

COURSE 9

**INVERSION OF LOW-FREQUENCY
ELECTROMAGNETIC DATA**

T. MADDEN

*Department of Earth, Atmospheric
and Planetary Sciences
Massachusetts Institut of Technology
Cambridge, Massachusetts, USA*

*Y. Desaubies, A. Tarantola and J. Zinn-Justin, eds.
Les Houches, Session L, 1988
Tomographie Océanographique et Géophysique /
Oceanographic and Geophysical Tomography
© Elsevier Science Publishers B. V., 1990*

Contents

1. Introduction: magnetotellurics, the need for interpretation methods.	379
2. Magnetotelluric equations: Maxwell's equations in conducting media	382
3. The maximum likelihood inverse	384
4. Sensitivity operators	388
5. The bilinear identity: adjoint operators and reciprocity.	391
6. Relaxing the inverse problem	397
References	407

1. Introduction: magnetotellurics, the need for interpretation methods.

Seismology is without question the dominant geophysical tool used to study the Earth's interior. This is partly due to the existence of powerful natural sources which allows one to observe waves that have travelled through virtually all parts of the Earth. The fact that travel time inversion methods have been in existence for almost a century has, however, also been an important factor in placing seismology in its dominant position.

The same is true for smaller scale exploration especially in the search for petroleum. Geophysical exploration, in the petroleum industry is dominated by reflection seismology, especially in the United States. Reflection seismology is especially well suited to studying subtle variations in the geometry of layered sequences of rocks, and is thus an ideal tool for exploring for petroleum in typical sedimentary basins. This has led to great improvements in the technology of acquiring and interpreting reflection data and further distanced these methods from other geophysical exploration techniques which are usually lumped together as nonseismic methods. Gravity and magnetic field measurements are routinely made, especially in preliminary assessments of previously unexplored areas. These methods have also been greatly helped by instrumental developments. Modern gravity meters can measure changes in the Earth's gravity field of one part in a billion, and yet they are small enough to be backpacked and rugged enough to survive such treatment. The two dimensional nature of gravity and magnetic data limits the information content of such data.

Electrical methods, however, can provide three dimensional data. Some of the electrical methods are among the oldest geophysical exploration techniques and their use is dominant in mining exploration, especially in the search for sulphide deposits which are responsible for most of the supplies of copper, lead, and zinc. Their use in petroleum exploration is much more limited, with the exception of work done in Eastern Europe and Russia. This is partly due to the more limited resolution of such measurements, but the methods also suffer from their more sporadic use and from the lack of sophisticated methods for interpreting the data. There are areas in the world, however, with potential for petroleum resources, which are difficult

to explore by reflection seismic methods, and such areas are being probed with electrical measurements. These areas usually involve either near surface basalt layers which cause very strong reflections and greatly hinder viewing the layers beneath them, or areas where tectonic events have disrupted the sedimentary layer geometries and greatly complicated the seismic signature making the seismic interpretation very difficult. There has also been a small but steady amount of work done using electrical methods to study the upper mantle.

In mining exploration the targets are often fairly shallow and the electrical measurements are usually active measurements. In active measurements the fields being measured are set up by a controlled source. As the targets get deeper the anomalous signals that one is looking for get smaller and eventually these signals are lost in the background noise. In electrical measurements this background noise is due mostly to externally produced changes in the Earth's magnetic field. These fluctuating magnetic fields induce electric fields and electric currents in the Earth which are called telluric currents. When the targets of interest are two to ten kilometers deep, it is very difficult to produce with active systems fields whose anomalies can be detected in the presence of these telluric fields. In such a situation one can turn the problem around and consider the telluric fields themselves as the source field. This is the basis of the electrical exploration method known as magnetotellurics.

The method was independently developed in Russia by Tikhonov and in France by Cagniard (Tikhonov, 1950; Cagniard, 1953). The basis of the method is quite simple. The electromagnetic fields induced by the external magnetic field variations must damp out with depth. This is known as the skin effect and is due to the damped nature of electromagnetic waves in a conducting medium. Thus from Amperes law we have that the magnetic field fluctuation amplitudes at the surface of the Earth determines the total current flowing in the Earth underneath. Once one knows the amount of current, one can use the electric field magnitude to determine the Earth electrical resistivity. This determination must of course take into account the vertical spread of the current due to the damping effect. Different frequencies have different depths of penetration and this allows one to study variations of conductivity with depth. By making measurements over a two dimensional grid of sites, one can also get information about the horizontal variations of conductivity, and thus in principle develop a three dimensional picture of the Earth conductivity structure. The data is usually presented in terms of the apparent resistivity and the E/H phase as a function of frequency. The apparent resistivity, a parameter which is commonly used in all electrical soundings, is simply the resistivity of the homogeneous

Earth that would have given one the same magnitude of the observed E signal. In an inhomogeneous Earth the apparent resistivity varies with frequency as well as with position.

The early work in magnetotellurics was usually confined to making one dimensional interpretations. When one considers that most of the very sophisticated treatment of seismic reflection data called migration is concentrated on making rather minor modifications of the one dimensional interpretation, one might think the one dimensional interpretation of magnetotelluric data would be a good first cut. Unfortunately this is not the case, and magnetotellurics has suffered greatly from the poor results obtained from such interpretations. It is true that magnetotellurics is a wave phenomenon and thus has analogies with seismic methods, but one must use the analogies with care. Magnetotellurics is actually very closely related to the seismic phenomena of surface waves. In both phenomena the depth variation information is subtly contained in the frequency response. One often hears the statement that the damped nature of such waves reduces the depth resolution to some rather large fraction of the depth. When the data is crude, this is a correct measure of the resolution, but the better the data, the more one can refine the interpretation and increase the resolution.

In recent years great improvements have been made in the instrumentation and data collection techniques for magnetotellurics, but now a serious limitation arises because the interpretation technology is still inadequate. This is due in part to the fact that much smaller efforts have been put into magnetotellurics than have been put into seismic methods, but it is also due to the fact that the magnetotelluric interpretation problems are intrinsically more difficult. First of all magnetotellurics is a very three dimensional phenomenon, and secondly because the electrical contrasts occurring in nature are very large, many useful simplifications such as perturbation expansions are not applicable.

The three dimensionality of the interpretation problem shows up strongly in two very different but common situations. Because of the continuity of current flow, small near surface regions with anomalous conductivities cause large changes in the electric field magnitude over these regions. This effect is called a static shift as it occurs for all frequencies whose fields penetrate deeper than the anomalous regions. The term is borrowed from reflection seismology where an anomalous near surface velocity change will cause a time shift of all reflections coming from below the anomalous zone. The phase of the electric field over such superficial regions will not change, however, and if one has data from neighboring positions, the static shift is easily spotted. This problem is well recognized, and is thus not a severe limitation. Using larger dipoles or better still contiguous dipoles allows

one to virtually eliminate the static shift problem. There is another common problem, however, whose severity is not as well recognized. This is the regional problem. The continuity of current which caused the static shift was also responsible for the lack of phase changes which exposed the superficial nature of this shift. The continuity of current implies that every current filament has its own phase which is invariant along its entire path. Thus some sort of current mixing is required for phase changes to be observed as a function of position. When different regions have different conductivity depth structures they should have different magnetotelluric phases, but if the conductivity structure discourages current mixing these differences are blunted. The worst example of such effects are seen at ocean continent margins. The ocean is a very good conductor and most of the telluric current at periods of up to 12 hours in oceanic regions is in the sea water. On continents at such long periods most of the current is in the mantle. The lower crust is usually a very poor conductor because the rocks are too tight to hold much pore fluids and not hot enough for solid state conduction. This resistive lower crustal layer prevents upper crustal or sea water currents from mixing easily with mantle currents, and as a result the telluric fields near the coast that run perpendicular to the coastline have phases one would associate with the oceans. These phase anomalies can persist inland for many hundreds of kilometers.

In petroleum exploration one is usually not interested in mantle properties and very long period data is not collected. Nevertheless in regions devoid of thick sedimentary covers at periods as short as 1 to 10 seconds most of the telluric currents should reside in the mantle, and if a sedimentary basin is surrounded by such a region, the current system in the basin can be considerably affected. To deal with this problem one needs regional data and one needs to be able to cope with the three dimensional interpretation problem. These same effects can occur within the sedimentary section itself, although their severity will be greatly reduced, since lower crustal like resistivities are rarely seen in sedimentary rocks, the one exception being salt beds. In studies of the upper mantle, however, one probably is always involved with these regional effects and using shortcuts in the interpretation of the magnetotelluric data usually leads to incorrect conclusions about the mantle conductivity structures.

2. Magnetotelluric equations: Maxwell's equations in conducting media

Magnetotellurics, which we will abbreviate as MT, is a low frequency elec-

tromagnetic phenomenon and is thus described by Maxwell's equations in a conducting medium:

$$\begin{pmatrix} -\sigma \\ \nabla \times \end{pmatrix} \begin{pmatrix} \nabla \times \\ -i\mu\omega \end{pmatrix} \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} = \begin{pmatrix} \mathbf{J} \\ \mathbf{M} \end{pmatrix}. \quad (2.1)$$

\mathbf{J} represents electric current sources and \mathbf{M} represents magnetic sources which can be created by current loops.

In isotropic homogeneous media solutions take the form

$$\begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} \approx \exp i\mathbf{k} \cdot \mathbf{R} - i\omega t \quad (\mathbf{k}^2 = i\mu\omega\sigma). \quad (2.2)$$

These solutions are diffusion equation solutions. The exponential decay of the solutions away from the surface is known as the skin effect and the $1/e$ damping distance is known as the skin depth. When the source field has large horizontal dimensions we obtain for the relationship between \mathbf{E} and \mathbf{H} , which is called an impedance,

$$E_x/H_y = Z_{xy} = (-i\mu\omega/\sigma)^{1/2}. \quad (2.3)$$

Z_{yx} is the same but with the opposite sign. The impedance data is usually presented as an apparent resistivity, a practice that is common to many electrical measurement techniques:

$$\rho_{\text{apparent}} = | -\mathbf{Z}^2 / i\mu\omega |. \quad (2.4)$$

When the electrical properties change with depth, ρ_{apparent} will vary with frequency and these variations can be used to solve for the resistivity variations with depth. This is known as the ID interpretation and is commonly used to interpret MT data. Unfortunately it almost never gives one correct results because of the strong effects of non ID and non isotropic conductivity distributions. The least well known effects are those due to the anisotropy. Layered media with conductivity variations act essentially as transversely anisotropic media with vertical resistivities which are greater than the horizontal resistivities. This causes the two fundamental modes, the poloidal mode and the toroidal mode, to have different behaviors whenever the fields have some finite horizontal wavelengths. For the toroidal \mathbf{E} , poloidal \mathbf{H} mode one has in homogeneous media

$$k_z^2 = i\mu\omega\sigma - (k_x^2 + k_y^2) \quad (2.5)$$

but for the poloidal \mathbf{E} , toroidal \mathbf{H} mode one has

$$k_z^2 = i\mu\omega\sigma - \rho\sigma(k_x^2 + k_y^2) \quad (2.6)$$

where ρ stands for the vertical resistivity and σ stands for the horizontal conductivity. The product $\rho\sigma$ is always greater than 1 when the anisotropy is due to layering. The poloidal \mathbf{E} mode with finite horizontal wavelengths can be excited by inhomogeneities in the medium. The conductive upper crustal rocks and the resistive lower crust produce a very large effective anisotropy and boundaries such as sedimentary basin edges or the ocean continent boundary can strongly excite the poloidal \mathbf{E} mode over wide areas. For these reasons MT interpretation has strong needs of 2D and 3D inversion methods.

3. The maximum likelihood inverse

One can divide inversion or tomographic procedures into two classes. One class is the class of direct inversions where a well defined operation on the data gives as a result a model which can explain the data. When the data and model are linearly related one always has such a relationship, but it is possible to have direct inversions in problems where the data is not linearly related to the model. Many such inversions are closely allied to layer stripping algorithms in that the accumulated information about the medium above some depth is used to determine the properties at that depth. Travel time inversion, which has been used for almost a century, is in this category. Levy has shown that the Gelfand-Levitan inversion procedure is also identical to a layer stripping procedure (Yagle and Levy, 1985).

These methods have to date been confined to one dimensional problems, but in principle they could be developed for two and three dimensional problems. They have the disadvantage of accumulating errors with depth, and being direct methods, of having less flexibility to incorporate known information about the medium. They have a distinct speed advantage, however. These methods usually assume an infinite data set and imply that the solution is unique, although there may be litches like the discontinuity in the travel time curve when low velocity layers are present or the finite viewing angle effects in tomographic inversions. A mathematical uniqueness, however, does not translate into an actual uniqueness when the realities of finite data, data noise, and false premises are introduced.

The second class of inversion procedures covers a wide range, but the procedures usually all involve some form of iteration or search, which makes them slower but does allow for some flexibility in influencing the results. Because of the very real problem of nonuniqueness one has in practice (if not in theory) it usually behooves one to use other sources of information to guide one in the inversion. In principle one might consider collecting the full set of acceptable models. This is impractical, especially if the problem is non linear, but one can determine bounds on the models. For one dimensional problems, this is a reasonable approach as one can graph such results and get some understanding of the nonuniqueness problem. This type of information, however, is limited by not exposing the correlations between the effects of different model parameters which can prevent certain paths within the bounds from being allowed. For linear problems one can in principle examine the null space of the model parameters which would expose all these effects if one could handle all the information. For nonlinear problems one could consider using the perturbation null space. In practice, however, the cost of generating the null spaces is too great and the information is too overwhelming for such an approach to be practical. An approach which is practical and which can help expose some of the non uniqueness is one which makes use of a priori model information or conjectures to influence the solutions and to compare the results obtained using a finite set of different constraints. This approach makes use of a procedure which I like to call the maximum likelihood inverse, but which Tarantola and Valette originally called the total inverse (Tarantola and Valette, 1982). The procedure generates the solution which maximizes the joint probability of fitting the data and fitting an a priori model. The procedure is an example of linearized inversion iterations.

If the forward solution is denoted by \mathbf{Fm} where

$$\mathbf{F} = \text{forward problem operator} , \quad (3.1a)$$

$$\mathbf{m} = \text{model parameters} , \quad (3.1b)$$

and if the effects of model perturbations are denoted by $\mathbf{A}\delta\mathbf{m}$ where

$$\mathbf{A} = \text{sensitivity operator, or Fréchet derivative} , \quad (3.1c)$$

$$\delta\mathbf{m} = \text{model parameter perturbations} , \quad (3.1d)$$

one can move towards a solution of the inverse problem by finding a model perturbation that improves the model prediction's fit of the data. If the

model is not too far from a solution and if the forward problem is not too locally nonlinear, a $\delta\mathbf{m}$ that solves

$$\mathbf{A}\delta\mathbf{m} = \mathbf{d} - \mathbf{F}\mathbf{m} , \quad (3.2)$$

where

$$\mathbf{d} = \text{data} \quad (3.1c)$$

will give us a new model, $\mathbf{m} + \delta\mathbf{m}$, which gives an improved fit to the data. When the forward problem is nonlinear

$$\mathbf{F}(\mathbf{m} + \delta\mathbf{m}) \neq \mathbf{F}\mathbf{m} + \mathbf{A}\delta\mathbf{m} \quad (3.3)$$

so that the solution of eq. (3.2) will not be the exact model perturbation needed to fit the data. It is assumed, however, that when \mathbf{m} is close to a solution, $\delta\mathbf{m}$ will be small and the error in (3.3) will become vanishingly small.

It is possible to generate higher order derivatives which could be used to make a better estimate of $\delta\mathbf{m}$, but, as we shall show below, it is not practical. Instead one uses the solution obtained for (3.2) to update the model and then one repeats or iterates the procedure until a satisfactory fit to the data is reached. This sort of procedure is called Newton's method.

When the problem is non unique there are a variety of approaches to finding solutions to (3.2). One of these is a weighted generalized inverse solution which is also sometimes called the maximum likelihood solution (Aki and Richards, 1980). This solution minimizes

$$\delta\mathbf{m}^T \mathbf{R}_{mm}^{-1} \delta\mathbf{m} \quad (3.4a)$$

subject to

$$[\mathbf{d} - (\mathbf{F}\mathbf{m} + \mathbf{A}\delta\mathbf{m})]^T \mathbf{R}_{dd}^{-1} [\mathbf{d} - (\mathbf{F}\mathbf{m} + \mathbf{A}\delta\mathbf{m})] \quad (3.4b)$$

being a minimum, where

$$\mathbf{R}_{dd} = \text{data covariance}$$

and

$$\mathbf{R}_{mm} = \text{model covariance} .$$

I have often complained to Kei Aki about calling a solution maximum likelihood if the results give an unlikely model. In practice one can find a wide range of models that give almost the minimum value for (3.4b) and some of these may be much more likely than the solution given by (3.4). Thus it seems more reasonable to use the term maximum likelihood for the solution that maximizes the joint probability of fitting the data and adhering to a likely model. Assuming Gaussian statistics this is accomplished by finding the $\delta\mathbf{m}$ that minimizes

$$\begin{aligned} & [\mathbf{d} - (\mathbf{F}\mathbf{m} + \mathbf{A}\delta\mathbf{m})]^T \mathbf{R}_{dd}^{-1} [\mathbf{d} - (\mathbf{F}\mathbf{m} + \mathbf{A}\delta\mathbf{m})] \\ & + [\mathbf{m}_0 - (\mathbf{m} + \delta\mathbf{m})]^T \mathbf{R}_{mm}^{-1} [\mathbf{m}_0 - (\mathbf{m} + \delta\mathbf{m})] \end{aligned} \quad (3.5)$$

where

$$\mathbf{m}_0 = \text{a priori model parameters} .$$

The solution to eq. (3.5) is given by

$$\delta\mathbf{m} = [\mathbf{A}^T \mathbf{R}_{dd}^{-1} \mathbf{A} + \mathbf{R}_{mm}^{-1}]^{-1} [\mathbf{A}^T \mathbf{R}_{dd}^{-1} (\mathbf{d} - \mathbf{F}\mathbf{m}) + \mathbf{R}_{mm}^{-1} (\mathbf{m}_0 - \mathbf{m})] \quad (3.6)$$

and updating \mathbf{m} and iterating the procedure should lead to a final model which minimizes

$$(\mathbf{d} - \mathbf{F}\mathbf{m})^T \mathbf{R}_{dd}^{-1} (\mathbf{d} - \mathbf{F}\mathbf{m}) + (\mathbf{m}_0 - \mathbf{m})^T \mathbf{R}_{mm}^{-1} (\mathbf{m}_0 - \mathbf{m}) . \quad (3.5a)$$

When the operators are complex we must use $\mathbf{A}^{*T} = \mathbf{A}^H$ (complex conjugate transpose or adjoint) in place of \mathbf{A}^T and if \mathbf{m} is real we must use

$$\begin{aligned} \delta\mathbf{m} = & [\text{Re}\{\mathbf{A}^H \mathbf{R}_{dd}^{-1} \mathbf{A} + \mathbf{R}_{mm}^{-1}\}]^{-1} \\ & [\text{Re}\{\mathbf{A}^H \mathbf{R}_{dd}^{-1} (\mathbf{d} - \mathbf{F}\mathbf{m}) + \mathbf{R}_{mm}^{-1} (\mathbf{m}_0 - \mathbf{m})\}] . \end{aligned} \quad (3.6a)$$

In electrical problems where the model parameters and the data values can vary by orders of magnitude it is best to use the logarithm of the parameters and of the data. In this case since $\ln Z = \ln |Z| + i\theta$ one can divide the data and model predictions into in amplitude and phase contributions and keep all the terms real. The logarithmic parameterization is useful even when the parameters do not have a wide range as the procedure guarantees the positiveness of the model parameters and the maximum eigenvalue of the inverse operator is always of order one.

When the forward problem is linear only one iteration is needed to obtain the solution and in this first step since $\mathbf{m} = \mathbf{m}_0$, we have

$$\delta \mathbf{m} = [\mathbf{A}^T \mathbf{R}_{dd}^{-1} \mathbf{A} + \mathbf{R}_{mm}^{-1}]^{-1} \mathbf{A}^T \mathbf{R}_{dd}^{-1} (\mathbf{d} - \mathbf{F} \mathbf{m}_0). \tag{3.7}$$

It can be shown that

$$[\mathbf{A}^T \mathbf{R}_{dd}^{-1} \mathbf{A} + \mathbf{R}_{mm}^{-1}]^{-1} \mathbf{A}^T \mathbf{R}_{dd}^{-1} = \mathbf{R}_{mm} \mathbf{A}^T [\mathbf{A} \mathbf{R}_{mm} \mathbf{A}^T + \mathbf{R}_{dd}]^{-1} \tag{3.8}$$

when \mathbf{R}_{mm} and \mathbf{R}_{dd} are positive definite.

Starting with the identity

$$\mathbf{A}^T \mathbf{R}_{dd}^{-1} \mathbf{A} \mathbf{R}_{mm} \mathbf{A}^T + \mathbf{A}^T = \mathbf{A}^T \mathbf{R}_{dd}^{-1} \mathbf{A} \mathbf{R}_{mm} \mathbf{A}^T + \mathbf{A}^T$$

and using $\mathbf{R} \mathbf{R}^{-1} = \mathbf{I} = \mathbf{R}^{-1} \mathbf{R}$ we can write

$$\begin{aligned} \mathbf{A}^T \mathbf{R}_{dd}^{-1} \mathbf{A} \mathbf{R}_{mm} \mathbf{A}^T + \mathbf{A}^T &= \mathbf{A}^T \mathbf{R}_{dd}^{-1} \mathbf{R}_{dd} \mathbf{A}^T \\ &= \mathbf{A}^T \mathbf{R}_{dd}^{-1} \mathbf{A} \mathbf{R}_{mm} \mathbf{A}^T + \mathbf{R}_{mm}^{-1} \mathbf{R}_{mm} \mathbf{A}^T. \end{aligned}$$

This can be factored as

$$\mathbf{A}^T \mathbf{R}_{dd}^{-1} (\mathbf{A} \mathbf{R}_{mm} \mathbf{A}^T + \mathbf{R}_{dd}) = (\mathbf{A}^T \mathbf{R}_{dd}^{-1} \mathbf{A} + \mathbf{R}_{mm}^{-1}) \mathbf{R}_{mm} \mathbf{A}^T.$$

Premultiplying by the term $(\mathbf{A}^T \mathbf{R}_{dd}^{-1} \mathbf{A} + \mathbf{R}_{mm}^{-1})^{-1}$ and post multiplying by $(\mathbf{A} \mathbf{R}_{mm} \mathbf{A}^T + \mathbf{R}_{dd})^{-1}$ gives us (3.8).

The right-hand side of (3.8) is the stochastic inverse as defined by Franklin (1970). There can be some confusion about the symbolism since \mathbf{R}_{dd} as we use it here is what Franklin labels \mathbf{R}_{nn} (data noise variance). In this form one can deal with continuous model parameterization as long as the forward problem operator, which will now be an integral, is known (since the problem is assumed linear \mathbf{A} and \mathbf{F} are identical). The $\mathbf{A} \mathbf{R}_{mm} \mathbf{A}^T$ integrations can then be performed and $(\mathbf{A} \mathbf{R}_{mm} \mathbf{A}^T + \mathbf{R}_{dd})$ is reduced to a matrix whose dimensions are given by the number of data points. In general, except for 1D problems one has to discretize the medium in order to obtain the forward solution and the sensitivity operators and the issue of a continuous medium description disappears.

We will refer to the procedure (3.6) as the maximum likelihood inverse.

4. Sensitivity operators

The heart of the iterative inversion is the sensitivity matrix \mathbf{A} , which is determined by a perturbation analysis. The terms of the sensitivity

matrix are also called Fréchet derivatives. In our electromagnetic case the unperturbed medium generates solutions that satisfy Maxwell's equations, which, when using an $e^{-i\omega t}$ time dependence can be written as

$$\begin{pmatrix} -\sigma & \nabla \times \\ \nabla \times & -i\mu\omega \end{pmatrix} \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} = \begin{pmatrix} \mathbf{J} \\ \mathbf{0} \end{pmatrix}, \tag{4.1}$$

where \mathbf{J} are current sources which for the magnetotelluric problem are assumed to form a current sheet far above the Earth's surface. The perturbed medium leads to perturbed solutions

$$\begin{pmatrix} -(\sigma + \delta\sigma) & \nabla \times \\ \nabla \times & -i\mu\omega \end{pmatrix} \begin{pmatrix} \mathbf{E} + \delta\mathbf{E} \\ \mathbf{H} + \delta\mathbf{H} \end{pmatrix} = \begin{pmatrix} \mathbf{J} \\ \mathbf{0} \end{pmatrix}. \tag{4.2}$$

Subtracting (4.1) from (4.2) and dropping second order terms leaves us

$$\begin{pmatrix} -\delta\sigma & \nabla \times \\ \nabla \times & -i\mu\omega \end{pmatrix} \begin{pmatrix} \delta\mathbf{E} \\ \delta\mathbf{H} \end{pmatrix} = \begin{pmatrix} \delta\sigma\mathbf{E} \\ \mathbf{0} \end{pmatrix}. \tag{4.3}$$

Thus, the solution perturbation satisfies the same equations as the original solution, but with source terms proportional to the medium perturbations times the original field. If we denote the Green's function as $G_i(r, s)_j$, where i represents the observed field component, r represents the observation point, s represents the source point, and j represents the source component, we can write

$$\begin{pmatrix} \delta E_x(r) \\ \vdots \\ \delta H_z(r) \end{pmatrix} = \int \delta\sigma(s) \begin{pmatrix} G_1(r, s)_1 & G_1(r, s)_2 & G_1(r, s)_3 \\ \vdots & \vdots & \vdots \\ G_6(r, s)_1 & G_6(r, s)_2 & G_6(r, s)_3 \end{pmatrix} \times \begin{pmatrix} E_x(s) \\ E_y(s) \\ E_z(s) \end{pmatrix} ds. \tag{4.4}$$

Thus, the elements of the sensitivity matrix involve $\mathbf{G}(r, s) \cdot \mathbf{E}(s)$. It would appear that one needs to compute a Green's function for every point in the medium where one is trying to determine the properties, but since the observation points are limited to the surface we can use reciprocity to greatly reduced the number of Green's functions that need to be determined. Even so, for three dimensional problems, where we have two-dimensional arrays of observation points, computing the sensitivity matrix is a large task. The reciprocity relationships (Lanczos, 1960), arise from the relationship

between the field equations (in our case Maxwell's equations) and their adjoint (Hermitian adjoint as defined by Morse and Feshbach, 1953).

If the Green's function of the adjoint is given as \underline{G} , then we have

$$\underline{G}^*(r, s)_k = G_k(s, r)_j. \quad (4.5a)$$

Since real constants and curl operators are self adjoint, the adjoint to Maxwell's equations are also electromagnetic equations, but with backwards time or negative frequencies. In the frequency domain the adjoint EM equations are simply complex conjugated EM equations. Therefore, we can rewrite (4.5a) as

$$G_j(r, s)_k = G_k(s, r)_j. \quad (4.5b)$$

Equation (4.5b) is not true in general but it is applicable to all the field equations used in geophysical exploration.

Because of the large variations in electrical properties that occur in the Earth it is better to use the logarithm of the conductivity or resistivity as parameters. This also guarantees the positiveness of the parameters and removes the bias associated with the choice of using resistivities or conductivities as parameters. The same holds true for the data, especially since the data is complex, as the logarithm separates into log amplitude and phase. Phase is a very important data parameter in magnetotellurics. Using the logarithmic parameterization, the terms of the sensitivity matrix take the form

$$A_{jk} = \frac{\sigma(s)G(r_j, s_k) \cdot E(s_k)}{E(r_j)}. \quad (4.6)$$

The error terms take the form $\ln(d/Fm)$. The \mathbf{A} terms are frequency dependent so that the column space of \mathbf{A} is increased to accommodate A_{jk} terms for every measurement frequency.

The same perturbation scheme can be used to establish all the higher order derivatives which opens up the possibility of going beyond linearized inversions, but for the higher order derivatives we need the full set of Green's functions and cannot make much use of reciprocity which would make the use of higher order derivatives impractical:

$$\begin{aligned} \frac{\partial^2 E(r_k)}{\partial \sigma(s_i) \partial \sigma(s_j)} &= \frac{\partial(G(r_k, s_j) \cdot E(s_j))}{\partial \sigma(s_i)} \\ &= \frac{\partial G(r_k, s_j)}{\partial \sigma(s_i)} \cdot E(s_j) + G(r_k, s_j) \cdot \frac{\partial E(s_j)}{\partial \sigma(s_i)} \\ &= G(r_k, s_i) \cdot G(s_i, s_j) \cdot E(s_j) \\ &\quad + G(r_k, s_j) \cdot G(s_j, s_i) \cdot E(s_i). \end{aligned} \quad (4.7)$$

The $\mathbf{A}^H \mathbf{R}_{dd}^{-1}$ (error) term of the maximum likelihood inverse (3.6) can now be seen to be an imaging operation which is similar to the Kirchhoff migration operator as used in reflection seismology. If the starting Earth model is homogeneous, no reflections would be predicted and the term $(\mathbf{d} - \mathbf{Fm}_0)$ would simply be the reflection seismogram. Since A_{jk} involves $G(r_j, s_k)$ and since $G(r_j, s_k)$ is equal to $G(s_k, r_j)$ the $\mathbf{A}^H \mathbf{d}$ operation is a sum over the receiver locations of the reflection data times the complex conjugate of the Green's function for surface sources. This is therefore a propagation in backwards time of the reflection data back into the Earth. It is not, however, the same as downwards continuation since the waves are damped and spread in the process while they are amplified and converged in downward continuation. The $G^*(r_j, s_k)$ term provides a negative delay as does the $E^*(s_k)$ term, and the sum of these delays is just equal to but opposite in sign to the delay of the reflection when the s_k point is at the same location as the source of the reflection. The operation also involves summing over all frequencies and since all frequencies will be at zero phase at the reflection point, the $\mathbf{A}^* \mathbf{T} \mathbf{D}$ operation images the sources of the reflections. The coherent summing at the image point is maintained when the \ln parameterization form of (2) is used. The intensity of the image is related to the intensity of the reflection, but because the migration is not downwards continuation and because of the $E^*(s_k)$ term, the amplitudes are distorted. The $[\mathbf{A}^T \mathbf{R}_{dd}^{-1} \mathbf{A} + \mathbf{R}_{mm}^{-1}]^{-1}$ term in (3.6) carries the necessary amplitude information that allows the solution of (3.6) to give us a good estimate of the medium perturbations. The term "filtered backprojection" is often used in describing tomographic inversions and it is also applicable to the maximum likelihood inverse, but since the operation involves a matrix inverse the word filtered carries a somewhat false connotation.

5. The bilinear identity: adjoint operators and reciprocity.

In order to establish the reciprocity relationships one must look at the solution of the adjoint problem (some texts use the term adjoint for what we have called transpose and Hermitian adjoint for what we call adjoint. We will use the term adjoint and the symbol \mathbf{A}^H to represent \mathbf{A}^{*T} .) For matrix operators the bilinear identity is given as

$$(\mathbf{u}^H \mathbf{D} \mathbf{v}) = (\mathbf{v}^H \mathbf{D}^H \mathbf{u})^* \quad (5.1)$$

or

$$(\mathbf{u}^T \mathbf{D} \mathbf{v}) = (\mathbf{v}^T \mathbf{D}^T \mathbf{u}).$$

Since $\mathbf{u}^T \mathbf{D} \mathbf{v}$ is a scalar the bilinear identity seems to be only a statement that the transpose of a scalar is equal to itself. Actually it is much more than this, and since (5.1) must hold for any arbitrary \mathbf{u} and \mathbf{v} it is actually a definition of \mathbf{D}^H . The bilinear identity must be modified when dealing with differential operators, as the operators are not completely defined without some boundary conditions. We can examine this by considering the finite difference approximation to the operator. Consider the one dimensional wave equation

$$\frac{\partial^2 \mathbf{H}}{\partial z^2} + k^2 \mathbf{H} = 0 \rightarrow \mathbf{H}_{i-1} + (k^2 \Delta z^2 - 2) \mathbf{H}_i + \mathbf{H}_{i+1} = 0, \quad (5.2)$$

if the boundary value \mathbf{H}_0 is given the first equation will be

$$(k^2 \Delta z^2 - 2) \mathbf{H}_1 + \mathbf{H}_2 = -\mathbf{H}_0. \quad (5.3a)$$

On the other hand if $\partial \mathbf{H} / \partial z$ is given, the first equation will be

$$(k^2 \Delta z^2 - 1) \mathbf{H}_1 + \mathbf{H}_2 = (\partial \mathbf{H} / \partial z)_0 \Delta z. \quad (5.3b)$$

Thus we see that the difference equation operator is changed by the boundary condition and the adjoint operator has its boundary conditions (not the boundary values) defined. The adjoint of (5.3a) will be the same as (5.3a) if k^2 is real (self adjoint) and will also have \mathbf{H} as a boundary condition. (5.3b) is also self adjoint. If both \mathbf{H}_0 and $(\partial \mathbf{H} / \partial z)_0$ are given, \mathbf{H}_1 is taken out of the vector space and the operator becomes rectangular (overdetermined). In this case the adjoint operator will not have an equation centered on z_1 and it will have no boundary condition (underdetermined). $\mathbf{u}^H \mathbf{D} \mathbf{v}$ in the differential form are integral operations and they require boundary terms in order to be fully defined. Thus the bilinear identity will not be identically zero as the boundary terms remain. In this case the bilinear identity takes the form

$$\iiint (\mathbf{u}^T \mathbf{D} \mathbf{v} - \mathbf{v}^T \mathbf{D}^T \mathbf{u}) dV = \iint (\text{boundary terms}) dS. \quad (5.4)$$

This means that the bilinear term must be an exact differential or in the general case it must be a pure divergence. The divergence condition identifies the adjoint operator and the boundary terms identify the adjoint boundary conditions. The rule for the boundary conditions can be stated as: if the boundary conditions are homogeneous the adjoint boundary conditions are those conditions that are necessary and sufficient to make the boundary terms vanish.

Consider again the one dimensional wave equation. Since

$$\mathbf{u} \cdot \left(\frac{\partial^2 \mathbf{v}}{\partial z^2} + k^2 \mathbf{v} \right) - \mathbf{v} \cdot \left(\frac{\partial^2 \mathbf{u}}{\partial z^2} + k^2 \mathbf{u} \right) = \frac{\partial \left(\mathbf{u} \frac{\partial \mathbf{v}}{\partial z} - \mathbf{v} \frac{\partial \mathbf{u}}{\partial z} \right)}{\partial z},$$

we have

$$\begin{aligned} \int_{\text{top}}^{\text{bottom}} \left[\mathbf{u} \cdot \left(\frac{\partial^2 \mathbf{v}}{\partial z^2} + k^2 \mathbf{v} \right) - \mathbf{v} \cdot \left(\frac{\partial^2 \mathbf{u}}{\partial z^2} + k^2 \mathbf{u} \right) \right] dz \\ = \left(\mathbf{u} \frac{\partial \mathbf{v}}{\partial z} - \mathbf{v} \frac{\partial \mathbf{u}}{\partial z} \right) \Big|_{\text{top}}^{\text{bottom}}. \end{aligned} \quad (5.5)$$

Thus the adjoint operator is also a wave equation and the adjoint boundary conditions are the ones that were discussed before. If \mathbf{v} or $\partial \mathbf{v} / \partial z$ is given as a boundary condition the operator is self adjoint and \mathbf{u} or $\partial \mathbf{u} / \partial z$ is given as a boundary condition for the adjoint operator. If both \mathbf{v} and $\partial \mathbf{v} / \partial z$ are given, the adjoint operator has no boundary conditions.

Examples of the transpose and boundary terms of common differential operators are given in table 1.

Table 1
Common differential operators.

D	D^T	Boundary terms
$\nabla \phi$	$-\nabla \cdot \mathbf{V}$	$\iint \phi \mathbf{V} \cdot dS$
$\nabla \cdot \mathbf{V}$	$-\nabla \phi$	$\iint \phi \mathbf{V} \cdot dS$
$\nabla \times \mathbf{V}$	$\nabla \times \mathbf{U}$	$\iint \mathbf{V} \times \mathbf{U} \cdot dS$
$\nabla^2 \mathbf{v}$	$\nabla^2 \mathbf{u}$	$\iint \left(\mathbf{u} \frac{\partial \mathbf{v}}{\partial n} - \mathbf{v} \frac{\partial \mathbf{u}}{\partial n} \right) \cdot dS$

For Maxwell's equations, using $\begin{pmatrix} \mathbf{e} \\ \mathbf{h} \end{pmatrix}$ for \mathbf{u} and $\begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix}$ for \mathbf{v} , we have

$$\mathbf{u} \cdot \mathbf{D} \mathbf{v} = -\mathbf{e} \cdot \sigma \mathbf{E} + \mathbf{e} \cdot \nabla \times \mathbf{H} + \mathbf{h} \cdot \mathbf{h} \cdot i\mu\omega \mathbf{H}. \quad (5.6)$$

Since

$$\mathbf{U} \cdot \nabla \times \mathbf{V} - \mathbf{V} \cdot \nabla \times \mathbf{U} = \text{div}(\mathbf{V} \times \mathbf{U}) ,$$

and

$$\mathbf{U} \cdot a\mathbf{V} - \mathbf{V} \cdot a\mathbf{U} = 0 ,$$

we find that curls and real constants are self adjoint. Thus the adjoint of Maxwell's equations are also E.M. equations but with backward time or negative frequencies (Maxwell's equations are equal to their own transpose). The bilinear identity takes the form

$$\begin{aligned} & \iiint \left[\begin{pmatrix} e \\ h \end{pmatrix} \cdot \begin{pmatrix} -\sigma & \nabla \times \\ \nabla \times & i\mu\omega \end{pmatrix} \begin{pmatrix} E \\ H \end{pmatrix} - \begin{pmatrix} E \\ H \end{pmatrix} \cdot \begin{pmatrix} -\sigma & \nabla \times \\ \nabla \times & i\mu\omega \end{pmatrix} \begin{pmatrix} e \\ h \end{pmatrix} \right] dV \\ & = \iint (\mathbf{E} \times \mathbf{h} - \mathbf{e} \times \mathbf{H}) \cdot d\mathbf{S} . \end{aligned} \tag{5.7}$$

In MT the usual boundary conditions are impedance boundary conditions:

$$\begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{pmatrix} 0 & Z_{xy} \\ Z_{yx} & 0 \end{pmatrix} \begin{pmatrix} H_x \\ H_y \end{pmatrix} . \tag{5.8}$$

In this case, the adjoint boundary conditions are the same and these boundary conditions are also homogeneous as they cause the boundary term in (5.7) to vanish.

The bilinear identity, since it holds for all \mathbf{u} and \mathbf{v} that satisfy their respective equations is a powerful statement that can be molded to formulate many special identities. For instance it can be used to expose the compatibility conditions of the source terms.

Consider Poisson's equation with Neumann boundary conditions

$$\left(\frac{\partial \mathbf{v}}{\partial n} \right) .$$

This is a self adjoint operator and the null space solutions of the adjoint problem are solutions of

$$\nabla^2 \mathbf{u}_0 = 0 \tag{5.9a}$$

with

$$\frac{\partial \mathbf{u}_0}{\partial n} = 0 . \tag{5.9b}$$

If

$$\nabla^2 \mathbf{v} = \mathbf{b} ,$$

the compatibility condition for \mathbf{b} is given by the bilinear identity using \mathbf{u}_0 :

$$\iiint \mathbf{u}_0 \cdot \mathbf{b} \, dV = \iint \mathbf{u}_0 \frac{\partial \mathbf{v}}{\partial n} \cdot d\mathbf{S} , \tag{5.10a}$$

or, since only $\mathbf{u}_0 = \text{constant}$ are null space solutions,

$$\iiint \mathbf{b} \, dV = \iint \frac{\partial \mathbf{v}}{\partial n} \cdot d\mathbf{S} . \tag{5.10b}$$

In electrical engineering, a good deal of use is made of Tellegen's theorem (Penfield et al, 1970), which is the bilinear identity applied to the operator divergence ($\nabla \cdot$) and its adjoint the negative gradient ($-\nabla$). When $\text{div } \mathbf{J} = 0$ but $\mathbf{J}_n \neq 0$ we have

$$\iiint -\mathbf{J} \cdot \nabla \vartheta \, dV = \iint \vartheta \mathbf{J}_n \cdot d\mathbf{S} ,$$

or

$$\iiint \mathbf{E} \cdot \mathbf{J} \, dV = \iint \vartheta \mathbf{J}_n \cdot d\mathbf{S} , \tag{5.11a}$$

when

$$\mathbf{E} = -\nabla \vartheta . \tag{5.11b}$$

This looks like an energy theorem in that the energy dissipation $\mathbf{E} \cdot \mathbf{J}$ is equal to the energy input across the boundaries $\vartheta \mathbf{J}_n$. However, it is more general than this and is actually a virtual energy theorem since the \mathbf{E} field and the \mathbf{J} field do not have to come from the same experiment, and in fact they do not even have to come from the same media, nor do the media have to be linear.

The bilinear identity also exposes the reciprocity relationships of the Green's functions. For vector equations we must deal with a matrix of Green's functions or a dyadic. Thus for the Fréchet derivatives of Maxwell's equations as written in (4.4) we would have terms like $G_3(r, s)_4$. In this case, $G_3(r, s)_4$ is the z component of the electric field at (r) due to a magnetic dipole at (s) oriented in the x direction.

The Green's functions satisfy homogeneous equations except for a delta function at the source location in the equation involving the source component:

$$D \begin{pmatrix} G_1(\zeta, a)_j \\ \vdots \\ G_m(\zeta, a)_j \end{pmatrix} = D\mathbf{G}(\zeta, a)_j = \begin{pmatrix} (0)_1 \\ \vdots \\ \delta(\zeta, a)_j \\ \vdots \\ (0)_n \end{pmatrix}. \quad (5.12)$$

If we label the Green's function of the adjoint operator as $\underline{\mathbf{G}}$,

$$D^H \begin{pmatrix} \underline{G}_1(\zeta, b)_k \\ \vdots \\ \underline{G}_n(\zeta, b)_k \end{pmatrix} = D^H \underline{\mathbf{G}}(\zeta, b)_k = \begin{pmatrix} (0)_1 \\ \vdots \\ \delta(\zeta, b)_k \\ \vdots \\ (0)_n \end{pmatrix}, \quad (5.13a)$$

or, since the delta function is real,

$$D^T \begin{pmatrix} \underline{G}_1^*(\zeta, b)_k \\ \vdots \\ \underline{G}_n^*(\zeta, b)_k \end{pmatrix} = D^T \underline{\mathbf{G}}^*(\zeta, b)_k = \begin{pmatrix} (0)_1 \\ \vdots \\ \delta(\zeta, b)_k \\ \vdots \\ (0)_n \end{pmatrix}. \quad (5.13b)$$

When the Green's functions satisfy homogeneous boundary conditions, the boundary term of the bilinear identity vanishes and we are left with

$$\iiint \left(\sum_{i=1}^m \underline{G}_i^*(\zeta, b)_k [D\mathbf{G}(\zeta, a)_j]_i - \sum_{i=1}^n G_i(\zeta, a)_j [D^T \underline{\mathbf{G}}^*(\zeta, b)_k]_i \right) dV(\zeta) = 0, \quad (5.14)$$

which, using (5.12) and (5.13b) gives

$$\underline{G}_j^*(a, b)_k - G_k(b, a)_j = 0. \quad (5.15)$$

Thus, reciprocity involves not only the *interchange of source and receiver position*, but also the *interchange of the adjoint function, complex conjugation*, and the *interchange of source and receiver orientation*. Since Maxwell's equations are symmetric but not Hermitian, and if we use self adjoint boundary conditions such as impedance conditions, $\underline{\mathbf{G}}^* = \mathbf{G}^T$, we have

$$\underline{G}_j^*(a, b)_k = G_k(b, a)_j. \quad (5.16)$$

However, this equation is not true in general.

The reciprocity relationship (5.16) for Maxwell's equations can be simply described if we use grounded wires to measure the electric field and coils to measure the magnetic field and grounded wires as electric dipole sources and coils as magnetic dipole sources. In this case the reciprocity geometry is given by simply interchanging the current source and the voltage receiver, leaving all the coils or wires fixed, and the reciprocity relationship can be stated as the equality of the V/I ratios obtained before and after the interchange (fig. 1).

The reciprocity allows one to compute sensitivity operators using only Green's functions for sources at the locations where the measurements are made, which takes a whole dimension out of the computational problem. For 1D problems only a single function is needed, for 2D problems only a line of source position functions are needed, and for 3D problems only a surface array of source position functions are needed.

In some applied source measurement techniques, such as dipole-dipole resistivity measurements, sources and receivers have occupied the same locations, and in such cases in doing the forward problems one generates all the terms of the sensitivity operator needed for the inverse problem.

The higher order sensitivity operators require Green's functions with source and receiver positions throughout the volume, and here reciprocity only reduces the computational load by a factor of two. Generating the terms involving products of Green's functions is also a time consuming operation.

6. Relaxing the inverse problem

Even though reciprocity greatly reduces the work in computing the sensitivity matrix terms, computing Green's functions for a two dimensional array of receiver points is still a large task. Furthermore generating $\mathbf{A}^H \mathbf{R}_{dd}^{-1} \mathbf{A}$ given \mathbf{A} is another large task, and inverting $(\mathbf{A}^H \mathbf{R}_{dd}^{-1} \mathbf{A} + \mathbf{R}_{mm}^{-1})$ is a still

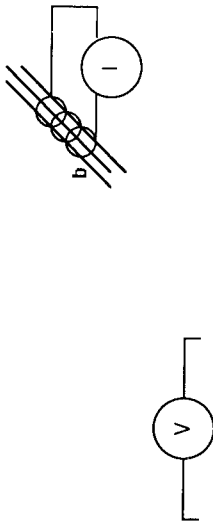
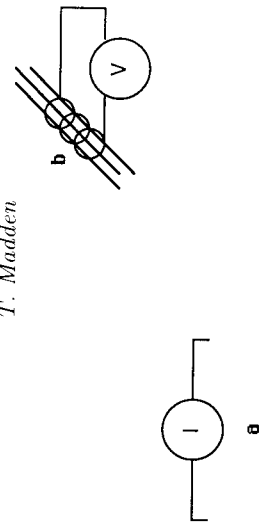


Fig. 1. Electromagnetic reciprocity. For the top panel, $V/I = G_5(b, a)_1$, and for the bottom panel $V/I = G_1(a, b)_5$.

larger task. Relaxation procedures can circumvent most of these tasks, however, and thus make three dimensional inversions practical. This possibility comes about because of two factors.

First of all in obtaining relaxation solutions to an equation

$$Ax = b \tag{6.1}$$

it is not necessary to know A , but only what Ax generates. Secondly, since each term in A involves a Green's function times some known factors, each operation involving A can be viewed as an electromagnetic problem involving a known source distribution. Thus if we have a forward problem procedure for solving the electromagnetic problem we do not need to compute the matrix of Green's functions. Claerbout's difference equation migration is an example of such a procedure (Claerbout, 1976).

If we rewrite (3.6) as

$$[A^T R_{dd}^{-1} A + R_{mm}^{-1}] \delta m = [A^T R_{dd}^{-1} (d - Fm) + R_{mm}^{-1} (m_0 - m)] \tag{6.2}$$

each relaxation step requires computing

$$y = Ax \tag{6.3a}$$

and then

$$A^H R_{dd}^{-1} y + R_{mm}^{-1} x \tag{6.3b}$$

which represents two forward problems. However, since once (6.2) is solved, one has to update the model and iterate the procedure, it does not seem necessary to obtain very accurate solutions to (6.2) until perhaps near the end of the iterative procedure.

To study the behavior of such relaxation procedures we have applied them to one and two dimensional inversions. We have no great faith in one dimensional inversions, and our relaxation tests of 1D examples were meant as simple tests of the relaxation procedure. In these tests we carried out only a single relaxation for each iteration, the relaxation step using the method of steepest descent. The results were amazing, but, as we shall see later, somewhat misleading. The simple relaxation procedure seemed to do as well as the full matrix inversion procedure. Figure 2 shows the progression of the relaxation iterations for a synthetic case involving a simple upper crust, lower crust, mantle model. Figure 3 compares the inversion convergence using the one relaxation step to the full matrix inversion procedure.

In figure 3 we do see some divergence of the two inversion procedures at the late stages, but by then the error terms are quite small. With more complicated models this divergence occurs at much larger errors and it will force one to use more relaxation steps and more efficient relaxation procedures.

The relaxation procedures are really only needed for the three dimensional problems, but then we will probably not be able to generate standard inversion results for comparisons. Two dimensional problems are much richer than one dimensional problems and they are still small enough to allow one to use standard procedures which can then be compared to the relaxation results. The presence of the $A^H A$ term in (6.2) causes problems for relaxation procedures because of the squaring of the eigenvalue spectrum of A . There are relaxation procedures which find the least squared

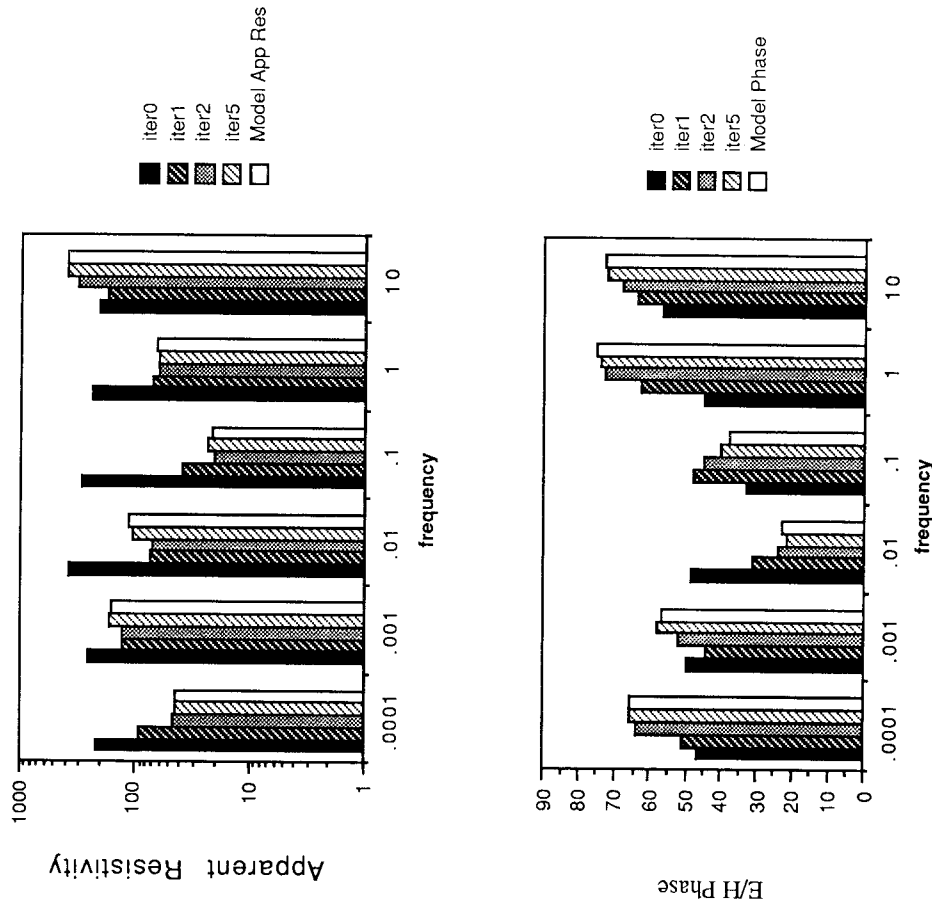


Fig. 2. Convergence of 1D inversion where each iteration step used a single relaxation.

error solution to $Ax = b$ without generating $A^H A$, but the presence of the $R_{mm}^{-1}(m_0 - m)$ term in (6.2) invalidates these methods (we are not solving $A\delta m = d - Fm$) and we do not want to abandon the maximum likelihood solution. Thus we find it is important to only use efficient relaxation procedures.

In figure 4 we show a simple model which can, however, cause some difficulties in the inversion since the major feature to be determined is the shape of the valley floor and magnetotellurics has the most problems with locating poor conductors. This model has 99 parameters, but, due to the fact that the model covariance matrix was set to put a strong 1D

Standard and Relaxation Inversion Convergence

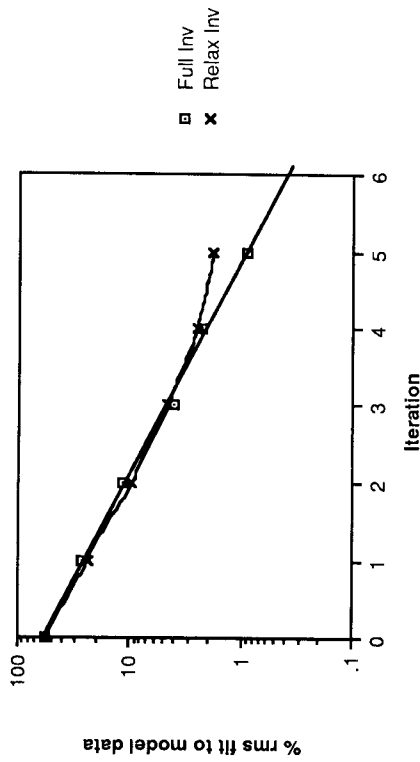
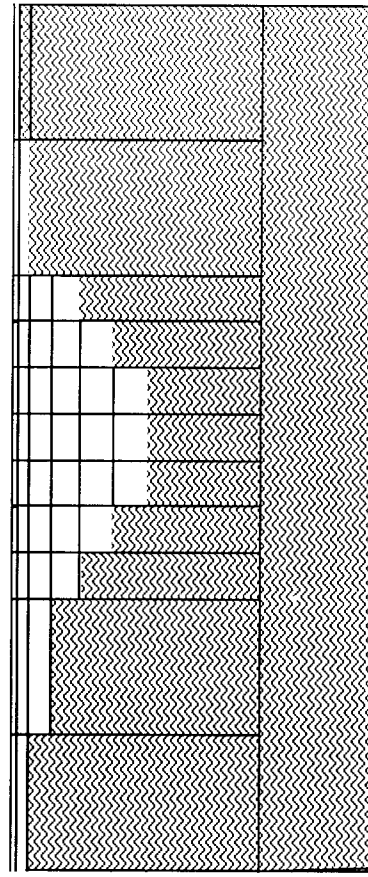


Fig. 3. Comparison of inversion convergence using one step relaxation and full matrix inversion steps for 1D MT example.



Model 5
2:1 vertical exaggeration

Fig. 4. 2D magnetotelluric model of sedimentary basin.

constraint on the region under the basement, it is really more like an 89 parameter model. Seven frequencies were used, two every decade from 3 to 10^{-1} Hz and one every decade from 10^{-2} to 10^{-4} Hz, so there are 154 data values. The final interpretation which produced a 1% rms fit to the data is shown in figure 5. The only significant errors in the model are the basement resistivity values and a somewhat fuzzy identification of the top of the basement in the middle of the valley.

depth in km

5 5 5 5 5 5 5 5 5 5 5
 5 5 5 5 5 5 5 5 5 5 5
 3000 5 5 5 5 5 5 5 5 5 5 5
 3000 3000 5 5 5 5 5 5 5 5 5 5
 3000 3000 3000 5 5 5 5 5 5 5 5 5
 3000 3000 3000 3000 5 5 5 5 5 5 5 5

9999 9999 9999 9999 9999 9999 9999 9999 9999 9999 9999

Model.5 resistivities

5 5 5 5 5 5 5 5 5 5 5 6
 6 5 5 5 5 5 5 5 5 5 5 6 195
 206 6 5 5 5 5 5 5 5 5 5 228 698
 433 89 7 4 5 5 4 4 4 8 537 1312
 653 178 42 12 5 6 7 14 34 932 2074
 594 187 49 16 5 4 4 12 33 1050 2331
 3944 2041 1192 1243 1375 1589 1886 2261 2705 8659 XXXX
 1593 5045 2407 2827 3295 3773 4219 4581 4804 XXXX XXXX

Model.5 standard inversion

5 5 5 5 5 5 5 5 5 5 5 6
 6 5 5 5 5 5 5 5 5 5 5 6 196
 213 6 5 5 5 5 5 5 5 5 5 185 704
 452 92 7 4 5 5 5 5 4 7 534 1334
 686 193 47 12 5 6 6 12 43 952 2080
 630 210 58 19 7 4 6 15 56 1085 2325
 4851 2163 1175 1247 1433 1684 2061 2553 3343 XXXX XXXX
 1642 5641 2325 2778 3294 3802 4304 4704 4976 XXXX XXXX

Model.5 relaxation inversion

Inversion results for Model 5.

Inversion of model.5

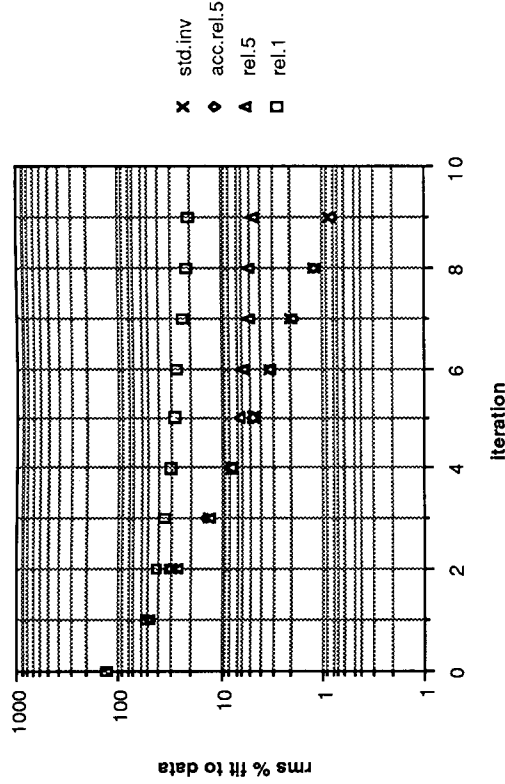
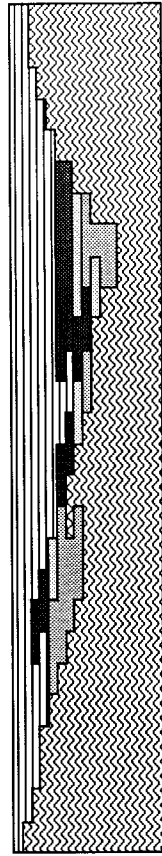


Fig. 6. Convergence of the inversion iterations for Model 5.



2-D Model OV.2

Fig. 7. Model Ov 2.

The convergence of the inversion iterations are shown in figure 6. We see here that the simple single relaxation for each iteration which worked for 1D problems is inadequate for the 2D problem. Using 5 conjugate gradient steps for each iteration improved the convergence, but it was still inadequate.

Relaxation methods have difficulty dealing with the small eigenvalue eigenvectors as these vectors are poorly represented in the residuals of the problem or to filter out the small eigenvalue vectors. Improving the eigenvalue spread is called preconditioning, but the preconditioning can be implemented by filtering each relaxation step (Golub and Van Loan, 1983) and, again, we can avoid having to evaluate the A or $A^H A$ opera-

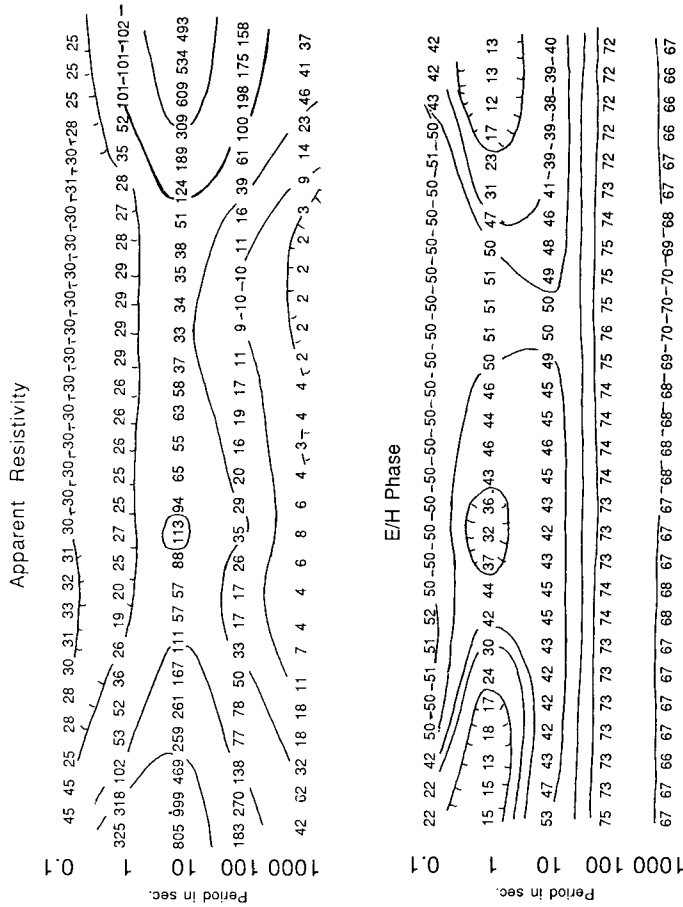


Fig. 8. MT pseudosection for model Ov.2.

tors. For the 2D MT inversion we used a matrix of subblocks, each block being the 1D inverse appropriate for each resistivity column of the model, as a preconditioner. Using this preconditioner with 5 conjugate gradient relaxation steps gave results quite comparable to the results obtained by a direct matrix inversion procedure.

Model 5, that was used in this study, although two dimensional, is still a rather small model and may not be a good test of what to expect in three dimensional problems. In figure 7 we show a somewhat larger and more complicated model with a vertical exaggeration of 1.6:1. This model is again a sedimentary basin, but there is some conductivity structure in the layering and these layers as well as the basement have been disrupted along several low angle faults. The data involves 5 frequencies between 10 and .001 Hz, so that we have 270 data points and again using a strong 1D constraint on the sub-basement structure we have essentially 325 parameters in our block representation of the model. Figure 8 shows the model data

Inversion of "Model.ov2"

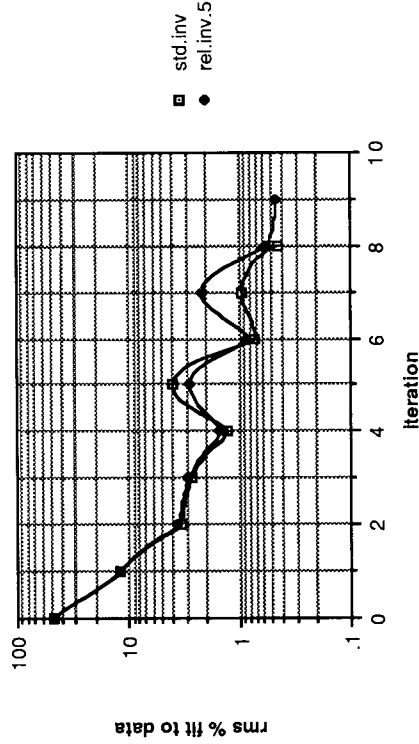
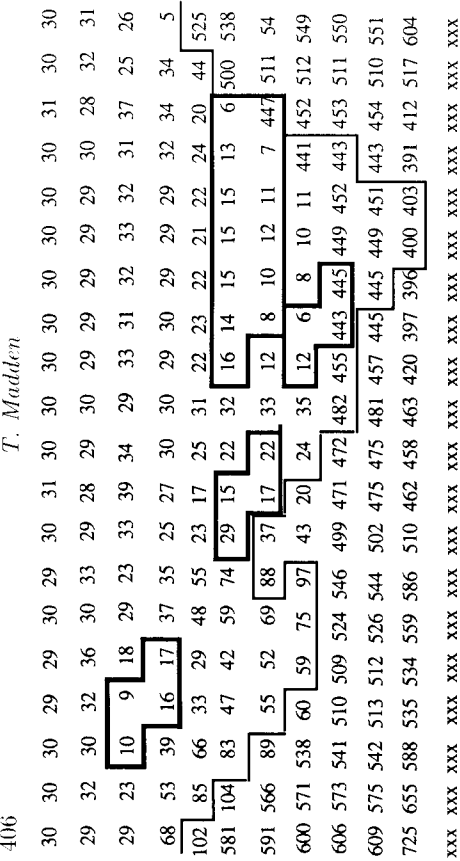


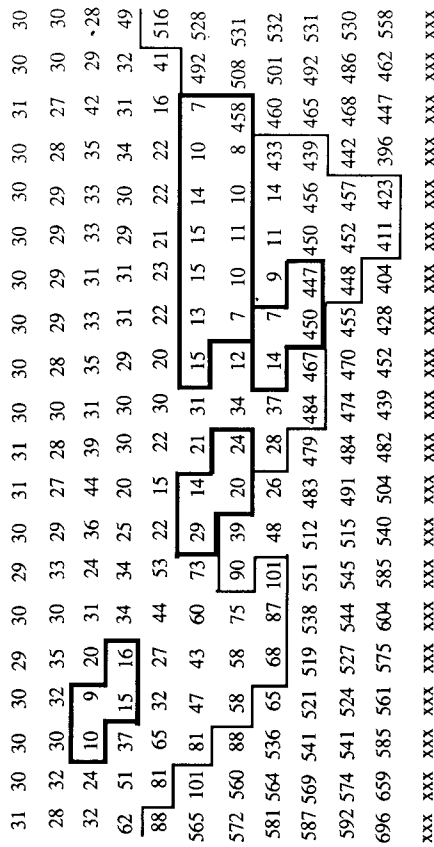
Fig. 9. Convergence of the inversion iterations for Model Ov.2.

as pseudosections. Usually near surface conductivity variations distort the apparent resistivity sections, but such variations which are easily handled by the inversion were left out of the model. The phase on the other hand is very robust with respect to near surface effects. The data clearly shows the major structure which is sediments over basement over mantle and also clearly shows the increased valley conductance (conductivity-thickness product) near the middle of the model. The effect of the uppermost section of the conductive layer is clearly seen at the 1 second period and the basement disruption and the conductive layers near the basement affect the 10 second period data and influence all the longer periods as well. The edge effects of the basin are seen at all periods. All the data for periods greater than 10 seconds are in the mantle branch. The lower crustal branch is in the 1-10 second period range, but it is poorly defined over the deepest part of the basin.

The iteration convergence is shown in figure 9. Again the five preconditioned conjugate gradient relaxation steps were quite comparable to the direct matrix inversions, although both methods showed some overshoot problems. This is a common effect when doing linearized inversions of non-linear problems. It can be moderated by damping. The R_{min}^{-1} term in the left-hand side of (6.2), since it is positive definite, acts like a damping term, but one usually finds additional damping necessary at the early stages of the inversion when large changes in the parameter values are made. We usually add an extra positive diagonal term whose magnitude is tied to the rms fit to the right-hand side of (6.2). This rms error term, unlike the rms fit to the data, goes to zero at the maximum likelihood solution point. It



Standard Inversion Results



Relaxation Inversion Results

Fig. 10. Model Ov 2 inversion results.

would appear in this example that one could have used a bit more damping.

The models obtained by the two inversion procedures were also quite comparable as seen in figure 10. All the segments showing resistivities greater than 390 were in the apriori model basement. The bold outlines show the true locations of the conductive sections and the narrow out-

lines show the true valley floor. The disruption of the conductive layer was clearly seen, but the conductivity of the layer was somewhat underestimated and the shapes of the segments were not always correct. The topography of the valley floor in the middle of the valley was barely hinted at. The basement-sediment contrast is very large and inversion has difficulties in moving the apriori model valley floor as we have seen before. In regions where the valley was much deeper than the apriori model we see how the inversion tries to keep the total conductance reasonable by increasing the conductivity of the bottom sediments rather than eroding away the apriori basement. The resistivities are seen to be higher in the vicinity of the basement overthrust, but this feature is very poorly resolved.

We think these results demonstrate that one can successfully relax the inverse problem and that this will help to make 3D inversions practical. The question of finding the best relaxation procedure was not really investigated, but it is quite evident that one needs an effective procedure.

Acknowledgements

I wish to thank Keiiti Aki from whom and with whom I learnt much of what I know about inverse theory. I also benefited from exposure to two experts in this field, Shimon Cohen and Albert Tarantola. There are many fellow and former students from whom it is difficult to separate out who learnt what from whom, and although we will not name them, they will understand. Support for our work on modelling and inverting magnetotelluric data was provided by Amoco, Arco, Chevron, Conoco, Sohio, Sun and Texaco, and we are grateful for both the financial and intellectual help they provided.

Special recognition must be paid to the influence Lanczos book on linear operators has had in developing our concepts about linear operators and inversion and much of the material in these notes is directly traceable to this reference.

References

- Aki, K., and Richards, P. G., 1980. Quantitative Seismology: Theory and Methods, W.H. Freeman and Co.
 Cagniard, E., 1963. Basic Theory of Magnetotelluric Method of Geophysical Prospecting, Geophysics, 18, 605-634
 Claerbout, J.F., 1976. Fundamentals of Geophysical Data Processing, McGraw-

- Hill.
- Franklin, J.N., 1970. Well posed stochastic extensions of ill posed linear problems, *J. Math. Anal. Applic.*, 31, 682-716
- Golub, G.H., and Van Loan, C.F., 1983. *Matrix Computations*, The John Hopkins University Press
- Lanczos, C., 1960. *Linear Differential Operators*, D. Van Nostrand
- Morse, P.M., and Feshbach, H., 1953. *Methods of Theoretical Physics*, McGraw-Hill
- Penfield, P. Jr., Spence, R., and Duinker, S., 1970. Telegen's Theorem and Electrical Networks, M.I.T. Press
- Reid, J. K., 1971. On the method of conjugate gradients for the solution of large sparse systems of linear equations, chapt in, *Large Sparse Sets of Linear Equations*, J.K. Reid edit, Academic Press
- Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, 49, 1259-1266
- Tarantola, A., and Valette, 1982. Generalized nonlinear inverse problems solved using the least-squares criterion, *Rev. Geophys. Space Phys.*, 20, 219-232
- Tikhonov, A. N., 1950. Determination of the electrical characteristics of the deep strata of the Earths Crust, *Doc. Acad. Nauk.(USSR)*, 73, 295-297
- Yagle, A.E., and Levy, B., 1985. A layer stripping solution of the inverse problem for a one-dimensional elastic medium, *Geophysics*, 50, 425-433