \end{cases}$$

$T^{(1)}$ corresponds to the Marquardt method, and as $N \to \infty$

$$t_i \to \begin{cases} 1 & \text{if} \quad \mu < k_i \\ \frac{1}{2} & \text{if} \quad \mu = k_i \\ 0 & \text{if} \quad \mu > k_i \end{cases}$$

$T^{(\infty)}$ clearly defines the Truncation method with threshold μ. In Fig. 1 we have drawn the graphs of the function

$$f_N(t) = \frac{t^{2N}}{t^{2N}+1} \quad \text{for} \quad N = 1, 2, 3, \infty.$$

The curves demonstrate how the sequence $T^{(N)}$ provides a tapered damping when $N = 1$ which gradually changes to a sharp cut-off at 1 as $N \to \infty$.

An important property of these methods is that the size of the correction step (2.7.1) is bounded and controlled by μ. That is

$$\begin{aligned} \|\delta \mathbf{x}\| &\leqslant \|B^+\| \, \|\boldsymbol{\varepsilon}\| \\ &= C\|\boldsymbol{\varepsilon}\| \end{aligned}$$

where

$$C = \max_{i=\overline{1, P}} \left| \frac{t_i}{s_i} \right| \leqslant \frac{1}{s_1 \mu} \quad \text{for} \quad T^{(N)}.$$

Moreover each of the corresponding ' approximate Inverses '

$$B^+ = V T^{(N)} S^+ U^T \quad N = 1, 2, \ldots, \infty$$

has the important property that

$$\lim_{\mu \to 0} B^+ = J^+.$$

When all parameters are Important, the corresponding numerical methods should behave like the generalized Gauss iterative method.

2.8. *The practical algorithm*

In practice, the choice of threshold, μ, and damping factors, T, depends on the problem considered.

A guide in the choice of threshold, and T, is the way in which the singular values of J cluster. For the resistivity problems studied, there is often a distinct separation of parameters into the Important, and Unimportant, divisions. The corresponding ' cliff ' in the singular values provides a natural threshold. (When the data error level is high, even the parameters above a natural threshold may be unstable and $\bar{\mu}$ must be raised as described in Section 2.9.) In these cases, the truncation and approximate truncation ($T^{(N)}$ for $N > 2$) methods have been successful. For the problems considered in the companion paper we used the T matrix

$$T^{(2)} = (t_i^{(2)})$$

where

$$t_i^{(2)} = \begin{cases} k_i^4/(k_i^4 + \mu^4) & k_i \geqslant \bar{\mu}^2 \\ 0 & k_i < \bar{\mu}^2 \end{cases}$$

and $\mu \geqslant \bar{\mu}$, the actual problem threshold.

The variable μ alters the local threshold above the lower limit $\bar{\mu}$ to control step

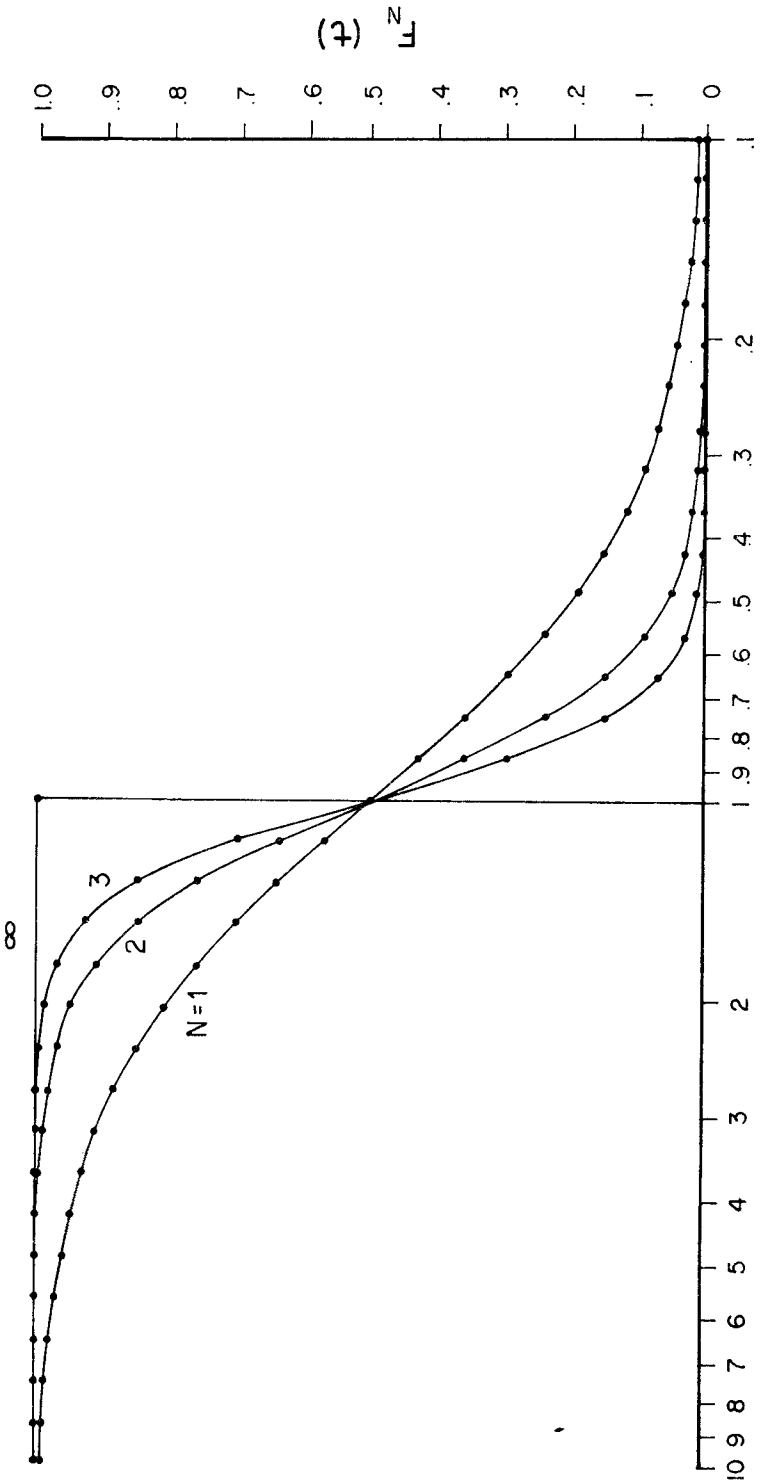

FIG. 1. Graph of $f_N(t) = t^{2N}/(1+t^{2N})$ for $N = 1,2,3$, and ∞, showing the attenuation provided by the damping factors.

size and direction, and eventually decreases to $\bar{\mu}$ if the steps are successful in reducing the error.

The overall strategy is similar to the Marquardt method, as implemented by Jennings & Osborne (1970). However, instead of orthogonal factorization, we have used the Singular Value Decomposition as published by Golub & Reinsch (1970) to construct B^+. The SVD algorithm uses Householder transformations to reduce J to upper bi-diagonal form, and then applies a variant of the QR method due to Francis (1961) to find the singular values of the bi-diagonal matrix. We have found the extra computing time involved only marginal as compared with the overheads of the rest of the program, and more than compensated by the extra information provided.

We chose the Marquardt methods for their control of the iteration from the initial point to the neighbourhood of the solution. However, instead of the full minimum condition (2.5.1) we wish to terminate the algorithm at the Truncation solution (2.6.1).

Therefore in order to combine the desirable features of the two methods, we need to define numerical convergence in such a way that they have the same terminal behaviour. A satisfactory choice is the measure

$$\delta L^2 = \sum_{i=1}^{P} t_i r_i^2$$

which, for the generalized Gauss, and Truncation methods is the predicted decrease in the residual error.

If $\delta L^2 \leqslant \text{tol}^2 \|\boldsymbol{\varepsilon}\|^2$ defines numerical convergence, where tol is a (small) relative tolerance, then

$$|r_i|^2 \leqslant (\text{tol}^2/t_i) \|\boldsymbol{\varepsilon}\|^2 \quad \text{for} \quad i = \overline{1, P}.$$

The Truncation method requires strict convergence for $i \leqslant l$, and makes no requirement on the error in directions corresponding to Unimportant parameters. The Marquardt method is not as strict for $|r_i|$ corresponding to a parameter just above the threshold, but requires $|r_i|$ corresponding to a parameter just below the threshold to be bounded.

If q is the residual error in the data at numerical convergence and we assume the data may be perturbed by amounts $\delta\mathbf{d}$ satisfying

$$\|\delta\mathbf{d}\| \leqslant q$$

then, modifying the results of Section 2.5 we find

$$|\delta p_i| \leqslant (t_i q/k_i) \quad \text{for} \quad i = \overline{1, P}. \tag{2.8.1}$$

These represent damped estimates of linear variation in the transformed parameters and it is important to notice that both well resolved, and poorly resolved, parameters have small variation.

To relate these bounds to the original parameters, we let

$$\delta\mathbf{x}^+ = VV^T\delta\mathbf{x} = (1/s_1)\, V\delta\mathbf{p}$$

be a ' generalized inverse ' estimate for $\delta\mathbf{x}$ corresponding to $\delta\mathbf{p}$.

If we put $e_i = 1/k_i$, and $\mathbf{e} = (e_1, e_2, \ldots, e_P)^T$ then the estimates

$$\pm(q/s_1)\,|V|\mathbf{e} \quad \text{where} \quad |V| = (|v_{ij}|)_{\substack{i=\overline{1,N} \\ j=\overline{1,N}}} \tag{2.8.2}$$

are coarse ' error bounds ' for the original parameters. These depend on the choice of T, and describe the ' useful ' variation in the parameters of the final model.

2.9 *Practical problems*

To derive the results of the previous sections, we assumed that the model was well approximated locally by the linear estimate (2.2.1) and that the data, and parameter, perturbations were ' small '. Implicitly, we assumed that the general residual error level (at least at the solution) was small, and that we were always ' near ' to the solution.

In practice, large general (gross) errors, and large impulsive (outlying) errors, occur in the data. Moreover, the final model is usually unknown, so that the initial model of the iteration will rarely be ' near ' the solution. In each of these situations, the local linear estimates can be misleading, for the following reason. Following Section 2.1, we can write (the squared) general error measure as

$$F^2(\mathbf{x}) = \boldsymbol{\varepsilon}^T W \boldsymbol{\varepsilon} = \|\boldsymbol{\varepsilon}\|_W{}^2.$$

The Hessian matrix (matrix of second partial derivatives) of $F^2(\mathbf{x})$ is a measure of the curvature of its contours, and

$$H = \left[\frac{\partial^2 F^2(\mathbf{x})}{\partial x_i \partial x_j}\right]_{i,\, j=\overline{1,N}} = -2R^T W\boldsymbol{\varepsilon} + 2J^T WJ.$$

The non-linear term, $-2R^T W\boldsymbol{\varepsilon}$, depends on both the tensor R, where

$$[R_{ijk}] = \left[\frac{\partial^2 g_k}{\partial x_i \partial x_j}\right]_{\substack{i,j=\overline{1,N} \\ k=\overline{1,M}}}$$

and the residual error in the data, $\boldsymbol{\varepsilon}$. When J is slowly varying (which means $\|R\|$ is small), and the residual error $F(\mathbf{x})$ is small, then the Gauss and related methods are highly efficient. Even when $\mathbf{g}(\mathbf{x})$ is reasonably non-linear (in that R is significant), the methods are efficient when $F(\mathbf{x})$ is small. The efficiency of the methods is a result of the accurate estimate of the curvature of the contours of $F(\mathbf{x})$ by $J^T WJ$. When $F(\mathbf{x})$ is not small, such as at a point remote from the solution, or when there is general gross error in the data, we find that non-linearity is most apparent in the directions corresponding to Unimportant parameters. The methods of Section 2.7 that attenuate the correction steps in the directions of the Unimportant parameters therefore ' filter ' out the components of the step that are unstable, and poorly estimate the curvature of $F(\mathbf{x})$.

At the solution, the residual error is orthogonal to directions corresponding to Important parameters. However, if it is still large, the ' error bounds ' and sensitivity analysis are only locally useful. For variations of the data up to the size of the residual error in very noisy data, they are only estimates of local parameter sensitivity, and not ' error bounds ' in any valuable sense.

The effective strategy in these problems is to raise the threshold level so that only the basic features of the model will be resolved in the first inversion. The process corresponds to first finding a rough two or three layer model, in manual inversion. All of the methods of the Marquardt type, described in Section 2.7 should use a high initial threshold μ which reduces the problem threshold, $\bar{\mu}$, if the iteration is successful. When there is gross error, $\bar{\mu}$ itself must be raised. In order to ensure a stable and useful iteration when the expected error q is large, we find that $\bar{\mu}$ must be approximately proportional to q. When q is RMS relative error, a good initial choice is to take $\bar{\mu} = q$.

The final practical problem we consider is that of ' outliers '. These are bad data points, and have a quite different effect on the solution from that of general error.

If $\mathbf{e}_i$ is the unit vector, with 1 in the i'th place, and zeros elsewhere, then the estimates

$$\delta\mathbf{x}^{(i)} = B^+ \mathbf{e}_i,$$

and

$$\delta\mathbf{p}^{(i)} = TK^+ U^T \mathbf{e}_i$$

represent the (damped) response of the parameters to a unit impulsive data error. Clearly, $\delta\mathbf{x}^{(i)}$ is the i'th column of B^+ and $\delta\mathbf{p}^{(i)}$ is the i'th column of $TK^+ U^T$.

In the same way as general errors, an outlier has a more drastic effect on Unimportant parameters than on Important parameters. However, if the outlier has a large component in directions corresponding to a large singular value (a large component in the corresponding column of $TK^+ U^T$), then damping the small singular values cannot mollify the resulting perturbation of that Important parameter.

The possible effects due to specific outliers, and combinations of outliers, on the well-resolved components of the solution can be estimated from B^+, or $T K^+ U^T$. The effect of local, and regional changes in slope, and curvature, can be estimated by applying difference operations to the rows of B^+, or $T K^+ U^T$. However, in order to overcome some of these problems, when they are particularly bad, a different estimate of $\delta\mathbf{x}$, based on the work of Claerbout (1973), could be made.

3. The thin resistive layer as an example

3.1. *The model and data*

A classical problem for the MT and DC survey methods is to interpret a thin resistive layer from surface apparent resistivity data. To study the outcome of part 2, we took as a model the structure of Fig. 2, where a layer of resistivity 10 Ohm-m, and thickness 10 m is embedded at a depth of 100 m, in a half-space of constant resistivity 1 Ohm-m.

Two ' fictitious ' layers, with boundaries at depths of 20 m, and 160 m serve to illustrate the consequent sensitivity, and stability, problems we have discussed.

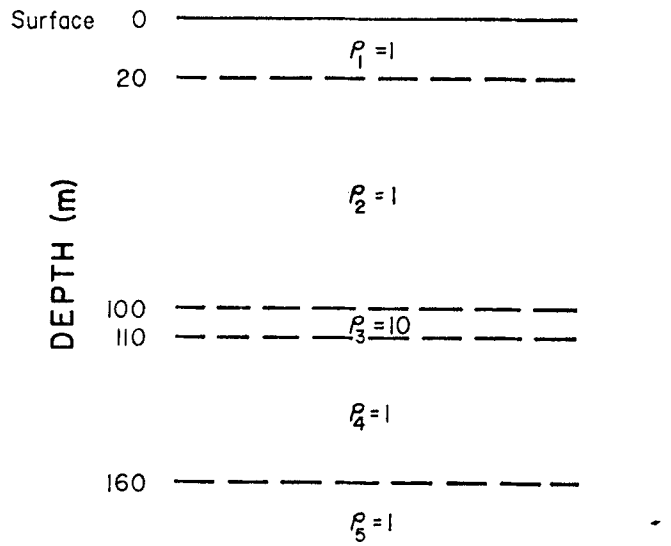

FIG. 2. The exact thin resistive layer model. Five layers are shown but two are not 'real'.

Table 1

Singular values, damping factors, and the V matrix for the exact model of Fig. 2.

Normalized singular values of Jacobian

0·100+01	0·834−00	0·285−00	0·128−00	0·105+00	0·399−01	0·341−02	0·104−07	0·790−09

Damping factors (1·000 per cent level)

0·100+01	0·100+01	0·100+01	0·100+01	0·100+01	0·996−00	0·133−01	0·0	0·0

Parameter space eigenvectors (V matrix)

	$\log q_1$	$\log q_2$	$\log q_3$	$\log q_4$	$\log q_5$	$\log q_6$	$\log q_7$	$\log q_8$	$\log q_9$
1. $\log \rho_1$	0·204	−0·392	0·815	−0·316	0·102	−0·172	0·011	0·000	0·000
2. $\log \rho_2$	0·490	−0·730	−0·313	0·264	−0·117	0·211	−0·013	0·000	0·000
3. $\log \rho_3$	0·049	−0·039	−0·220	−0·515	0·458	0·284	−0·626	0·000	0·000
4. $\log \rho_4$	0·103	−0·035	−0·261	−0·535	−0·646	−0·446	−0·132	0·000	0·000
5. $\log \rho_5$	0·835	0·544	0·065	0·044	0·037	0·004	0·001	0·000	0·000
6. $\log h_1$	−0·016	0·027	0·060	−0·034	−0·130	0·190	−0·005	0·970	0·003
7. $\log h_2$	−0·063	0·109	0·240	−0·138	−0·520	0·760	−0·019	−0·243	−0·001
8. $\log h_3$	0·060	−0·042	−0·235	−0·507	0·245	0·180	0·768	0·000	0·000
9. $\log h_4$	0·000	0·000	0·000	0·000	0·000	0·000	0·000	−0·003	1·000

The parameters $\mathbf{x}$ of Section 2.1 are the logarithms of layer resistivities, and thicknesses,

$$\mathbf{x} = (\log \rho_1, \log \rho_2, \ldots \log \rho_5, \log h_1, \ \ldots \log h_4)^T.$$

Taking logarithms of the layer parameters imposes the constraints

$$\begin{cases} \rho_i \geqslant 0 & i = \overline{1,5} \\ h_i \geqslant 0 & i = \overline{1,4} \end{cases}$$

in a natural way, makes the Jacobian free of dimensions of the parameters and makes the problem stable if any layer becomes very thin.

The forward problem, and computation of partial derivatives of the model data with respect to the x_i, are discussed in the companion paper. We used the forward mode to generate 10 DC (Schlumberger) and 16 MT apparent resistivities, sampled densely enough to contain most of the information the separate methods are able to provide.

These data were perturbed by generated ' Gaussian ' relative error to the three per cent level, and inverted to illustrate various aspects of the general problem.

The inverse problem we consider is the joint problem described in the companion paper. For the present we regard this as one problem containing both sets of data. For example, if J_1 and $\boldsymbol{\varepsilon}_1$ are the Jacobian, and error vector for the MT and J_2 and $\boldsymbol{\varepsilon}_2$ are the Jacobian and error vector for the DC problem, then the joint problem matches

$$\begin{bmatrix} W_1^{\frac{1}{2}} J_1 \\ W_2^{\frac{1}{2}} J_2 \end{bmatrix} \delta \mathbf{x} \simeq \begin{bmatrix} W_1^{\frac{1}{2}} \boldsymbol{\varepsilon}_1 \\ W_2^{\frac{1}{2}} \boldsymbol{\varepsilon}_2 \end{bmatrix}$$

where W_1 and W_2 are diagonal matrices corresponding to the mean relative error measure (2.1.1) of Section 2.1.

3.2. *Parameter classification and ' error bounds '*

Table 1 contains the normalized singular values (k_i, $i = \overline{1,9}$), the damping factors ($t_i^{(2)}$, $i = \overline{1,9}$) and the V matrix from the SVD of the weighted Jacobian of the joint exact model. Here, we have taken

$$t_i^{(2)} = \frac{k_i^4}{k_i^4 + \bar{\mu}^4}, \text{ with } \bar{\mu} = 0{\cdot}01 \text{ (1 per cent)}$$

Parameters with $k_i \geqslant 0{\cdot}01$ are called Important at ' the 1 per cent level ' and in Table 1 there are six Important parameters.

To interpret the nature of the Important, and Unimportant parameters, they must be related to the original layer parameters. The expression

$$\mathbf{p} = s_1 V^T \mathbf{x}$$

defines the (linear) relation between the original, and transformed, parameter spaces, and for our problem, we have

$$\frac{1}{s_1} p_j = \log q_j = \sum_{i=1}^{5} v_{ij} \log \rho_i + \sum_{i=6}^{9} v_{ij} \log h_{i-5}$$

or

$$q_j = \prod_{i=1}^{5} \rho_i^{v_{ij}} \prod_{i=6}^{9} h_{i-5}^{v_{ij}}$$

Table 2

Singular values, damping factors, and the V matrix for the final model of Fig. 3.

Normalized singular values of Jaocbian

0·100+01	0·808−00	0·264−00	0·127−00	0·991−01	0·350−01	0·375−02	0·905−03	0·383−03

Damping factors (1·000 per cent level)

0·100+01	0·100+01	0·100+01	0·100+01	0·100+01	0·993−00	0·194−01	0·671−04	0·215−05

Parameter space eigenvectors (V matrix)

	$\log q_1$	$\log q_2$	$\log q_3$	$\log q_4$	$\log q_5$	$\log q_6$	$\log q_7$	$\log q_8$	$\log q_9$
1. $\log \rho_1$	0·161	−0·430	0·778	−0·335	0·172	−0·207	0·016	−0·008	0·002
2. $\log \rho_2$	0·381	−0·776	−0·288	0·252	−0·184	0·268	−0·015	−0·009	0·005
3. $\log \rho_3$	0·046	0·049	−0·248	−0·349	0·616	0·241	−0·561	−0·054	0·238
4. $\log \rho_4$	0·109	0·061	−0·343	−0·660	−0·486	−0·421	−0·100	0·015	0·092
5. $\log \rho_5$	0·899	0·428	0·068	0·053	0·024	0·003	0·001	0·000	0·000
6. $\log h_1$	−0·019	0·047	0·057	−0·095	−0·118	0·210	0·198	−0·898	0·284
7. $\log h_2$	−0·054	0·134	0·237	−0·332	−0·397	0·763	−0·076	0·249	−0·072
8. $\log h_3$	0·056	−0·054	−0·262	−0·379	0·387	0·152	0·717	0·067	−0·301
9. $\log h_4$	0·005	−0·002	−0·015	−0·037	−0·022	−0·025	−0·341	−0·351	−0·871

Table 3

'Error bounds' at the 2·5 per cent error level for the exact model of Fig. 2

2·5000 per cent error level

		Bound (1)	Bound (2)
ρ_1	1·0	0·61	1·65
ρ_2	1·0	0·62	1·63
ρ_3	10·0	4·35	22·99
ρ_4	1·0	0·36	2·78
ρ_5	1·0	0·89	1·12
h_1	20·0	14·97	26·71
h_2	80·0	25·14	254·61
h_3	10·0	5·21	19·19
h_4	50·0	50·00	50·00

This structure is convenient as significant parameter combinations in resistivity are often products and ratios. The factors ρ_i, h_i, $\rho_i h_i$, h_i/ρ_i, and ρ_{i+1}/ρ_i may all appear as transformed parameters, and we normally interpret the V matrix in terms of these (familiar) combinations.

For example the second transformed parameter $\log q_2$, corresponds to the second column of V.

$$\log q_2 = -0{\cdot}392 \log \rho_1 - 0{\cdot}730 \log \rho_2 - 0{\cdot}039 \log \rho_3 - 0{\cdot}035 \log \rho_4 + 0{\cdot}544 \log \rho_5 + 0{\cdot}027 \log h_1 + 0{\cdot}109 \log h_2 - 0{\cdot}042 \log h_3.$$

For interpretation we say

$$q_2 \simeq \rho_2/\rho_5 \text{ (with } \rho_2 \text{ predominant).}$$

In a similar way the six Important parameter combinations are, ρ_5, ρ_2/ρ_5, ρ_1, q_4, q_5, h_2/ρ_4, and the three Unimportant parameters are h_3/ρ_3, h_1, and h_4. The parameters q_4 and q_5 are more complicated. q_5 is roughly $h_2 \rho_4/\rho_3$, and q_4 is roughly $\rho_4 \rho_3 h_3$. q_5 is an important interactive parameter and is above the threshold at the 0·1, or 10 per cent level.

We may also use the V matrix to assess how important a given layer parameter is. The rows of V represent the way the original parameters are spread among the transformed parameters. For example, ρ_3 corresponds to the third row of V, and has significant components in columns corresponding to q_4, q_5 and q_7 with most ' energy '

Table 4

'Error bounds' at the 2·5 per cent error level for the final model of Fig. 3.

2·5000 per cent error level

		Bound (1)	Bound (2)
ρ_1	0·97	0·53	1·79
ρ_2	0·99	0·54	1·83
ρ_3	9·1	3·83	21·63
ρ_4	1·06	0·37	3·04
ρ_5	1·0	0·90	1·11
h_1	19·1	12·71	28·71
h_2	67·1	18·17	247·73
h_3	9·27	4·60	18·71
h_4	50·0	43·70	57·21

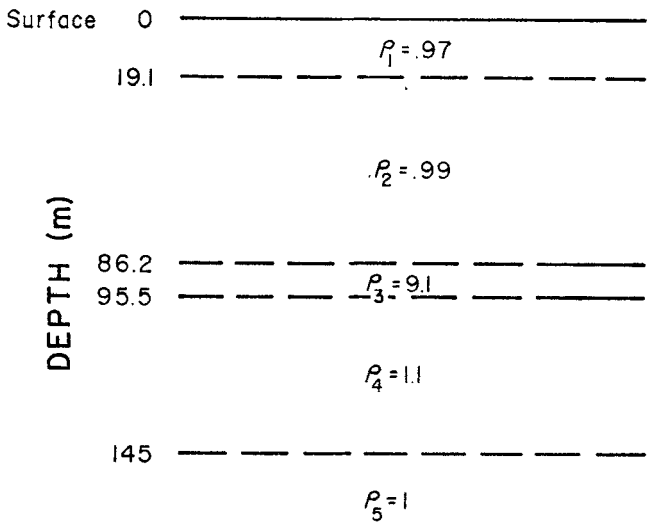

FIG. 3. The final model after inversion against data with 3 per cent added Gaussian noise.

in the q_4 and q_7 columns. As noted above, q_4 is roughly the resistivity thickness $(\rho_3 h_3)$ and q_7 the conductivity thickness (h_3/ρ_3) for the third layer.

Not every situation can be interpreted in this way. If the depth to a layer

$$Z_i = h_1 + h_2 + \ldots + h_{i-1}$$

or the reflection coefficient

$$= \frac{\rho_{i+1} - \rho_i}{\rho_{i+1} + \rho_i}$$

is a well-resolved feature of the model it cannot be simply constructed from the logarithms of layer parameters. This is not, however, a great limitation.

Following Section 2.8 the damped ' error bounds ' for the layer parameters are given by the inequalities (2.8.2). That is,

$$\pm (r/s_1)\,|V|\mathbf{e}$$

where r is a mean relative error level, which is normally of the order of the final, or expected residual error after inversion. If $\bar{\rho}_j$ and $\bar{h}_{j-5}$ represent given, or ' mean ' values for the layer parameters, we may write

$$\log \rho_j = \log \bar{\rho}_j \pm e_j \quad j = \overline{1, 5}$$

and

$$\log h_{j-5} = \log \bar{h}_{j-5} \pm e_j \quad j = \overline{6, 9}$$

where

$$e_j = \frac{r}{s_1} \sum_{i=1}^{9} v_{ji}\, t_i / k_i$$

Table 3 displays the corresponding bounds for the exact model, with $r = 0{\cdot}25$ (2·5 per cent error). Within the limits of small errors, these represent bounds in which the ' solution ' would be contained if the inversion were started from the correct model, and 2·5 per cent error put on the data. The bounds on the Unimportant

parameters are small, because our method (in the absence of good reason to change them) leaves them at the starting values, or varies them only as much as they interact with the Important parameters.

3.3. *Result of inversion from the 'exact' model*

Starting from the exact model, with $\mu = 0{\cdot}01$, and with data perturbed to the 3 per cent relative error level, we obtained, after inversion, the final model of Fig. 3. The residual error was 2·46 per cent, which was mainly the original error, but shows the model fits some of the original error at the expense of its 'truth'.

Table 2 records the singular values, the damping factors, and the V matrix, while Table 4 displays the 'error bounds', all of which may be interpreted as described in Section 3.2.

There has been a shift in the nature of the 'transformed' parameters, but the structure of Important /Unimportant parameters is essentially the same. The parameters q_4 and q_5, corresponding to columns 4 and 5 of V, are much more interactive for this model, but are made up of q_4 and q_5 of the exact model.

From Table 3, the final model is within the 2·5 per cent coarse 'error bounds' on the exact model, and the exact model is within the 2·5 per cent coarse 'error bounds' of the final model.

The example is, of course, artificial, and all of the Unimportant parameters are nicely set at their actual (unseen) values. It does, however, illustrate the nature of the analysis in the theoretical case, and provide a basis for its extension to more realistic situations.

Acknowledgments

The Australian Research Grants Committee and Macquarie University supported this research. We are indebted to Dr Jack Cribb of the CSR Research Department and Dr B. David Johnson of Macquarie University for their critical comments.

References

Backus, G. E. & Gilbert, J. F., 1967. Numerical application of a formalism for geophysical inverse problems, *Geophys. J. R. astr. Soc.*, **13,** 247–276.

Backus, G. E. & Gilbert, J. F., 1968. The resolving power of gross earth data, *Geophys. J. R. astr. Soc.*, **16,** 169–205.

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

Claerbout, J., 1973. Robust modeling with erratic data. Geophysics, **38,** (5), 826–844.

Faddeev, D. K. & Faddeeva, V. N., 1969. Stability in linear Algebra problems, in *Information processing* 68, North Holland, Amsterdam.

Forsythe, G. & Moler, C., 1967. *Computer solution of linear algebraic systems*, Prentice-Hall, New Jersey.

Francis, J., 1961. The QR transformation. A unitary analog to the LR transformation, *Computer J.*, **4,** 265–271.

Golub, G. H. & Reinsch, C., 1970. Singular value decomposition and least-squares solutions, *Num. Math.*, **14,** 403–420.

Hanson, R. J., 1971. A numerical method for solving Fredholm integral equations of the first kind using singular values, *SIAM J. Num. Anal.*, **8,** 616–622.

Jackson, D. D., 1972. Interpretation of inaccurate, insufficient and inconsistent data, *Geophys. J. R. astr. Soc.*, **28,** 97–109.

Jennings, L. S. & Osborne, M. R., 1970. Applications of orthogonal Matrix transformations to the solution of systems of linear and non-linear equations, *Tech. Report No.* 37, *Aust. Nat. Univ.*, Canberra.

Kowalik, J. & Osborne, M. R., 1968. *Methods for unconstrained optimization problems*, American Elsevier.

Lanczos, C., 1958. Linear Systems in self-adjoint form, *Am. math. Monthly*, **65,** 665–679.

Lanczos, C., 1960. *Linear differential operators*, D. Van Nostrand, London.

McMechan, G. A. & Wiggins, R. A., 1972, Depth limits in body wave inversion, *Geophys. J. R. astr. Soc.*, **28,** 459–473.

Madden, T. R., 1972. Transmission systems and network analogies to geophysical forward and inverse problems, *Tech. Rept. No.* 72–3 *Dept. of Earth and Planetary Sciences*, M.I.T.

Marquardt, D. W., 1963. An algorithm for least-squares estimation of non-linear parameters, *J. SIAM*, **11,** 431–441.

Moore, E. H., 1920. On the reciprocal of the general algebraic Matrix, *Bull. Am. math. Soc.*, **26,** 394–395.

Morrison, D. D., 1960. Methods for non-linear least-squares problems, and convergence proofs, tracking programs and orbit determination, *Proc. Jet propulsion Lab. Seminar*, 1–9.

Osborne, M. R., 1972. A class of methods for minimising a sum of squares, *Aust. Comp. J.*, **4,** 164–169.

Penrose, R., 1955. A generalized inverse for matrices, *Proc. Camb. Phil. Soc.*, **51,** 406–413.

Varah, J. H., 1973. On the numerical solution of ill-conditioned linear systems with applications to ill-posed problems, *SIAM J., Num. Anal.*, **10,** (2) 257–267.

Vozoff, K. & Jupp, D. L. B., 1975; Joint inversion of geophysical data, *Geophys. J. R. astr. Soc.*, **42,** 977–991

Wiggins, R. A., 1972. The general linear inverse problems: implication of surface waves and free oscillations for earth structure, *Rev. Geophys. Space Phys.*, **10,** 251–285.