GEOPHYSICS, VOL. 52, NO. 3 (MARCH 1987); P. 289–300, 6 FIGS., 5 TABLES.

Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data

Steven C. Constable*, Robert L. Parker*, and Catherine G. Constable*

ABSTRACT

The inversion of electromagnetic sounding data does not yield a unique solution, but inevitably a single model to interpret the observations is sought. We recommend that this model be as simple, or smooth, as possible, in order to reduce the temptation to overinterpret the data and to eliminate arbitrary discontinuities in simple layered models.

To obtain smooth models, the nonlinear forward problem is linearized about a starting model in the usual way, but it is then solved explicitly for the desired model rather than for a model correction. By parameterizing the model in terms of its first or second derivative with depth, the minimum norm solution yields the smoothest possible model.

Rather than fitting the experimental data as well as possible (which maximizes the roughness of the model), the smoothest model which fits the data to within an expected tolerance is sought. A practical scheme is developed which optimizes the step size at each iteration and retains the computational efficiency of layered models, resulting in a stable and rapidly convergent algorithm. The inversion of both magnetotelluric and Schlumberger sounding field data, and a joint magnetotelluric-resistivity inversion, demonstrate the method and show it to have practical application.

INTRODUCTION

The inversion of actual field data from a geoelectrical sounding experiment (that is, magnetotelluric, dc resistivity or controlled-source electromagnetic) cannot yield a unique solution even though it has been proven (e.g., Langer, 1933) that ideal observations can yield such a solution. It would be unfair to say that geophysicists are unaware of this fact, but they usually shield themselves from its implications by imposing constraints on the models they seek, to stabilize the solution and give it the illusion of uniqueness. A very common example is restriction of the solution to the class of models consisting of a small number of layers (often less than five). We call these "simple layered models." However, this approach produces solutions that are dependent upon the class of models chosen. For example, selecting a small number of layers tempts one to believe there really are large discontinuities between the layers at the depths discovered by the computer program. Even if the modeler exercises proper caution, readers of his work may not.

It may be argued that in some circumstances it is reasonable to represent the earth by a simple layered model, for example when trying to establish depth to a water table or the basement of a sedimentary sequence. This does not alter the dependence of solutions on the parameterization, and so before the solution reflects the true structure of the earth, the parameterization and starting model must be close to being correct. Furthermore, if more than four or five layers are suggested by a priori information (by a well log, say), a least-squares inversion is unlikely to constrain such a highly parameterized model.

There has been no great success in overcoming the uniqueness problem associated with practical (that is uncertain, incomplete) data. The Monte Carlo method, in which a huge number of randomly generated models are tested against the data, has been used for resistivity (Sternberg, 1979) and magnetotelluric (MT) (Jones and Hutton, 1979b) soundings in an attempt to characterize all models which agree with the observations. Such computations can never be exhaustive, and even calculations ranging over the class of simple layered models are computationally extravagant in the light of the insight obtained from them. In the absence of a universally valid description of the set of models consistent with a given geoelectric data set, the best policy may be to seek a model whose features are in some way essential characteristics of any of the possible solutions, one of which presumably is the true structure.

We propose finding the smoothest model in a special sense so that its features depart from the simplest case only as far as is necessary to fit the data. Other, more exciting models will be able to satisfy observations, but many of them will be far

Manuscript received by the Editor February 28, 1986; revised manuscript received July 30, 1986.
*Scripps Institution of Oceanography, University of California at San Diego, Mail Code A030, La Jolla, CA 92093.
© 1987 Society of Exploration Geophysicists. All rights reserved.

more provocative and attractive than reality; our approach guarantees that the real profile must be at least as rich in structure as the profile found but never less complex in structure. It appears that earlier iterative techniques for finding even simple models still suffer from accidental discovery of unnecessarily complex solutions.

The quest for simple solutions is well founded. In the early fourteenth century William of Occam wrote that "it is vain to do with more what can be done with fewer" (see Russell, 1946, ch. 14). What has become known as Occam's razor has also become a fundamental tenet of modern science; hypotheses should be neither unnecessarily complicated nor unnecessarily numerous.

The basic motivation for seeking smooth models is that we do not wish to be misled by features that appear in the model but are not essential in matching the observations. Figure 1 shows two models which fit a set of Schlumberger data equally well: a slowly varying model obtained by the methods described here, and a model obtained by the popular Marquardt method (Marquardt, 1963). The Marquardt technique, also called ridge regression, is stable and efficient for parameter fitting and has been used extensively to interpret geoelectrical data (e.g., Inman, 1975; Petrick et al., 1977). However, by allowing 27 layers the normal restrictions of a simple layered model have been lost, resulting in spurious high-conductivity zones 1 and 10 km deep and a general fine structure which is completely meaningless. Clearly, to attach any importance to the low-conductivity zones would be a mistake, because even though we cannot preclude their existence, they are not demanded by the data. The other model fits the data equally well without them.

On the other hand, we do have some justification for thinking that a feature appearing in the most featureless solution has some significance. Thus, we hope that the low-resistivity and high-resistivity regions of the smooth model are a simplified but reasonable representation of the real earth. Another way to consider the situation is to realize that electromagnetic (EM) sounding experiments cannot resolve sharp boundaries or thin layers; the diffusive nature of the energy propagation "smears out" the real earth structure. We believe it is appropriate to construct models that reflect this limitation of the experiment. We stress, however, that on the basis of the Schlumberger data alone there is no reason to believe the smooth model in Figure 1 to be any closer to the real earth than the Marquardt model. If the measurements are to be our sole guide, there is nothing to choose between the two profiles because they do an equally good job of predicting the data.

An advantage of inverting for maximally smooth models is that we obtain a specific model whose characteristics we have chosen; the solution does not depend upon some arbitrary starting guess or some accident of the computer program.

LAYERED MODELS AND SMOOTH MODELS

Conventional least-squares inversion for a simple layered model derives its stability from the essential smoothness of the conductivity function within the layers. Thus, a four-layer model requires the conductivity function to be piecewise perfectly smooth with three discontinuities. This restriction is relaxed as the number of layers increases, and at some point the layer thickness will be below the resolving capabilities of the data. This is the point at which the model will begin to exhibit

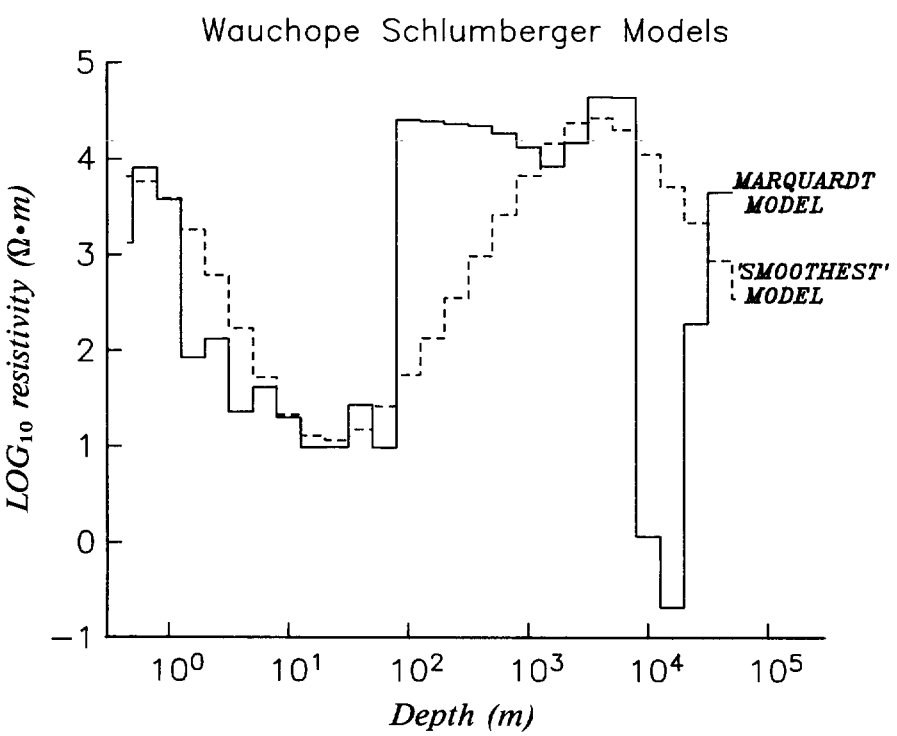

FIG. 1. Two models which fit a set of Schlumberger resistivity data (described by Constable et al., 1984) equally well (with an rms misfit of 1.0). The solid line is a model obtained by Marquardt inversion with a 27 layer starting model, and the broken line is the smoothest model in the sense of the second derivative, found from the same starting model.

oscillations not required by the sounding data. Fitting simple layered models is therefore a delicate balance between suppressing significant structure by including too few parameters in the model and introducing spurious structure by including too many parameters.

We assert that it is more satisfactory to allow the model to be as flexible as possible, but to suppress complexity explicitly. We can do this for continuous profiles by defining roughness, the converse of smoothness, as the integrated square of the first or second derivative with respect to depth:

$$R_1 = \int (dm/dz)^2 \, dz$$

or (1)

$$R_2 = \int (d^2m/dz^2)^2 \, dz,$$

where $m(z)$ could be resistivity or log resistivity. The strategy is to find the solution agreeing with measurement that has the smallest possible roughness. This idea is familiar from the modern methods of data interpolation, because, for example, cubic spline interpolation is just the curve passing through a given series of measurement points with the smallest possible R_2 (see de Boor, 1978). Exact fitting is often assumed, but splines allowing misfit are also employed for data smoothing. The original idea of a penalty for complexity seems to be due to Tikhonov, who named the general procedure "regularization," introducing it in order to overcome mathematical difficulties in the theory of ill-posed problems (see Tikhonov and Arsenin, 1977). It is our contention that regularization has enormous practical benefits in the interpretation of experimental data.

Although equation (1) requires a smoothly varying function as the model, a realizable computer algorithm is more readily based upon a set of piecewise constant layers to perform the calculations of the forward problem more efficiently. We use the notation

$$m(z) = m_i, \quad z_{i-1} < z \leq z_i, \quad i = 1, 2, \cdots, N,$$

where $z_0 = 0$ and in practice N typically falls in the range 20 to 100. Because of the inevitable loss of resolution with depth, it is sensible to arrange z_{i-1}/z_i to be some constant less than unity. A uniform half-space terminates the system. At this point we may think of m_i as a resistivity or a conductivity. An equivalent roughness in the discrete representation is based upon difference rather than differential operators. Let us say that

$$R_1 = \sum_{i=2}^{N} (m_i - m_{i-1})^2$$

and (2)

$$R_2 = \sum_{i=2}^{N-1} (m_{i-1} - 2m_i + m_{i+1})^2.$$

Suppose now there are M data, $d_1, d_2, \ldots, d_M$. These may be apparent resistivities and phases at various frequencies, or apparent resistivities at different electrode spacings, or any combination of these in a joint inversion. It will be assumed that an error estimate σ_j is associated with each datum. Forward modeling allows prediction of the values of these measurements from our discrete model via the functionals $F_j[m]$. We assess the goodness of fit of the model predictions to the actual values with the usual weighted least-squares criterion

$$X^2 = \sum_{j=1}^{M} (d_j - F_j[m])^2/\sigma_j^2, \tag{3}$$

where σ_j is the uncertainty in the jth datum (assuming statistical independence in the error).

We are now ready to state the mathematical problem to be solved: for given data d_j and the associated uncertainties, we must find the model m_i that makes R_1 or R_2 as small as possible, while X^2 achieves an acceptable value. This is a nonlinear optimization problem (in contrast to spline smoothing, for example, which is linear). Because of the nonlinearity, there is no guarantee any m_i will be able to make X^2 small enough, and it is virtually certain that in every practical case exact fitting ($X^2 = 0$) is impossible. However, we assume that the approximations of one-dimensionality, large-scale source fields, etc., are all good enough that a reasonable fit to the observations is possible. The problem of finding the smallest achievable X^2 associated with an arbitrary 1-D profile for MT or dc resistivity data has been completely solved (Parker, 1980; Parker and Whaler, 1981; Parker, 1984), so in these cases we can begin by calculating a lower limit on the values considered.

To explain how we find the smoothest model in the *nonlinear* case, it is convenient to consider first the much easier question of a *linear* forward problem. Notational compactness is gained by adoption of vector notation; our model is written as $\mathbf{m} \in E^N$, a vector in the Euclidean space of dimension N. Similarly, the measurements which comprise M real values will be denoted by $\mathbf{d} \in E^M$. In general, the solution to the forward problem can always be expressed

$$d_j = F_j[\mathbf{m}], \qquad j = 1, 2, \ldots, M,$$

where F_j is the (usually nonlinear) forward functional associated with the jth datum; in vector notation,

$$\mathbf{d} = \mathbf{F}[\mathbf{m}].$$

When the forward problem is linear, this may be replaced by

$$\mathbf{d} = \underline{\mathbf{G}}\mathbf{m}$$

where $\underline{\mathbf{G}}$ is an $M \times N$ matrix whose elements may be calculated from the theory of the forward problem. For this simple linear case, the misfit X^2 defined in equation (3) can be written

$$X^2 = \| \underline{\mathbf{W}}\mathbf{d} - \underline{\mathbf{W}}\underline{\mathbf{G}}\mathbf{m} \|^2, \tag{4}$$

where $\underline{\mathbf{W}}$ is the diagonal $M \times M$ matrix

$$\underline{\mathbf{W}} = \text{diag}\,\{1/\sigma_1, 1/\sigma_2, \ldots, 1/\sigma_M\}$$

and $\| \cdot \|$ denotes the usual Euclidean norm. Continuing in this vein, we can write the roughness in terms of simple matrix operations; for example, R_1 given in equation (2) is

$$R_1 = \| \underline{\partial}\mathbf{m} \|^2 \tag{5}$$

where $\underline{\partial}$ is an $N \times N$ matrix given by

$$\underline{\partial} = \begin{bmatrix} 0 & & & & \\ -1 & 1 & & & 0 \\ & -1 & 1 & & \\ & & \cdots & & \\ & 0 & & -1 & 1 \end{bmatrix}.$$

Notice that R_2 in the discrete case is just

$$R_2 = \|\underline{\partial}\underline{\partial}\mathbf{m}\|^2 = \|\underline{\partial}^2\mathbf{m}\|^2.$$

For the linear discrete case now under consideration, the mathematical minimization problem is this: We must minimize R_1 of equation (5) subject to the condition that the misfit X^2 in equation (3) is equal to X^2_* (a value deemed acceptable in view of the uncertainties). If the uncertainties are due to a zero-mean, Gaussian process that is independent in each of the observations, and σ_j are the associated standard deviations, then X^2 is well known to be distributed as χ^2. This is a great deal to assume about the noise, but often the uncertainties are rather poorly known and more refined statistical models may not be worth the considerable additional labor. With the Gaussian model, the expected value of X^2 is just M, the number of data, and is equivalent to an rms misfit of 1. It is unlikely that any other kind of distribution function would yield a value for the expected X^2 that is widely different. M is

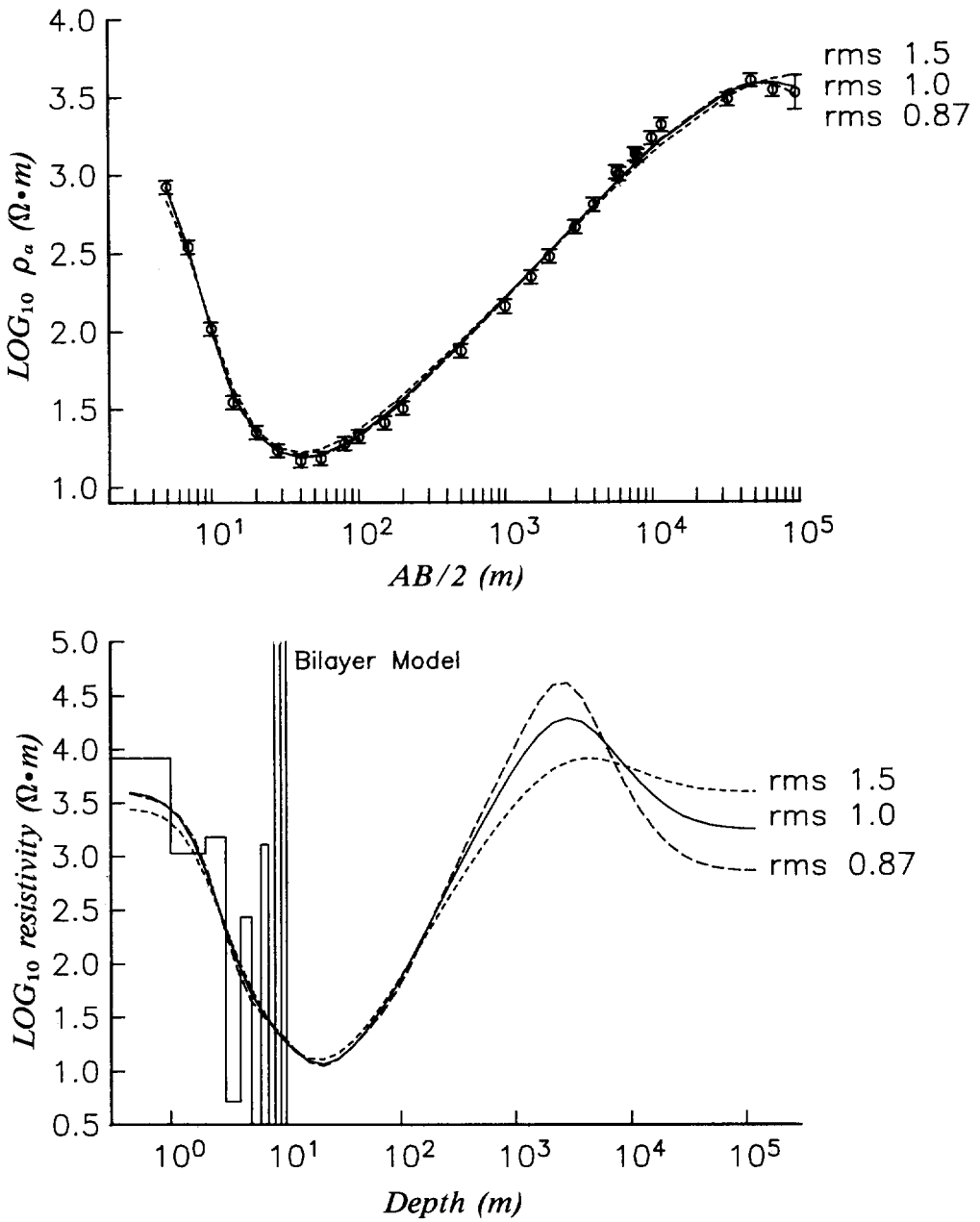

FIG. 2. Three maximally smooth models (below), in a first derivative sense, and their responses (above) for three different rms misfits. A marginal improvement in fit to the data requires a substantial increase in model structure. The data are the same as for Figure 1. Also shown is the best-fitting 1-D (bilayer) model, which produces a misfit of 0.75. The resistivities of the bilayer model exceed the range used for this plot.

the number we usually adopt for X^2. In any case, X_*^2 should not be chosen to be too close to the smallest achievable value. Models corresponding to the smallest possible X^2 are rough to the point of being physically unreasonable; they are delta functions in the case of MT (Parker, 1980) and arbitrarily thin layers in the case of resistivity sounding (Parker, 1984). As one approaches small values of X^2, a substantial increase in roughness is required to achieve only marginal improvement in fit. The degree of increase is clear in Figure 2, which shows the smoothest models and response functions for three levels of desired misfit, as well as the best-fitting 1-D model.

The optimization is performed as follows. To minimize a functional subject to a constraint, we use the method of Lagrange multipliers (see Smith, 1974): the constraint equation is rearranged to form an expression equal to zero; that expression is multiplied by a parameter, the Lagrange multiplier, and added to the functional to be minimized; the original functional is minimum where the new one, varying with its original parameters and the Lagrange multiplier, is stationary without constraint. It is convenient to call the Lagrange multiplier μ^{-1}. Then the unconstrained functional is

$$U = \| \underline{\partial}\mathbf{m} \|^2 + \mu^{-1}\{\| \underline{W}\mathbf{d} - \underline{W}\underline{G}\mathbf{m} \|^2 - X_*^2\},$$

where the first term on the right is the roughness and the second the misfit, weighted by the Lagrange multiplier. For any value of μ this functional of $\mathbf{m}$ is stationary when $\nabla_m U$, the gradient of U with respect to $\mathbf{m}$, vanishes. After a little algebra, we find

$$\mu^{-1}(\underline{W}\underline{G})^T\underline{W}\underline{G}\mathbf{m} - \mu^{-1}(\underline{W}\underline{G})^T\underline{W}\mathbf{d} + \underline{\partial}^T\underline{\partial}\mathbf{m} = 0. \quad (6)$$

Rearrangement gives

$$\mathbf{m} = \left[\mu\underline{\partial}^T\underline{\partial} + (\underline{W}\underline{G})^T\underline{W}\underline{G}\right]^{-1}(\underline{W}\underline{G})^T\underline{W}\mathbf{d}. \quad (7)$$

Variation with respect to μ yields the original constraint condition. Because μ is not known, equation (7) does not completely solve the problem; μ must be selected so that when equation (7) is substituted into equation (4), the desired X^2, namely X_*^2, is obtained. An almost identical problem arises in construction of optimally smooth magnetic fields after downward continuation (Shure et al., 1982), but we defer discussion of this search because the question deserves special attention in the actual nonlinear problem which we discuss later. It is useful to interpret μ as a kind of smoothing parameter: when μ is large, we see from the definition of U that the solution to equation (7) is not influenced much by the data misfit; it is a very smooth function. Alternatively, as μ tends to zero, the roughness term is of little significance in the minimization problem, and $\mathbf{m}$ will satisfy the data constraints at whatever cost in roughness.

THE NONLINEAR PROBLEM

When the full nonlinear problem is considered, the functional to be minimized is still R_1 given by equation (5), but the expression for the data misfit, equation (4), becomes

$$X^2 = \| \underline{W}\mathbf{d} - \underline{W}\mathbf{F}[\mathbf{m}] \|^2.$$

The theory of constrained minimization, however, instructs us to proceed in the same way: an unconstrained functional U is formed by means of a Lagrangian multiplier:

$$U = \| \underline{\partial}\mathbf{m} \|^2 + \mu^{-1}\{\| \underline{W}\mathbf{d} - \underline{W}\mathbf{F}[\mathbf{m}] \|^2 - X_*^2\}. \quad (8)$$

The extremal values of R_1 will be found at the stationary points of U as before; taking the gradient, we find the vectors $\mathbf{m}$ that cause U to be stationary obey

$$\mu^{-1}(\underline{W}\underline{J})^T\underline{W}\underline{J}\mathbf{m} - \mu^{-1}(\underline{W}\underline{J})^T\underline{W}\mathbf{d} + \underline{\partial}^T\underline{\partial}\mathbf{m} = 0, \quad (9)$$

where the $M \times N$ matrix $\underline{J}$ is the Jacobian or gradient matrix:

$$\underline{J} = \nabla_m \mathbf{F}.$$

Expressed as components

$$J_{ij} = \frac{\partial F_i[\mathbf{m}]}{\partial m_j}.$$

(See the Appendix for details of the computations.) The rows of $\underline{J}$ are the discrete equivalents of the Fréchet derivatives in the continuous profile problem (Parker, 1977). In the linear problem, $\underline{G} = \underline{J}$; what makes the solution of equation (9) much more difficult than equation (6) is that, while $\underline{G}$ is a constant matrix, $\underline{J}$ *depends upon* $\mathbf{m}$. Thus, instead of a simple set of linear equations similar to equation (7), we must solve a simultaneous nonlinear system for $\mathbf{m}$. One way to proceed is to attack equation (9) directly and solve the system by Newton's method; unfortunately, this requires differentiation of $\underline{J}$ to find the second derivative of $\mathbf{F}$, a very tedious piece of algebra.

A simple alternative is to return to equation (8) and examine the minimization problem created by linearization about a particular model. Most solutions of nonlinear systems require an initial guess for the answer, i.e., a starting model, from which an iterative process begins a refinement procedure; here we posit an initial model $\mathbf{m}_1$. Elementary calculus says that if $\mathbf{F}$ is differentiable at $\mathbf{m}_1$ (as we shall always assume it is) for sufficiently small vectors Δ

$$\mathbf{F}[\mathbf{m}_1 + \Delta] = \mathbf{F}[\mathbf{m}_1] + \underline{J}_1\Delta + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\varepsilon}$ is a vector whose magnitude is $o\|\Delta\|$ and $\underline{J}_1$ is $\underline{J}[\mathbf{m}_1]$, the Jacobian matrix evaluated at the vector $\mathbf{m}_1$. Suppose we approximate $\mathbf{F}$ by dropping the remainder term $\boldsymbol{\varepsilon}$ and write

$$\Delta = \mathbf{m}_2 - \mathbf{m}_1.$$

If this approximate expression is substituted into equation (8), we have returned to a problem linear in $\mathbf{m}_2$:

$$U = \| \underline{\partial}\mathbf{m}_2 \|^2 + \mu^{-1}\left\{\| \underline{W}(\mathbf{d} - \mathbf{F}[\mathbf{m}_1] + \underline{J}_1\mathbf{m}_1) - \underline{W}\underline{J}_1\mathbf{m}_2 \|^2 - X_*^2\right\},$$

where the expression in parentheses in the second term is a kind of data vector which we will call $\hat{\mathbf{d}}_1$. Now we define $\mathbf{m}_2$ as the model that minimizes U under this approximation; then we find from the linear theory that

$$\mathbf{m}_2 = \left[\mu\underline{\partial}^T\underline{\partial} + (\underline{W}\underline{J}_1)^T\underline{W}\underline{J}_1\right]^{-1}(\underline{W}\underline{J}_1)^T\underline{W}\hat{\mathbf{d}}_1.$$

If we select μ to yield the desired data misfit as computed by the linear approximation, we have an obvious basis for an iterative scheme: $\mathbf{m}_2$ is used as the next member of a sequence $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \cdots$, in which each preceding vector is used as the starting approximation for the next.

It may be shown that this scheme, if it converges, solves the minimization of the original nonlinear functional, and therefore the final answer should be independent of the starting guess (provided the minimum is unique). The result will be the model of smallest roughness with the specified misfit. This minimization scheme is in contrast with the approach usually adopted in the geophysical literature, in which the functional to be minimized is simply the misfit itself. Because of the nonlinearity of the problem, a straightforward minimization scheme is most likely to diverge, and this tendency must be counteracted by damping the process in some way. Damping consists of systematically reducing the size of the change in the model from one iteration to the next. The Marquardt method modifies the matrix $\mathbf{J}^T\mathbf{J}$ by adding a constant value to the diagonal, thus decreasing the size of the computed perturbation; the constant is chosen to decrease the misfit at each iteration, for a fixed Jacobian. Iteration continues until the misfit has been brought down to some acceptable level, not to the minimum possible value. The final solution lies close to the initial guess, because the modified Jacobian keeps the changes small at each step of the process; thus the resultant

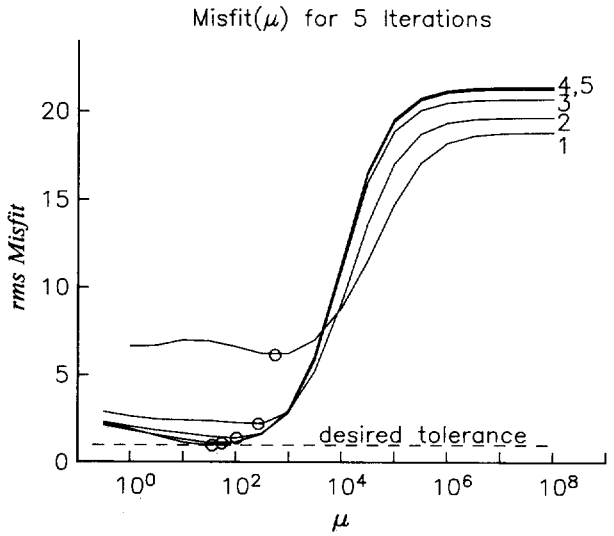

FIG. 3. The rms misfit as a function of the Lagrange multiplier μ at five iterations of an inversion. The values of μ and misfit chosen for each iteration are marked by symbols and also given in Table 1. The corresponding intermediate models are shown in Figure 4.

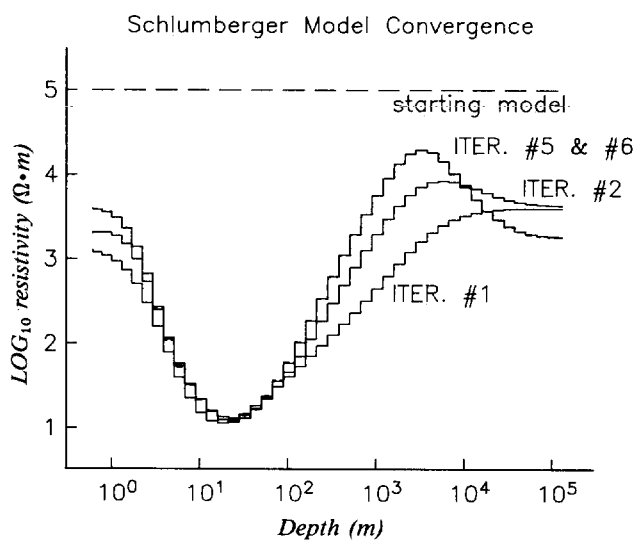

FIG. 4. The intermediate models produced during inversion of the Schlumberger data (Figure 2) for a smooth model (first derivative sense) fitting the data to rms misfit of 1.0. The starting model was a half-space of $10^5\ \Omega \cdot m$.

Table 1. Parameters for the iteration shown in Figures 3 and 4. The values of μ are those chosen to give the intermediate models.

No.	X^2	rms	μ	$\|R_1\|^2$	$\|\Delta\|^2$
0	11 250	63.0		0.0	
1	1 062	6.16	571	0.75	309
2	138.5	2.22	273	1.04	5.01
3	56.0	1.404	106	1.31	0.921
4	34.4	1.111	55.2	1.48	0.660
5	29.4	1.007	46.1	1.50	0.071
6	28.9	1.001	52.6	1.50	0.005

Table 2. Data used for the Schlumberger sounding example, from Constable et al. (1984).

AB/2 (m)	$\log_{10} \rho_a$ (Ω·m)	$\sigma_{\log \rho}$ (Ω·m)
5.0	2.923	0.043
7.0	2.539	0.043
10.0	2.017	0.043
14.0	1.546	0.043
20.0	1.354	0.043
28.0	1.235	0.043
40.0	1.170	0.043
55.0	1.184	0.043
80.0	1.281	0.043
100.0	1.326	0.043
150.0	1.413	0.043
200.0	1.505	0.043
500.0	1.873	0.043
1 000.0	2.158	0.043
1 500.0	2.345	0.043
2 000.0	2.476	0.043
3 000.0	2.661	0.043
4 000.0	2.804	0.043
5 733.0	3.011	0.043
6 000.0	2.997	0.043
7 705.0	3.126	0.043
8 000.0	3.116	0.043
10 000.0	3.231	0.043
11 660.0	3.317	0.043
32 979.9	3.479	0.043
47 780.0	3.601	0.043
67 369.9	3.538	0.043
95 360.0	3.521	0.110

model is strongly influenced by the initial choice. This is clearly seen in Figure 1 where the large discontinuity at 80 m depth is a relic of the original $\mathbf{m}_1$. When a small number of parameters is employed, for example in the simple layered model, the best-fitting solution in the limited class is sought and simplicity is forced by the small number of free parameters. Sometimes the misfit cannot be reduced sufficiently, so more parameters may have to be introduced. Large-amplitude oscillations may develop, and there is no ready means of controlling them in the Marquardt scheme.

We have suggested that the successive iterates of our scheme, $\mathbf{m}_{k+1}$, are to be found by choosing μ in such a way that the linear approximation to the misfit would be arranged to match the desired tolerance. Unless one is close to a solution, the linear approximation is unlikely to be an accurate reflection of the true misfit, and, since deriving a solution using linear approximations itself requires a number of iterative steps, it seems this approach is unnecessarily time-consuming. In fact, experience with this direct approach suggests that it rarely converges in its unmodified form because the starting approximations are not sufficiently near a solution.

We propose an alternative scheme which has proven very effective in practice. Suppose the kth iterate has been computed; define the vector

$$\mathbf{m}_{k+1}(\mu) = \left[\mu \underline{\partial}^T \underline{\partial} + (\underline{\mathbf{W}}\underline{\mathbf{J}}_k)^T \underline{\mathbf{W}}\underline{\mathbf{J}}_k\right]^{-1} (\underline{\mathbf{W}}\underline{\mathbf{J}}_k)^T \underline{\mathbf{W}}\hat{\mathbf{d}}_k. \quad (10)$$

We sweep through values of μ computing the true misfit of the model $\mathbf{m}_{k+1}(\mu)$

$$X_{k+1}(\mu) = \| \underline{\mathbf{W}}\mathbf{d} - \underline{\mathbf{W}}\mathbf{F}\left[\mathbf{m}_{k+1}(\mu)\right] \|.$$

In the initial phase of the calculation, the main task is to reduce the misfit, because the initial guess usually lies far from any model which has adequate agreement with the observations, and whatever value of μ is selected, $X_k(\mu)$ is larger than X_*. An obvious way to proceed is to choose μ to minimize $X_k(\mu)$, which is readily accomplished with a 1-D line search. Although there is no guarantee that the X_k-minimizing model fits better than $\mathbf{m}_k$, we have found the scheme to be very satisfactory. After a number of iterations, μ can be selected to make X_k match X_* exactly. In fact, there probably will be more than one such value; if there is more than one value the largest successful μ is correct because it causes the roughness to be least.

Table 3. The COPROD data set used for the MT example, from Jones and Hutton (1979a).

Period (s)	$\log_{10} \rho_a$	$\sigma_{\log \rho}$	Phase (degrees)	σ_{phase} (degrees)
28.5	2.315	0.072 1	57.19	22.95
38.5	2.254	0.042 5	58.19	22.95
52.0	2.229	0.024 4	61.39	4.96
70.5	2.188	0.021 0	59.09	4.46
95.5	2.180	0.016 4	59.89	5.96
129.0	2.162	0.017 3	51.19	22.95
174.6	2.151	0.028 7	46.89	22.95
236.2	2.208	0.032 8	42.79	2.46
319.6	2.194	0.019 3	36.89	1.65
432.5	2.299	0.027 0	32.00	22.95
585.1	2.338	0.059 1	44.00	6.37
791.7	2.420	0.050 6	32.00	2.46
1 071.1	2.405	0.082 5	37.59	22.95
1 449.2	2.308	0.123 3	45.29	4.15
1 960.7	2.397	0.092 7	50.09	22.95

Table 4. Schlumberger data used in joint inversion, from Constable (1985).

AB/2 (m)	$\log_{10} \rho_a$	$\sigma_{\log \rho}$
3.0	1.528	0.043
5.0	1.093	0.043
7.0	0.944	0.043
10.0	0.919	0.043
14.0	0.903	0.043
20.0	0.869	0.043
28.0	0.826	0.043
40.0	0.819	0.043
55.0	0.838	0.043
80.0	0.857	0.043
100.0	0.851	0.043
150.0	0.934	0.043
200.0	0.959	0.043
300.0	0.944	0.043
400.0	0.982	0.043
500.0	1.049	0.043
1 000.0	1.356	0.082
1 500.0	1.502	0.043
2 000.0	1.622	0.043
3 000.0	1.761	0.043
4 000.0	1.902	0.043
6 000.0	2.089	0.043
8 000.0	2.192	0.043
10 000.0	2.174	0.086

Table 5. MT data used in joint inversion, from Cull (1985).

Period (s)	$\log_{10} \rho_a$	$\sigma_{\log \rho}$	Phase	σ_{phase}
0.020	0.712	0.043 4	20.50	4.50
0.050	0.813	0.140 6	22.00	19.98
0.079	0.964	0.113 8	19.28	16.08
0.126	1.148	0.160 9	20.14	11.03
0.199	1.030	0.304 6	35.65	23.57
0.316	1.238	0.488 0	34.44	20.66
0.501	1.503	0.324 4	39.71	16.60
0.794	1.520	0.477 2	41.90	23.40
1.258	1.675	0.430 6	47.28	16.60
1.995	1.944	0.441 2	42.78	16.77
3.162	1.844	0.408 1	48.50	18.41
5.011	1.737	0.523 6	56.87	18.05
7.943	1.614	0.199 5	58.80	17.63
12.580	1.592	0.478 0	58.88	18.10
19.950	1.336	0.222 8	61.44	18.31
31.620	1.358	0.447 1	61.90	16.58
50.110	1.345	0.352 4	51.48	25.52
79.430	0.935	0.104 1	65.50	10.51
125.800	0.821	0.096 1	58.20	23.53
199.500	0.535	0.400 0	54.00	20.00
794.299	0.312	0.400 0	86.00	30.00
1 258.000	−0.107	0.556 1	41.00	32.00
1 995.000	0.390	0.861 2	58.66	21.63

Figure 3 shows rms misfit, which is $(X_k^2/M)^{1/2}$, at successive iterations in the inversion of the Schlumberger data. The figure also shows the locations of the μ chosen at each iteration; the minima were found using a golden section search, and the intercepts were found using the bisection method (see, for example, Gill et al., 1981).

CONVERGENCE AND STABILITY

Because we seek a well defined, specific model (i.e., the smoothest model possible), our iterative scheme is very stable. That is, the models found at each iteration will not contain very large or very small conductivities unless they are absolutely required in order to fit the data.

The convergence of our scheme is also impressive. Figure 4 shows the starting model and the models for each of the five iterations required to fit the Schlumberger data with a maximally smooth model. Table 1 gives the values of X^2, rms error, μ_k, $\| R_1 \|^2$, and $\| \Delta \|^2$ (the step size) at each iteration. The starting model was a half-space of $10^5\ \Omega \cdot$m, and the smoothest model (in a first derivative sense) fitting the data to a tolerance of an rms error of 1 was sought. This model was found in only five iterations. A sixth iteration verified that the procedure had converged and the algorithm stopped on the criteria that $\| \Delta \|^2 < 0.01$ and $|\text{rms error} - \text{required rms error}| < 0.05$.

EXAMPLES

We present a few examples of the application of our inversion technique. In these examples the model was parameterized as $\log_{10}$ (layer resistivities), with the layer thickness held constant in the log domain. All the data, except for MT phases, were also parameterized in $\log_{10}$ domain, so the rms tolerances refer to the misfit in log space. The actual data used in these inversions are given in Tables 2 through 5.

The Schlumberger data already presented are from Constable et al. (1984). These data are from a deep sounding conducted on the central Australian shield which employed overland telephone lines to achieve a maximum electrode spacing of 200 km. Constable et al. (1984) presented a six-layer model obtained by a Marquardt method which fits the data to an rms tolerance of 0.81 when calculated in the manner just described. The best 1-D fit to these data is obtained with a bilayer model (Parker, 1984) which predicts the data to within an rms misfit of 0.75 (Figure 2). The bilayer model is physically unrealistic, being only 10 m thick and containing layer resistivities from $10^{-1}\ \Omega \cdot$m to $10^9\ \Omega \cdot$m, but it does highlight the nonuniqueness problem of electrical sounding inversion. Models which fit the Schlumberger data to an rms misfit of 1 and which are maximally smooth in a second derivative (Figure 1) and first derivative (Figure 2) sense are shown. The conclusions drawn by Constable et al. (1984) from their Mar-

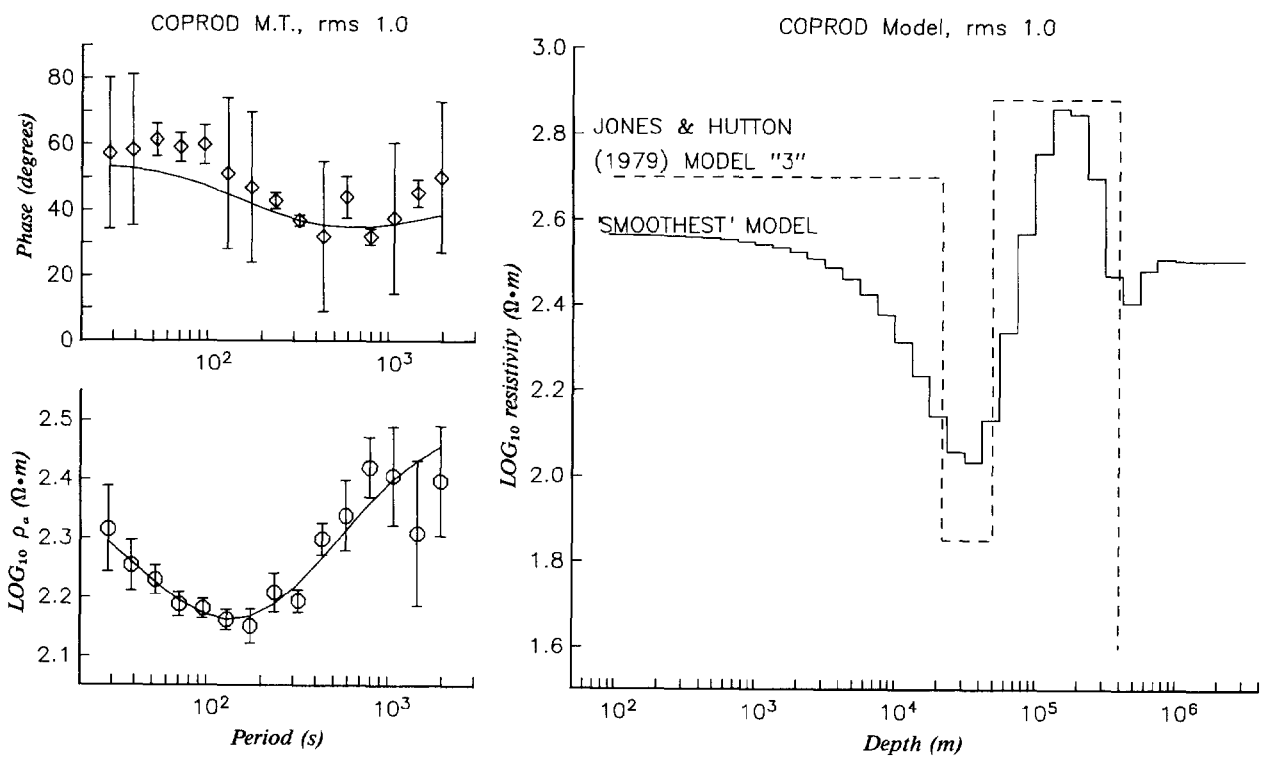

FIG. 5. The COPROD data described by Jones and Hutton (1979a), and the smallest first derivative model fitting these data to an rms tolerance of 1.0. The large error bars in phase represent measurements where phase error was indeterminate. Also shown is a simple layered model given in Jones and Hutton (1979a) for the same data. The terminating half-space of the Jones and Hutton model has a resistivity of 1 $\Omega \cdot$m.

quardt model would also have been drawn from the smooth models. In particular, the drop in resistivity at 2–10 km appears in all but the model with rms misfit of 1.5. If the data errors have been well assessed and the earth is truly one-dimensional, then the probability of the real earth generating such poorly fitting data is only 0.025. We therefore interpret the drop in resistivity at depth as being significant.

As an example of MT inversion, we present the COPROD data circulated by Dr. Alan Jones, collected at a site near Newcastleton in Britain and described in Jones and Hutton (1979a). The model that is smoothest in a first derivative sense and that fits the data to an rms misfit of 1.0 is shown in Figure 5, along with a simple layered model from Jones and Hutton (1979a). The smooth model contains all the features of their layered model, but has a more conservative resistivity for the deep conductive region. Parker (1982) shows that the maximum depth to which any model is constrained by these data is about 300 km. Our smooth model has little structure below 400 km and no structure below 700 km, in general agreement with Parker's result.

As a final example of the versatility of our inversion algorithm, Figure 6 shows the result of inverting Schlumberger and MT data simultaneously. The data are from a site in south-central Australia where both a 20 km electrode spacing Schlumberger sounding (Constable, 1985) and a wide-band MT sounding (Cull, 1985) were conducted. The joint inversion of resistivity and MT data is not a new idea (see Vozoff and Jupp, 1975); we merely wish to demonstrate the tolerance of our inversion routine to the nature of the forward problem and the consistent results that can be obtained from data of mixed type. Figure 6 shows the joint model and the responses fitting the various sounding data to a combined rms error of 1. For comparison, the smooth models with rms errors of 1 obtained when the two data sets are inverted independently are also shown (the responses are not given here). The only significant difference between the joint fits and individual fits occurs in the MT phase, which has a slightly poorer fit below a period of 1 s for the joint model. However, the errors on the MT data are probably too large, resulting in a preferential fit to the Schlumberger data. Note that there is no reason to believe the structure at this site is one-dimensional; the inversion was done as an exercise to test the modeling algorithm. However, a 1-D model fits both data sets, and Cull (1985) inferred the presence of the deep conductive layer on the basis of several additional MT stations to the east of the site represented here.

CONCLUSIONS

Although the nonuniqueness in inverting electromagnetic sounding data is well known, we usually require a preferred model to represent and interpret our data. It is most desirable to avoid including features in that model which are not actu-

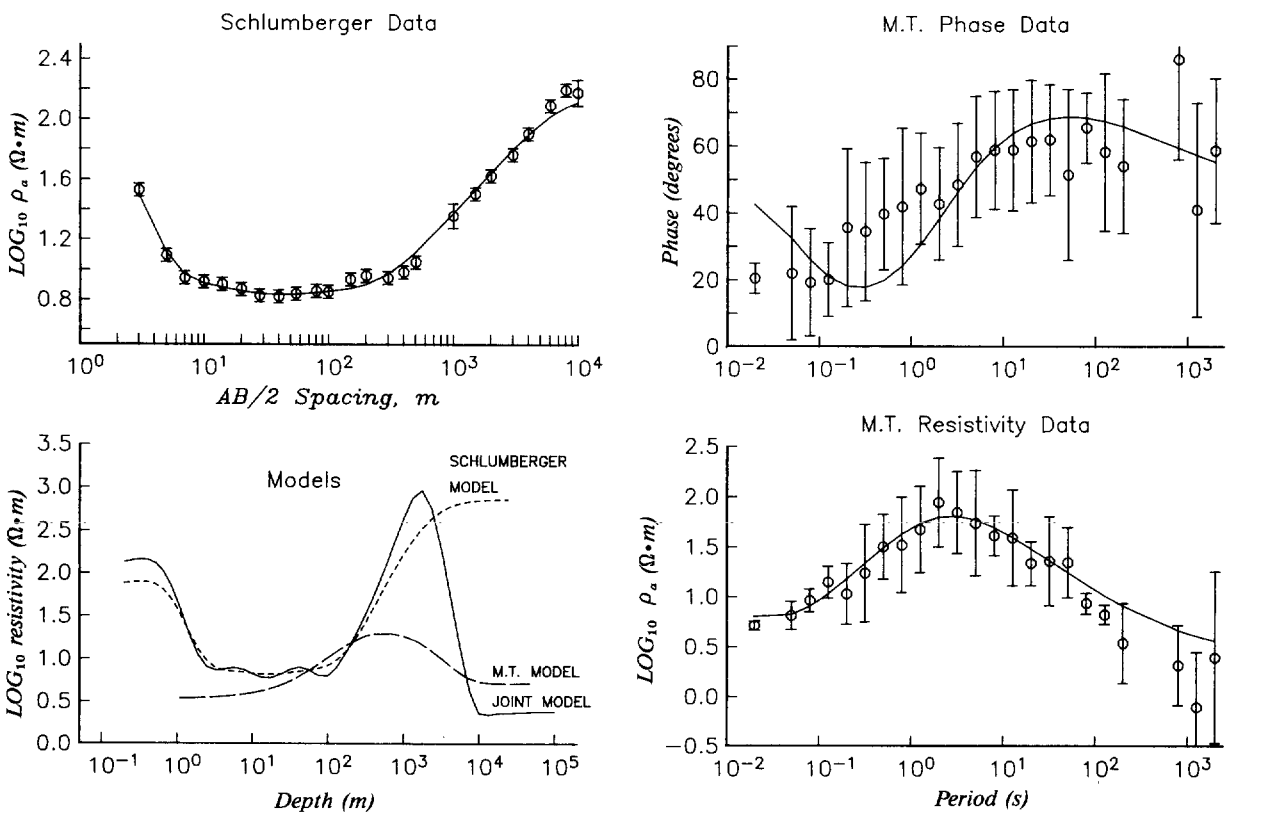

FIG. 6. The joint inversion of Schlumberger (Constable, 1985) and MT (Cull, 1985) data. The model responses are all for the joint model (smooth in a first derivative sense), but models obtained by fitting the data sets separately are also shown. The steps in the models due to layering have been smoothed for clarity.

ally required by the data, and for the model not to depend on the number of layers used or the starting model chosen. We have shown how to accomplish these goals by finding the smoothest model which fits the data to a prescribed tolerance. If the data errors have been well estimated, have zero mean, and are at least approximately Gaussian, then the expected tolerance is equivalent to an rms misfit of 1.0. Since the model found will be, to a large extent, dependent upon the data errors as well as the data themselves, every effort should be made in the field to collect accurate estimates of error. Of course, this is the case for any reasonable method of data interpretation.

Our inversion scheme is stable and typically converges in five or six iterations for resistivity or MT problems. We have demonstrated its use with resistivity, MT, and combined resistivity-MT data. However, our algorithm depends little on the nature of the forward function, and could be used for any problem where a smooth 1-D solution is required.

ACKNOWLEDGMENTS

This work was funded in part by a consortium of oil companies which includes Amoco, ARCO, and Elf Aquitane. Philip Stark critically read the manuscript.

REFERENCES

de Boor, C., 1978, A practical guide to splines: Springer-Verlag, Inc.
Constable, S. C., 1985, Resistivity studies over the Flinders conductivity anomaly: Geophys. J. Roy. Astr. Soc., **83**, 775–786.
Constable, S. C., McElhinny, M. W., and McFadden, P. L., 1984, Deep Schlumberger sounding and the crustal resistivity structure of central Australia: Geophys. J. Roy. Astr. Soc., **79**, 893–910.
Cull, J. P., 1985, Magnetotelluric soundings over a Precambrian contact in Australia: Geophys. J. Roy. Astr. Soc., **80**, 661–675.
Ghosh, D. P., 1971, Inverse filter coefficients for the computation of apparent resistivity standard curves for a horizontally layered earth: Geophys. Prosp., **19**, 769–775.
Gill, P. E., Murray, W., and Wright, M. H., 1981, Practical optimization: Academic Press, Inc.
Guptasarma, D., 1982, Optimization of short digital linear filters for increased accuracy: Geophys. Prosp., **30**, 501–514.
Inman, J. R., 1975, Resistivity inversion with ridge regression: Geophysics, **40**, 798–817.
Johansen, H. K., 1975, An interactive computer/graphic-display-terminal system for interpretation of resistivity soundings: Geophys. Prosp., **23**, 449–458.
Jones, A. G., and Hutton, R., 1979a, A multi-station magnetotelluric study in southern Scotland I. Fieldwork, data analysis and results: Geophys. J. Roy. Astr. Soc., **56**, 329–349.
——— 1979b, A multi-station magnetotelluric study in southern Scotland II. Monte-Carlo inversion of the data and its geophysical and tectonic implication: Geophys. J. Roy. Astr. Soc., **56**, 351–358.
Koefoed, O., 1970, A fast method for determining the layer distribution from the raised kernel function: Geophys. Prosp., **18**, 564–570.
Langer, R. E., 1933, An inverse problem in differential equations: Bull., Am. Math. Soc., ser. 2, **29**, 814–820.
Marquardt, D. W., 1963, An algorithm for least-squares estimation of non-linear parameters: J. Soc. Ind. Appl. Math., **11**, 431–441.
Parker, R. L., 1977, The Frechét derivative for the electromagnetic induction problem: Geophys. J. Roy. Astr. Soc., **49**, 543–547.
——— 1980, The inverse problem of electromagnetic induction: existence and construction of solutions based on incomplete data: J. Geophys. Res., **85**, 4421–4428.
——— 1982, The existence of a region inaccessible to magnetotelluric sounding: Geophys. J. Roy. Astr. Soc., **68**, 165–170.
——— 1984, The inverse problem of resistivity sounding: Geophysics, **49**, 2143–2158.
Parker, R. L., and Whaler, K. A., 1981, Numerical methods for establishing solutions to the inverse problem of electromagnetic induction: J. Geophys. Res., **86**, 9574–9584.
Petrick, W. R., Pelton, W. M., and Ward, S. H., 1977, Ridge regression inversion applied to crustal resistivity sounding from South Africa: Geophysics, **42**, 995 1005.
Russell, B., 1946, History of western philosophy: George Allen and Unwin, Ltd.
Shure, L., Parker, R. L., and Backus, G. E., 1982, Harmonic splines for geomagnetic modeling: Phys. Earth Plan. Int., **32**, 215–229.
Smith, D. R., 1974, Variational methods in optimization: Prentice-Hall, Inc.
Sternberg, B. K., 1979, Electrical resistivity of the crust in the southern extension of the Canadian shield—layered earth models: J. Geophys. Res., **84**, 212–228.
Tikhonov, A. N., and Arsenin, V. Y., 1977, Solutions of ill-posed problems: John Wiley and Sons, Inc.
Vozoff, K., and Jupp, D. L. B., 1975, Joint inversion of geophysical data: Geophys. J. Roy. Astr. Soc., **42**, 977–991.

APPENDIX

COMPUTATION OF THE JACOBIAN MATRIX

The inversion scheme developed here, and other inversion methods based on the linearization of nonlinear functions, depend critically on the ability to compute the Jacobian matrix. The most generally applicable way of estimating the required partial derivatives is to use a finite-difference technique. Gill et al. (1981) present a useful discussion of finite-difference methods. In particular, note that the central difference

$$\frac{F(m_i + \delta) - F(m_i - \delta)}{2\delta} = \frac{\partial F}{\partial m_i} + O(\delta^2) \qquad \text{(A-1)}$$

is accurate to second order, whereas the forward difference

$$\frac{F(m_i + \delta) - F(m_i)}{\delta} = \frac{\partial F}{\partial m_i} + O(\delta) \qquad \text{(A-2)}$$

is accurate only to first order. However, since $F(\mathbf{m})$ is required anyway and is common to all the $\partial F/\partial m_i$ ($i = 1, 2, \ldots, N$), computation of the partials matrix using equation (A-1) requires twice as many evaluations of the forward function as the computation using equation (A-2).

Errors in the estimation of $\partial F/\partial m_i$ lead to instabilities during the inversion, and so wherever possible analytical expressions should be used instead of finite-difference schemes. Analytical derivatives for layered models have the added bonus of often being much cheaper to compute, because usually $\partial F/\partial m_i$ can be computed from the intermediate results of evaluating $\partial F/\partial m_{i+1}$. We show this computation for the Schlumberger and MT problems. As an example of the computational savings possible, the partials matrices for the 45-layer Schlumberger inversions presented here took $3t$ s to

complete, where t is the time taken to do the forward calculation $\mathbf{F}(\mathbf{m})$. If a finite-difference algorithm were used, the partials matrix would take at least $45t$ s.

Derivatives for the Schlumberger problem

The forward calculation for Schlumberger apparent resistivities over a layered model may be written

$$\rho_a = \left(\frac{AB}{2}\right)^2 \int_0^\infty T_1(\lambda) J_1\left(\frac{AB}{2}\lambda\right)\lambda \, d\lambda, \qquad \text{(A-3)}$$

where J_1 is the first-order Bessel function of the first kind, $AB/2$ is the half-electrode spacing, and $T_1(\lambda)$ is the Koefoed resistivity transform (Koefoed, 1970). $T_1(\lambda)$ may be calculated from the recurrence relation

$$T_i = \frac{T_{i+1} + \rho_i \tanh(\lambda t_i)}{1 + T_{i+1}\tanh(\lambda t_i)/\rho_i}, \qquad \text{(A-4)}$$

where the ρ_i and the t_i are the layer resistivities and thicknesses and T_i is the transform evaluated at the top of the ith layer. Starting with $T_N = \rho_n$ at the top of the terminating half-space, equation (A-4) may be evaluated repeatedly to get $T_1(\lambda)$. It is common to evaluate equation (A-3) using the filter nethod (Ghosh, 1971), rewriting the equation as

$$\rho_a = \sum_{k=k_{\min}}^{k_{\max}} T_1(\lambda_k) f_k, \qquad \text{(A-5)}$$

where the f_k are coefficients of a moving average filter. There have been at least a dozen sets of filter coefficients published, and many of them perform very poorly. The 141-point filter of Johansen (1975) and the shorter filters of Guptasarma (1982) are recommended.

The derivatives $\partial\rho_a/\partial\rho_j$ may be obtained by differentiating equation (A-5) to get

$$\frac{\partial\rho_a}{\partial\rho_j} = \sum_k \frac{\partial T_1(\lambda_k)}{\partial\rho_j} f_k. \qquad \text{(A-6)}$$

To evaluate $\partial T_1/\partial\rho_j$, we write

$$\frac{\partial T_1}{\partial\rho_j} = \frac{\partial T_1}{\partial T_2}\frac{\partial T_2}{\partial T_3}\frac{\partial T_3}{\partial T_4}\cdots\frac{\partial T_{j-1}}{\partial T_j}\frac{\partial T_j}{\partial\rho_j}. \qquad \text{(A-7)}$$

Differentiating equation (A-4) gives expressions for $\partial T_i/\partial T_{i+1}$ and $\partial T_j/\partial\rho_j$:

$$\frac{\partial T_i}{\partial T_{i+1}} = \left[1 - \tanh^2(t_i\lambda_k)\right]\Big/ c_i,$$

and

$$\frac{\partial T_i}{\partial\rho_i} = \tanh(t_i\lambda_k)\left[1 + T_{i+1}^2/\rho_i^2 + 2T_{i+1}\tanh(t_i\lambda_k)/\rho_i\right]\Big/ c_i,$$

where

$$c_i = \left[1 + \tanh(t_i\lambda_k)T_{i+1}/\rho_i\right]^2.$$

The recursion may be started by noting that

$$\frac{\partial T_N}{\partial\rho_N} = 1.$$

Clearly, when evaluating $\partial T_1/\partial\rho_N$ using equation (A-7), the intermediate results $\partial T_1/\partial T_2$, $\partial T_2/\partial T_3$, ..., $\partial T_{N-1}/\partial T_N$ may be used to compute $\partial T_1/\partial\rho_{N-1}$, $\partial T_1/\partial\rho_{N-2}$, ..., $\partial T_1/\partial\rho_1$ with little extra computational effort. This results in great efficiency when entire rows of the partials matrix are computed at once.

Derivatives for the MT problem

The forward problem for the MT case is given by another recurrence relation:

$$c = \frac{1}{k_1}\coth\left[k_1 t_1 + \coth^{-1}\left(\frac{k_1}{k_2}\coth\left\{k_2 t_2 \cdots \coth\left[k_{N-1}t_{N-1} + \coth^{-1}\left(\frac{k_{N-1}}{k_N}\right)\right]\cdots\right\}\right)\right], \qquad \text{(A-8)}$$

where $k_j = (i2\pi f\mu_0/\rho_j)^{1/2}$ and f is the frequency of interest. The quantity c is given the unimaginative name of "complex c value" and is a natural way to present MT data. However, most workers in this field prefer to use apparent resistivity and phase, which are given by

$$\rho_a = 2\pi f\mu_0 c^2, \qquad \Phi = \tan^{-1}\left[\frac{\mathrm{Re}\,(c)}{-\mathrm{Im}\,(c)}\right],$$

where Re (c) and Im (c) are the real and imaginary parts of c. Before differentiating equation (A-8), we simplify the notation by defining

$$c_\ell \equiv f\left(k_\ell, f\left\{k_{\ell+1}, f\left[k_{\ell+2}, \ldots, f(k_{N-1}, 1/k_{N-1})\right]\cdots\right\}\right),$$

where $f(k_i, x) \equiv 1/k_i \coth[k_i t_i + \coth^{-1}(k_i x)]$. Now we may let $c = c_1$ and write equation (A-8) as

$$\frac{\partial c_1}{\partial k_j} = \frac{\partial c_1}{\partial c_2}\frac{\partial c_2}{\partial c_3}\frac{\partial c_3}{\partial c_4}\cdots\frac{\partial c_{j-1}}{\partial c_j}\frac{\partial c_j}{\partial k_j}. \qquad \text{(A-9)}$$

Now

$$\frac{\partial c_i}{\partial c_{i+1}} = \coth'\left[k_i t_i + \coth^{-1}(k_i c_{i+1})\right]\mathrm{arccoth}'(k_i c_{i+1})$$

and

$$\frac{\partial c_j}{\partial k_j} = -\frac{1}{k_j^2}\coth\left[k_j t_j + \coth^{-1}(k_j c_{j+1})\right] + \frac{1}{k_j}\coth'\left[k_j t_j + \coth^{-1}(k_j c_{j+1})\right] \times \left[t_j + \mathrm{arccoth}'(k_j c_{j+1})\, c_{j+1}\right],$$

where

$$\coth' = \frac{\partial \coth z}{\partial z} = -\left(\frac{2}{e^z - e^{-z}}\right)^2, \qquad z \neq 0,$$

and

$$\text{arccoth}' = \frac{\partial \coth^{-1} z}{\partial z} = (1 - z^2)^{-1}, \qquad z \neq 1.$$

Note also that $\partial c_N/\partial k_N = -1/k_N^2$. The derivatives with respect to layer resistivity are simply

$$\frac{\partial c_1}{\partial \rho_j} = \frac{\partial c}{\partial k_j}\frac{\partial k_j}{\partial \rho_j} = -\frac{1}{2}\left(\frac{i2\pi f\mu_0}{\rho_j^3}\right)^{1/2}\frac{\partial c}{\partial k_j}.$$

As in the Schlumberger case, the intermediate values used to compute $\partial c_1/\partial \rho_N$ may be used to compute $\partial c_1/\partial \rho_{N-1}$, $\partial c_1/\partial \rho_{N-2}$, etc. If apparent resistivity and phase data are being inverted, one will need to know that

$$\frac{\partial \rho_a}{\partial \rho_j} = 2\pi f\mu_0\left[\text{Re}\,(c_1)\,\text{Re}\left(\frac{\partial c_1}{\partial \rho_j}\right) + \text{Im}\,(c_1)\,\text{Im}\left(\frac{\partial c_1}{\partial \rho_j}\right)\right]$$

and

$$\frac{\partial \Phi}{\partial \rho_j} = \frac{1}{c_1^2}\left[\text{Re}\,(c_1)\,\text{Im}\left(\frac{\partial c_1}{\partial \rho_j}\right) - \text{Im}\,(c_1)\,\text{Re}\left(\frac{\partial c_1}{\partial \rho_j}\right)\right].$$