

# The 2-D magnetotelluric inverse problem solved with optimization

Ashley E. Van Beusekom,<sup>1</sup> Robert L. Parker,<sup>1</sup> Randolph E. Bank,<sup>2</sup> Philip E. Gill<sup>2</sup>  
and Steven Constable<sup>1</sup>

<sup>1</sup>IGPP, Scripps Institution of Oceanography, La Jolla, CA 92093-0225. E-mail: beusekom@usgs.gov

<sup>2</sup>Department of Mathematics, University of California San Diego, La Jolla, CA 92093-0112

Accepted 2010 November 15. Received 2010 October 25; in original form 2010 January 5

## SUMMARY

The practical 2-D magnetotelluric inverse problem seeks to determine the shallow-Earth conductivity structure using finite and uncertain data collected on the ground surface. We present an approach based on using PLTMG (Piecewise Linear Triangular MultiGrid), a special-purpose code for optimization with second-order partial differential equation (PDE) constraints. At each frequency, the electromagnetic field and conductivity are treated as unknowns in an optimization problem in which the data misfit is minimized subject to constraints that include Maxwell's equations and the boundary conditions. Within this framework it is straightforward to accommodate upper and lower bounds or other conditions on the conductivity. In addition, as the underlying inverse problem is ill-posed, constraints may be used to apply various kinds of regularization. We discuss some of the advantages and difficulties associated with using PDE-constrained optimization as the basis for solving large-scale nonlinear geophysical inverse problems.

Combined transverse electric and transverse magnetic complex admittances from the CO-PROD2 data are inverted. First, we invert penalizing size and roughness giving solutions that are similar to those found previously. In a second example, conventional regularization is replaced by a technique that imposes upper and lower bounds on the model. In both examples the data misfit is better than that obtained previously, without any increase in model complexity.

**Key words:** Numerical solutions; Inverse theory; Non-linear differential equations; Magnetotelluric; Geomagnetic induction.

## 1 INTRODUCTION

In magnetotelluric (MT) sounding, measurements of time-varying surface magnetic and electric fields allow us to learn about the conductivity structure of the Earth. However, the measurements are necessarily finite and uncertain and in the face of these limitations, the ultimate goal of an inversion must be to quantify the information that the measurements contain about the electrical structure of the Earth. Because the MT inverse problem is also non-linear and ill-posed, there is a paucity of rigorous methods that can extract the truly essential features of conductivity models. One such solution to the MT problem in one spatial dimension has been developed by us using optimization theory to unify the treatment of the differential equations and the inversion (Medin *et al.* 2007). Our method was applied to long-period observations and the question of the conductivity in the upper and mid-mantle paper. Our purpose here is to extend (as far as possible) that technique to a 2-D geometry.

Our earlier work does not rely on regularization to construct a plausible model from which to draw conclusions. Instead, the observations and the differential equations are regarded as constraints

in an optimization problem in which a conductivity function is minimized subject to inequality constraints that keep  $\sigma$  positive (for example). Subject to these constraints, bounds are sought on the average conductivity in intervals of particular geophysical interest, such as the seismic transition zone. Whereas the 1-D problem is motivated by whole-Earth issues such as possible conductivity associated with water in the transition zone, 2-D problems are concerned more with visualization of the crust and uppermost mantle. With this perspective, although definitive conclusions are the ultimate objective, the best that can be achieved with current techniques is the identification of general features rather than precise inference. Our work on the 1-D problem used the optimization package SNOPT (Gill *et al.* 2005). Here we employ the general-purpose partial differential equation (PDE)-solver PLTMG (Piecewise Linear Triangular MultiGrid), which combines the finite element method with an interior method for constrained optimization. As we will explain, fitting our problem to this code is far from a trivial matter, but there are advantages in using a flexible, debugged code that is built on firm theoretical foundations. The utility of incorporating inequality constraints, something easily done with PLTMG, is one lesson we take from the 1-D study.

In Section 5, we give results using conventional smoothing regularization and using inequality constraints. The data set, COPROD, was collected in southern Saskatchewan and Manitoba, Canada over the prominent North American Central Plain (NACP) conductive anomaly ( $<10 \Omega\text{m}$ ). COPROD has been the subject of numerous inversion studies, making it ideal for validating our method (see Jones 1993, for the solutions of other researchers). Because an optimization problem with regularization most closely resembles the previous researchers' problem setups, we first present our results on COPROD2 with regularization. Our code is able to find a better overall fit to the data compared with earlier solutions without increased model complexity; it can handle large numbers of sites without significant degradation in computational speed. For a second example of the value of our new method, the data set COPROD2 is inverted with additional inequality constraints on the structure in lieu of traditional regularization. The success of this constraint-based regularization indicates the benefits of PDE-constrained optimization for solving inverse problems.

## 2 OPTIMIZATION THEORY IN INVERSION

The theory for the 2-D MT inverse problem is far less advanced than that for the corresponding 1-D case. For example, conditions for the existence of a unique solution with exact observations are not known, whereas conditions for the solution of the 1-D case have been known for more than 30 years (Weidelt 1972). The existence problem in 1-D, that is, the consistency of a given finite, noisy data set with the 1-D approximation, was also settled some time ago (Parker 1980, 1982), through the ability to construct the best-fitting conductivity profile with depth, the  $D^+$  model comprised of a finite number of delta functions. The solution of the  $D^+$  problem cannot be extended to two dimensions because the residues at the poles in the complex  $c$  plane are no longer necessarily positive (Weidelt & Kaikkonen 1994). Only for very simple and specific inverse problems can one find the exact, or analytical, 2-D solution that gives the minimum data misfit, (see Weidelt 1975; Everett 1996).

In contrast, there are many purely numerical methods for finding a 2-D MT regularized solution (e.g. deGroot-Hedlin & Constable 1990; Rodi & Mackie 2001; Oldenburg & Ellis 1993; Smith & Booker 1991; Romo *et al.* 2005). These methods can usually find a smooth solution that fits the data, but they cannot identify properties common to all models that fit the data. Towards this goal, we present a 2-D inversion method based entirely on an optimization approach that treats the conductivity and the electric and magnetic fields at all observed frequencies as unknowns in addition to the conductivity. This 'all-at-once' approach was first introduced in the geophysical literature by the UBC-GIF group to solve 1- and 3-D regularized inversions of electrical problems as well as 2-D regularized inversions of gravity problems with additional inequality constraints on density (Haber *et al.* 2000, 2004; Oldenburg 2004; Lelièvre *et al.* 2008). We adopt the same basic philosophy, but exploit the versatility of PLTMG to apply different types of regularization, including constraint-based regularization.

## 3 SOLVER SPECIFICS

We briefly outline details of the general-purpose package PLTMG. More technical information may be found in (Bank 2007; Bank *et al.* 2003; Bank & Holst 2003; Bank & Smith 2002; Gill *et al.* 1981). PLTMG is designed to find the real scalar variable  $u$  that

satisfies the elliptic PDE:

$$-\nabla a(x, z, u, \nabla u) + f(x, z, u, \nabla u) = 0 \quad (1)$$

in the domain  $\Omega \subset \mathcal{R}^2$ , subject to boundary conditions

$$u = B_d(x, z) \text{ on } \Gamma_d \quad (2)$$

$$an = g(x, z, u) \text{ on } \Gamma_n \quad (3)$$

$$u, a.n \text{ continuous on } \Gamma_0 \quad (4)$$

for the Dirichlet, Neumann and periodic boundary conditions; here  $n$  is the unit normal,  $a$  is the vector  $(a_x \ a_z)^T$  and  $a_x, a_z, f, B_d, g$  are scalar functions. The equations and a schematic of the domain are depicted in Fig. 1. Then eqs (1)–(4) can be converted into the weak form used in the finite-element approximation, that is,  $A^*(u, \Lambda) = 0$  with the test functions  $\Lambda$ . Here  $A^*$  is given by

$$\begin{aligned} A^*(u, \Lambda) &= \int_{\Omega} [a(u, \nabla u) \nabla \Lambda + f(u, \nabla u) \Lambda] dx \ dz - \int_{\Gamma} g(u) \Lambda \ ds \\ &= \langle a(u, \nabla u), \nabla \Lambda \rangle_{\Omega} + \langle f(u, \nabla u), \Lambda \rangle_{\Omega} - \langle g(u), \Lambda \rangle_{\Gamma}, \end{aligned} \quad (5)$$

where  $\langle \cdot \rangle_{\Omega}$  and  $\langle \cdot \rangle_{\Gamma}$  are inner products over the domain and the boundary.

As the name indicates, PLTMG is a multigrid code that automatically refines the finite element triangular grid, beginning with a coarse approximation, then adding elements in regions where they are most needed in a sequence of iterations. However, PLTMG does not merely solve (1)–(4); it allows those equations, or equivalently  $A^*(u, \Lambda) = 0$ , to be applied as constraints in the minimization of an objective of our choosing:

$$p(u) = \int_{\Omega} p_1(x, z, u, \nabla u) \ dx \ dz + \int_{\Gamma} p_2(x, z, u, \nabla u) \ ds, \quad (6)$$

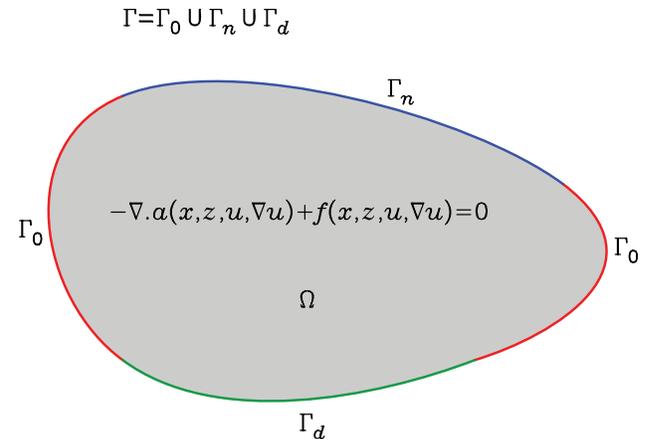
where  $p_1$  and  $p_2$  are scalar functions. Furthermore, the solution  $u$  can be subjected to upper and lower bounds:  $\underline{u}(x, z) \leq u \leq \bar{u}(x, z)$ . Thus we can state the basic problem solved by PLTMG as

$$\min_{u \in \mathcal{R}^N} p(u) \quad (7)$$

subject to

$$A^*(u, \Lambda) = 0. \quad (8)$$

$$\bar{u} \geq u \geq \underline{u}. \quad (9)$$



**Figure 1.** The domain of the general boundary value problem for the scalar elliptic partial differential equation, (1)–(4).

PLTMG solves this system by minimizing a single Lagrangian function in which the objective  $p$  and constraints all appear in suitably weighted terms.

This basic formulation does not cover the 2-D MT inverse problem because  $u$  must be a real scalar function. For example, in the Transverse Electric (TE) mode, the complex electric field components along strike (direction  $y$ ) satisfy

$$-\nabla\nabla E_k(x, z) + i\omega_k\mu_0\sigma(x, z)E_k(x, z) = 0 \quad (10)$$

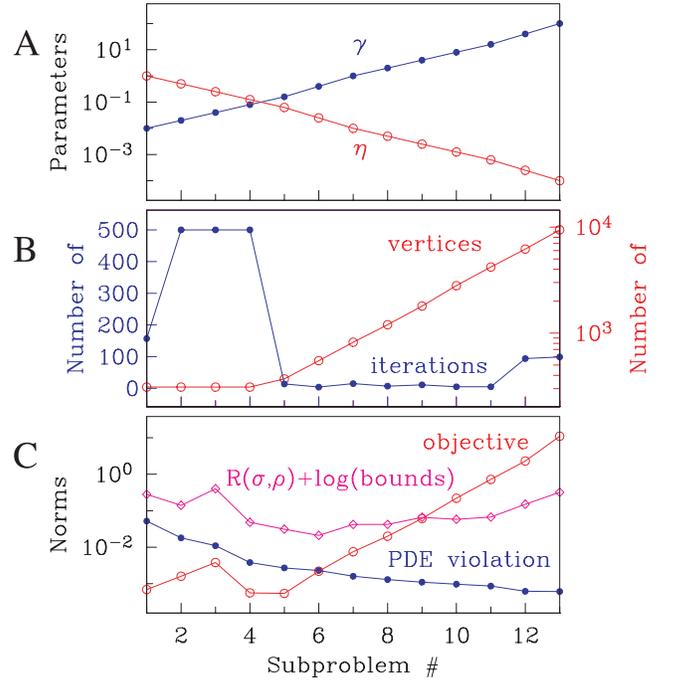
for each of the angular frequencies  $\omega_k$ ,  $k = 1, 2, \dots, K$ . Expanded in real and imaginary parts, these equations comprise a set of  $2K$  coupled elliptic equations. We can adopt the PLTMG machinery to our purposes by applying a framework called FOSLS, for First-Order System Least-Squares, developed for general second-order PDEs at the University of Colorado, Boulder and by Cai *et al.* (1994, 1997). We rewrite the coupled elliptic equations as a system of loosely coupled first-order equations by defining new variables for the partial derivatives. There are six real unknowns for each complex field in the original equations. For example, we define  $E_y = u + iv$  and introduce four real variables  $U = \partial_x u$ ,  $V = \partial_x v$ ,  $W = \partial_z u$  and  $Y = \partial_z v$ . These definitions allow the original equations to be expressed as a set of first-order equations; see eqs (A2)–(A7) of Appendix A. A modified version of PLTMG is used to minimize a functional consisting of the sum of the squared 2-norms of the residuals of the first-order equations. For convenience in exposition, we set  $u = (u \ U \ W \ v \ V \ Y \ \sigma)^T$ .

The elliptic vector problem obtained from the FOSLS discretization is solved as a PDE-constrained optimization problem using PLTMG. For this purpose PLTMG employs a primal interior method to minimize the penalty functional. A sequence of unconstrained subproblems is solved in which a new function is minimized, the Lagrangian of the form

$$\mathcal{L}_{\gamma\eta} = \{\text{objective}\} + \gamma \{\text{PDE violation}\} - \eta \log(\sigma). \quad (11)$$

A complete description of the objective function is given in the next section. The term that penalizes the violation of the field equations is described in Appendix A. In the successive subproblems, the parameters  $\gamma \rightarrow \infty$  and  $\eta \rightarrow 0$ ; the solution of the previous subproblem serves as a starting point for the next. As the iterations proceed, the minimizer of  $\mathcal{L}_{\gamma\eta}$  approaches the solution of the PDEs while maintaining strictly feasible values of  $\sigma$ . The positive parameter  $\eta$ , is called a barrier parameter (see Chap. 6, Gill *et al.* 1981); the logarithmic term enforces the positivity of  $\sigma$ , the conductivity. Additional equality constraints appear as objective terms penalized by  $\gamma$ . Additional inequality constraints, such as those in (9), would be entered as logarithmic terms multiplied by  $\eta$ .

The use of a sequence of unconstrained subproblems, introduced to solve the constrained optimization problem, also permits us to improve the resolution and accuracy of the finite element approximations in a natural way. At each step, as well as increasing  $\gamma$  and decreasing  $\eta$ , PLTMG refines the grid, adding more triangular elements as necessary. We make  $\gamma$  inversely proportional to  $\eta$ , while  $\gamma = O(\sqrt{N})$ , where  $N$  is the number of vertices in the grid. In this way errors in satisfying the constraints are of the same order as those in the finite element approximation to the PDE (Bank 2007). This approach is computationally efficient since we are not solving the subproblems with large  $\eta$  on a fine mesh; the large  $\eta$  problems normally require more iterations. As the mesh is refined, the current solution is interpolated onto the refined mesh and becomes the initial guess for the iteration on that mesh. Because of the improving quality of these initial guesses, we expect that fewer iterations will be required for the finest meshes. The values of the pa-



**Figure 2.** Convergence history of the COPROD2 joint TE—TM data inversion. A and C show parameters  $\gamma$  and  $\eta$  and the L2 norms of eq. (11). The objective here only contains the data misfit. The regularization term  $R(\sigma, \rho)$  will be discussed later. B shows the number of vertices in the mesh and the number of iterations required for each subproblem’s convergence.

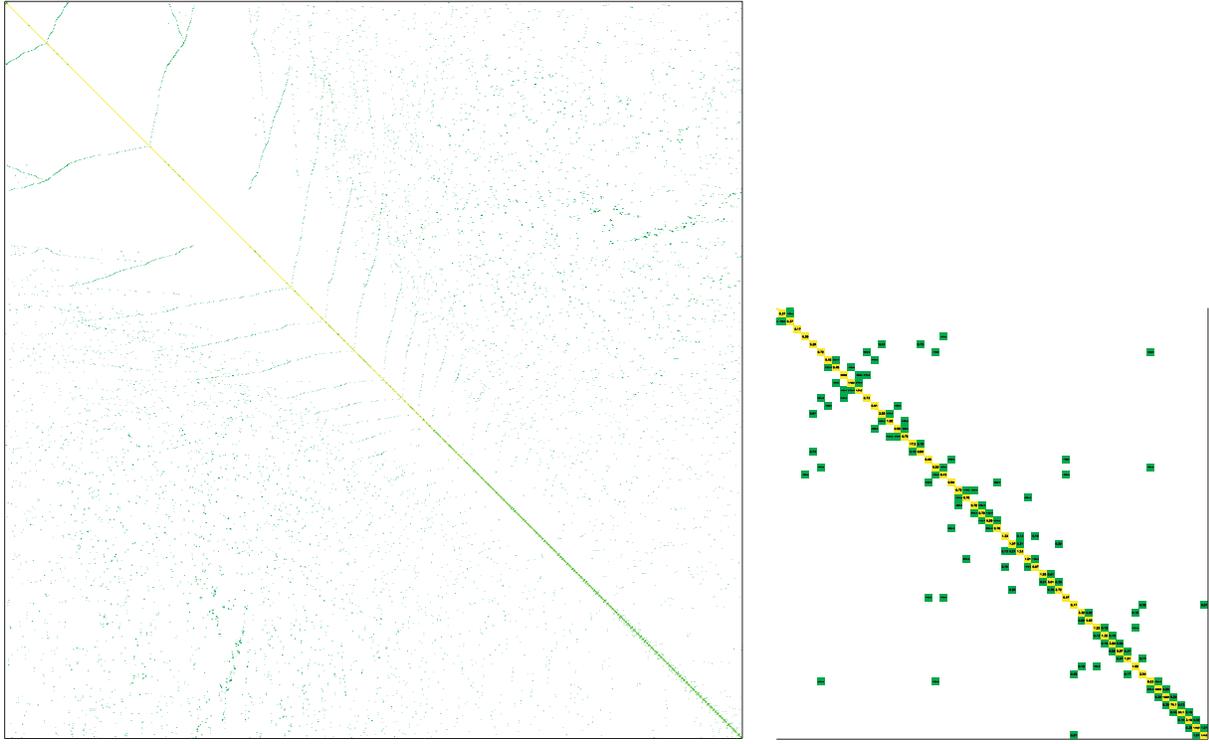
rameters and norms throughout convergence of a specific problem are shown in Fig. 2. (The problem is the COPROD2 joint TE—TM data inversion; the equations are (13)–(26) and the solution for resistivity is Fig. 4).

In each subproblem we seek the stationary point of the Lagrangian, the point where  $\nabla\mathcal{L}_{\gamma\eta} = 0$ , which is a necessary condition for optimality. To solve this system of nonlinear equations PLTMG uses an approximate Newton iterative procedure with line search. We write a perturbation to the current solution vector as  $t$ , then a first-order Taylor expansion yields

$$t\nabla\nabla\mathcal{L}_{\gamma\eta}(u) = -\nabla\mathcal{L}_{\gamma\eta}(u). \quad (12)$$

This sparse linear system is solved approximately for  $t$ . Here  $\nabla\mathcal{L}_{\gamma\eta}$  is the Jacobian of the Lagrangian and  $\nabla\nabla\mathcal{L}_{\gamma\eta}$  is its Hessian. Note, however, that  $\nabla\nabla\mathcal{L}_{\gamma\eta}$  is the Jacobian of the system  $\nabla\mathcal{L}_{\gamma\eta} = 0$  and we refer to this matrix as the Jacobian for the problem in its role in the Newton iterations. In our approach, the Jacobian is approximated by a block diagonal matrix, with diagonal blocks corresponding to each scalar field and conductivity, or  $u = (u \ U \ W \ v \ V \ Y \ \sigma)^T$  for each frequency. These diagonal blocks typically have a form equivalent to that arising from discretization of a second-order scalar elliptic PDE of the type normally handled by PLTMG. This diagonal approximation significantly reduces the time needed to solve the linear system for  $t$ . The next estimate of the solution is  $u_+ = u + \beta t$ , where  $\beta$  is a positive step chosen to yield a sufficient decrease in the Lagrangian. A single step of the iterative Newton cycle for each subproblem involves recalculating the Lagrangian and its derivatives at  $u_+$  and the calculation of an approximate solution of the linear system for  $t$ .

The linear system (12) is very large for even a modest-size problem. However, the Jacobian is sparse and we only compute its



**Figure 3.** Green and yellow are the non-zero values in the sparse Jacobian matrix for a small TE inverse problem with 2500 vertices. The right figure is a zoom on 0.01 per cent of the matrix, the lower right corner.

diagonal blocks, which makes storage of  $\nabla\nabla\mathcal{L}_{\gamma\eta}$  and solution of the equations possible. Fig. 3 shows a diagonal block of the Jacobian  $\nabla\nabla\mathcal{L}_{\gamma\eta}$  for a small TE inverse problem with coarse mesh of 2500 vertices (the matrix is accordingly a square  $2500 \times 2500$ ). Furthermore, only one of the sparse diagonal Jacobian blocks need be computed at a time. By storing the non-zero elements of only one sparse diagonal block at any given time, the same storage may be shared by all the diagonal blocks and the space required is that same as that for a single self-adjoint scalar PDE equation. This is a very big storage savings when there are as many scalar unknowns, as there are in our problem.

#### 4 OPTIMIZATION OF THE INVERSE PROBLEM

We now discuss how to set up a finite-dimensional problem that approximates all the unknowns simultaneously, that is, it gives the electric field  $E$ , the magnetic field  $H$ , the conductivity  $\sigma$  and the resistivity  $\rho = 1/\sigma$ . In principle, there are many ways to formulate an optimization problem in the PLTMG setting that will solve the MT inverse problem. Each of these alternatives is non-linear and non-convex and so the iterative scheme may converge to only a local minimizer of the optimization problem. Moreover, because inverse problems are inherently ill-conditioned, relatively minor differences in the formulation of the problem can make the difference between rapid convergence and failure. For example, consider the the single variable equation  $1/x = 1$ . Newton's method for a root of  $f(x) = 1/x - 1 = 0$  diverges for starting points  $x \geq 2$  or  $x \leq 0$ . However, if the problem is reformulated as  $G(x) = x - 1 = 0$ , it may be solved in one Newton step from any starting point. The proper form

in a problem as complex as ours is far from obvious and we have experimented with many variants.

In general the measurements are of complex admittances,  $c(x_j, \omega_k)$ , in the both TE and TM modes. The data are collected at  $z = 0$  on the Earth's surface, at  $j = 1, \dots, J$  points  $x_j$  and  $k = 1, \dots, K$  frequencies  $\omega_k$ . An idealized optimization problem to solve the joint TE–TM inverse problem is given by the objective function

$$\min_{\substack{\sigma > 0 \\ \rho > 0}} \sum_{k,j} \left\{ \left| \frac{c_{\text{TE}}(x_j, \omega_k) - E(x_j, 0, \omega_k) / (-\partial_z E(x_j, 0, \omega_k))}{\Theta_{\text{TE}}(x_j, \omega_k)} \right|^2 + \left| \frac{c_{\text{TM}}(x_j, \omega_k) - \rho(x_j, 0) \partial_z H(x_j, 0, \omega_k)}{\Theta_{\text{TM}}(x_j, \omega_k)} \right|^2 \right\} + R(\sigma, \rho), \quad (13)$$

where the double summations are performed with respect to the indices  $j = 1, \dots, J$  and  $k = 1, \dots, K$ . The two sums provide the discrepancies between the observations and the model predictions. Each sum is weighted by the inverse of its error,  $\Theta_{\text{TE}}$  and  $\Theta_{\text{TM}}$ . The functional  $R$  is a regularizing term defined later. The objective function is minimized subject to three sets of constraints. The first is the differential equations for the electromagnetic fields:

$$-\nabla\nabla E_k + i\omega_k \sigma \mu_0 E_k = 0 \quad (14)$$

$$-\nabla\rho\nabla H_k + i\omega_k \mu_0 H_k = 0, \quad (15)$$

where  $E_k = E(x, z, \omega_k)$  and  $H_k = H(x, z, \omega_k)$ , the fields at the set of angular frequencies  $\omega_k$ . These constraints are imposed with top ( $z = -\tau$ ) and bottom ( $z = +\tau$ ) field and conductivity conditions for both modes:

$$E(x, \tau, \omega_k) = 0 \quad (16)$$

$$\sigma(x, \tau) = \sigma_0 \quad (17)$$

$$\partial_z H(x, \tau, \omega_k) = 0 \quad (18)$$

$$\rho(x, \tau) = \rho_0 \quad (19)$$

$$\partial_z E(x, -\tau, \omega_k) = -1 \quad (20)$$

$$H(x, -\tau, \omega_k) = \frac{1}{\mu_0 \omega_k} i. \quad (21)$$

In addition, we impose periodic left ( $x = -L$ ) and right ( $x = +L$ ) conditions for the fields and conductivities

$$E(L, z, \omega_k) = E(-L, z, \omega_k) \quad (22)$$

$$\sigma(L, z) = \sigma(-L, z) \quad (23)$$

$$H(L, z, \omega_k) = H(-L, z, \omega_k) \quad (24)$$

$$\rho(L, z) = \rho(-L, z). \quad (25)$$

In (13)–(25) we treat  $\sigma$  and  $\rho$  as independent unknowns and then we link them with a constraint equation:

$$\sigma\rho - 1 = 0. \quad (26)$$

Note that since the constraints are satisfied as  $\gamma$  of (11) approaches infinity,  $\sigma$  and  $\rho$  will only be reciprocals at the solution. Equation (26) is the first example of some additional restrictions that are needed to obtain a PLTMG-FOSLS solution. If either  $\sigma$  or  $\rho$  is eliminated, we find that the joint TE–TM inversion never converges to a sensible solution. When one of the unknowns appears in the denominator of one of the PDEs there is a solution in which that variable tends to zero everywhere, along with the other field variables and of course this is not the solution we want. However, it is the one to which the PLTMG-FOSLS iterations usually tend. Thus dividing by a variable that has an attraction to zero is not a good idea in the current PLTMG-FOSLS code. The root of the difficulties may lie in our not retaining the off-diagonal blocks of the Jacobian. Keeping only multiplications in the PDEs and connecting  $\rho$  and  $\sigma$  through (26) appears to overcome the problem. The cost of introducing an additional unknown and one more equation into the system is very slight; recall each additional frequency requires six new variables.

In the computer code, all variables are scaled so that (13)–(26) are the same magnitude. For example, lengths are divided by  $L$  and frequencies by an average  $\omega$ .

Next, we consider the modifications that are necessary to adapt the objective equation (13) to our approach. The objective is the weighted squared 2-norm misfit between the complex TE and TM data,  $c_{TE}$  and  $c_{TM}$  and the model values at the data collection points with the addition of regularization functionals  $R$ . We have weighted the deviations by the inverse of the error  $\Theta(x_j, \omega_k)$  and thus the sum is in  $\chi^2$  form. The TM part of the objective is easily handled as written above, but the TE part is subject to numerical instability resulting from the division by the  $z$  gradient of  $E$ . To mitigate this instability, the TE misfit terms are multiplied by  $\partial_z E$ , which rescales the objective. At first sight this appears to give TE penalty with the wrong units, but the boundary condition (20) makes the gradients of  $E$  dimensionless. The revised objective is effectively the same when the misfit is small, since  $|\partial_z E|$  at the surface does not

deviate very far from unity in the final solution. Consider a single term in the TE part of the objective. For simplicity assume that the uncertainties in the real and imaginary are the same (something easily revised if necessary), so that  $\Theta_{TE}(x_j, \omega_k) = (1 + i)\Theta/\sqrt{2}$ ; write  $c_{TE}(x_j, \omega_k) = a + ib$ ,  $E = u + iv$  and  $\partial_z E(x_j, 0, \omega_k) = W + iY$ . Then a typical term in the original TE objective sum is

$$\frac{1}{\Theta^2} \left| a + ib - \left( -\frac{u + iv}{W + iY} \right) \right|^2. \quad (27)$$

We multiply though by  $(W + iY)^2$  and then break into real and imaginary parts; this yields an alternative TE misfit terms:

$$\Theta^{-2}(-aW + bY - u)^2 + \Theta^{-2}(-aY + bW - v)^2. \quad (28)$$

Even with this reformulation, during the early iterations (i.e. when  $\gamma$  is small in (11)) there is still a tendency to converge to a solution in which  $\partial_z E(x_j, 0, \omega_k)$  is very small. This instability may be avoided by using a regularization term weighted by  $\eta$  (the reciprocal of  $\gamma$ ). This scheme is shown qualitatively in (11) which biases early solutions towards those with  $W + iY = -1$  in the air ( $z < 0$ ).

We follow the traditional formulation of including an air layer above the conducting ground. This approach is often used to allow for the fact that the correct boundary condition for the TE mode at the Earth's surface (Schmucker 1971) cannot be written in differential form. For this reason the exact condition is incompatible with the requirements of PLTMG for boundary conditions. The fields vary slowly in the air and the number of elements devoted to this region is small. We can simplify things in the air layer, where the conductivity is known. Rather than solving the true PDEs (14) and (15) there, we impose the conditions  $\nabla \nabla \cdot E = 0$  and  $H - (i/\omega_k) = 0$ , since physically in a perfect insulator  $H$  does not vary. Because PLTMG assumes every variable has a solution over the entire domain, we artificially set  $\sigma = \rho = 1$  in the air to avoid the scaling problems associated with setting  $\sigma = 0$ ,  $\rho = \infty$ . We do not need to include  $\sigma$  and  $\rho$  in the equations when we solve for  $E$  in the air layer.

The presence of Dirichlet boundary conditions on  $\rho$  and  $\sigma$ , (17) and (19), requires some comment. If we do not set a Dirichlet boundary on  $\rho$ , we find that no reasonable amount of regularization on  $\|\partial\rho/\partial z\|_2^2$  keeps  $\rho$  from going to zero at the bottom boundary  $z = \tau$ . Because of the way in which FOSLS handles the Neumann boundary condition (18) (see Appendix A), without an explicit condition on  $\rho$ , we find that  $\rho(x, \tau) = 0$  is the solution to the problem we have posed; but then  $\sigma = 1/\rho = \infty$  causes instability. The presence of the Dirichlet condition in  $\rho$  necessitates one in  $\sigma$  also. As long as the bottom boundary is far from the surface no pathological behaviour from these boundary conditions is observed.

Finally, we consider the regularization term in the objective. The MT inverse problem is ill-posed because small-scale variations in conductivity at depth have vanishingly small influence on the observations and so, unless they are suppressed somehow, the solution to the inverse problem is unstable. The function  $R$  in (13) is designed to penalize large amplitudes and short wavelength components in  $\sigma$  and  $\rho$ ; see (29). This discrimination can be tailored to increase with depth, since growth of the unstable parts increases exponentially far from the surface. Another way to look at regularization (Parker 1994) is to say that with it we obtain the smoothest possible solution (the one with the smallest  $R$ ) for the particular misfit obtained and thus all other models with the same misfit are necessarily more complex. Keeping complexity of the models to a minimum is desirable for interpretation in terms of geological structure. One can do without regularization, however, if one applies strong inequality

constraints, for then the optimization problem becomes one of trading misfit against a bound on conductivity and there is every reason to believe that extremal solution pressing up against conductivity bounds are rather smooth overall. We illustrate these ideas in the next section.

## 5 RESULTS

We apply the optimization method outlined above to analyse a data set. During early testing we applied the code to artificial data generated from a known structure: a hemicylindrical conductor lying on a perfectly conducting base. We used 3 frequencies and 7 sites of TE and TM data. The algorithm returned a smoothed version of the correct structure. For more details see Medin (2008). Here we invert a set of field observations, COPROD2 (Jones & Savage 1986; Jones 1988), which contains a subset of the frequencies drawn from COPROD. The complete set consists of 40 frequencies of TE and TM data collected at 35 sites along a 407 km profile in southern Saskatchewan and Manitoba, Canada. We will be comparing our results mainly with those of deGroot-Hedlin & Constable (1993), whose data selection we follow, namely, the TE and TM responses for 6 of the lower frequencies. The data set we used was already corrected for static shifts by the earlier investigators.

### 5.1 COPROD2 joint TE—TM data inversion

We run the joint inversion on 6 frequencies of the COPROD2 data set. With 6 frequencies of both the TE and TM modes in the FOSLS format on a final mesh of 9400 vertices (stopping at  $\nu = 100$ ), we are solving for  $(2 \times 6 \times 6 + 2) \times 9400 = 676\,800$  unknowns with  $2 \times 6 \times 8 + 1 = 97$  constraint equations. The convergence history is shown in Fig. 2. This problem took approximately 16.9 hr to run on a 2 processor 3 GHz Intel Xeon desktop with 32 GB RAM. The solution is shown in Fig. 4. We will report the frequencies as their reciprocals, periods in seconds and the conductivity/resistivity values as resistivity in  $\Omega\text{m}$ ; this is the traditional way to discuss MT systems.

In the optimization problem eqs (13)–(26) we use  $\Theta_{\text{TE}}(x_j, \omega_k) = 0.1$ ,  $\Theta_{\text{TM}}(x_j, \omega_k) = 0.1|c_{\text{TM}}(x_j, \omega_k)|$  and

$$R(\sigma, \rho) = f(\sigma, \sigma_0) + f(\rho, \rho_0) \quad (29)$$

with

$$f(\beta, \beta_0) = w_1 \eta \int_{\Omega} \exp(10z + x^2/4) [(\beta(x, z) - \beta_0)^2 + |\nabla\beta|^2] dx dz, \quad (30)$$

where  $w_1 = 10^{-6}$ ,  $\rho_0 = 10 \Omega\text{m}$  (scaled  $\sigma_0 = 10$  and scaled  $\rho_0 = 0.1$ ) and  $\eta$  is a parameter that is driven to zero as the solutions of the successive subproblems converge, as shown qualitatively in (11). This regularization is biased towards regions that are far from the data (i.e. where the problem is the most ill-posed and hence subject to difficulties with singularities). Due to our weighting with parameter  $\eta \rightarrow 0$  as we converge on a solution, the regularization is only large at the beginning. It was verified experimentally that the choice of  $\sigma_0$  and  $\rho_0$  does not change the major features of the solution.

The data fit is shown together with the resistivity. Constraint (26) is accurately satisfied so that  $\rho$  is indistinguishable from  $1/\sigma$ . The RMS misfit is 0.9 per cent. Of note, there is a lot of short wavelength, high-frequency variability in the model responses of the TM COPROD2 data which are not reflective of anything in the data. This behaviour is caused by shallow structure in the resistivity model

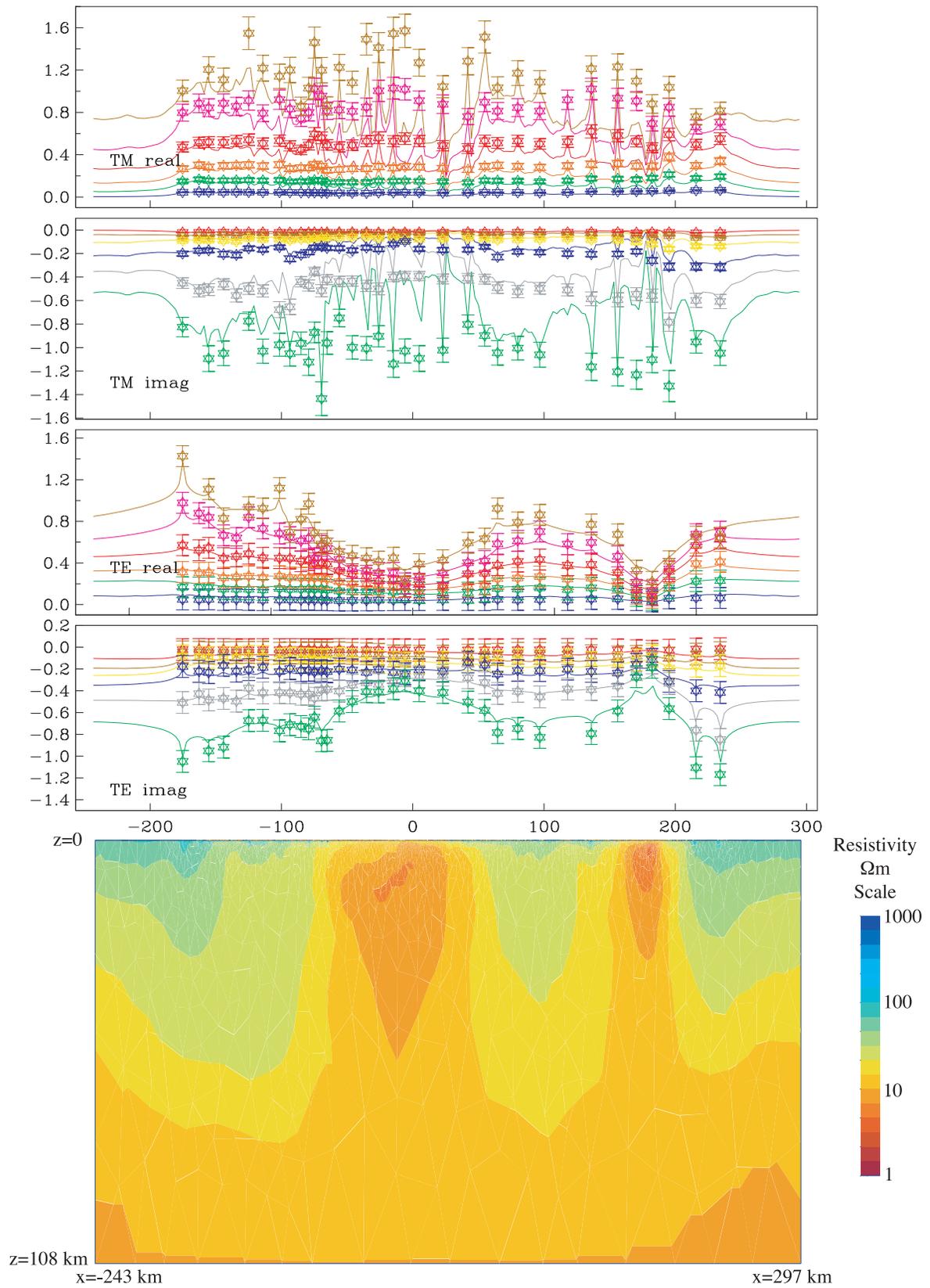
which is not visible on the scale of the plot. The TM COPROD2 data does not tell us much about the deep structure; we find that in the joint inversion the addition of the TM data to the TE data smoothes the solution found from the TE data inversion alone. The TM data inversion alone does not resolve the known conductive anomalies. Previous researchers (e.g. Berdichevsky *et al.* 1998) have found that TM data tend to resolve resistive structure ( $> 1000 \Omega\text{m}$ ) while TE data tend to resolve conductive structure ( $< 10 \Omega\text{m}$ ), so we are not surprised that in a region with conductive anomalies the TE inversion identifies the structure.

We compare our solution to the joint inversions by deGroot-Hedlin & Constable (1993) and Romo *et al.* (2005). These solutions are respectively A, B and C of Fig. 5. The scale of the three models is the same; the vertical dashed lines are the horizontal extent of our plot on the same scale. We place the figures between the lines in respect to where they lie in the range  $-243 \leq x \leq 297 \text{ km}$ . A direct comparison of computational speed is not possible because the different computers involved in each of the calculations.

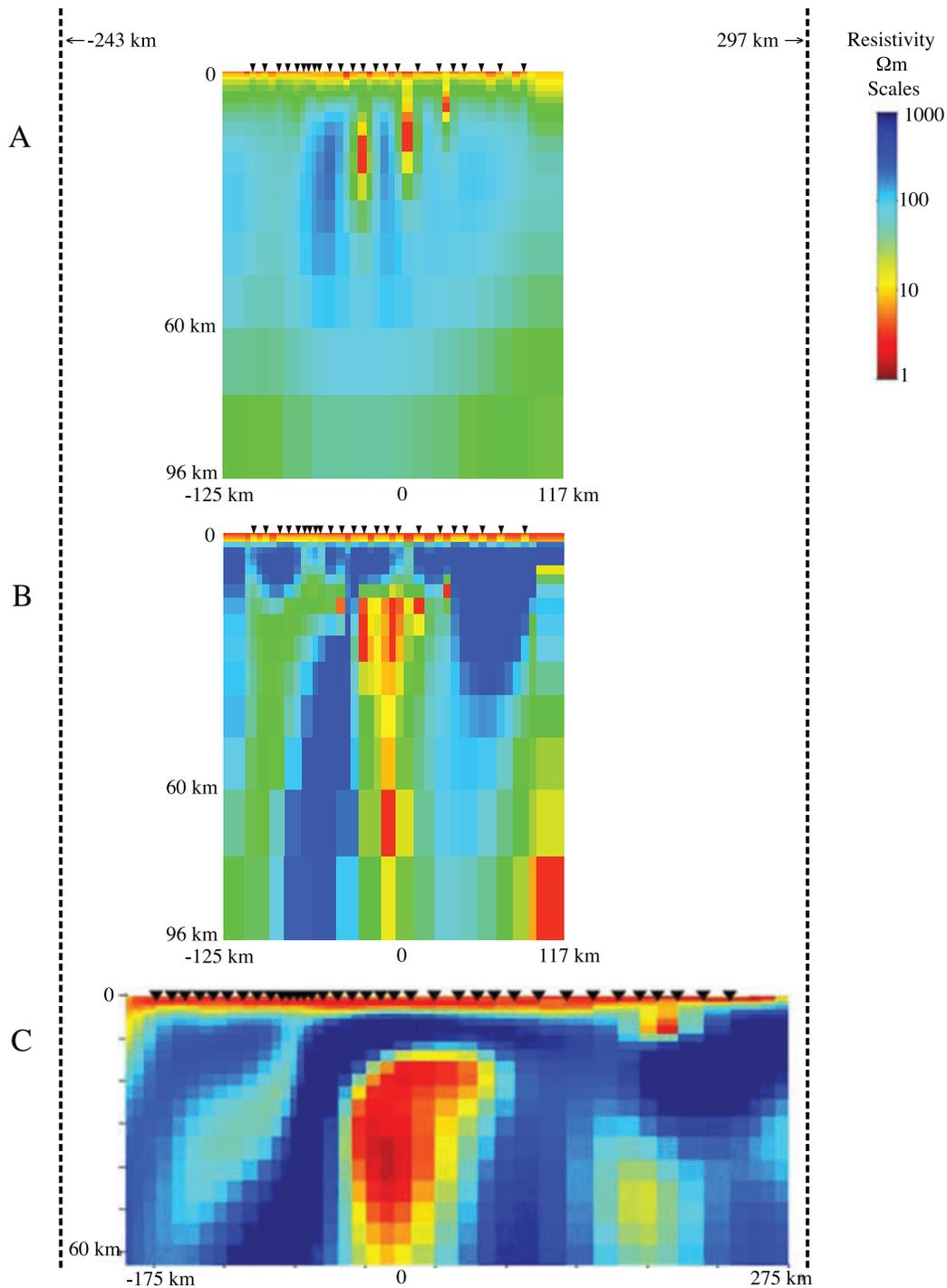
The inversions by deGroot-Hedlin & Constable (1993) used the same periods plus 28 s, for the middle 23 data sites in the popular 2-D MT inversion code Occam2-D (deGroot-Hedlin & Constable 1990). Our code has an advantage over Occam2-D in that it only refines heavily around data sites instead of also at large-depth. So there is little computational expense added when more sites are included, consequently we use all 35 data sites. Occam2-D finds a smooth model that fits the data to some desired misfit. Plot A shows Occam2-D's solution dominated by a discontinuous conductive body from  $-50 < x < 50 \text{ km}$ , centred at depth between  $8 < z < 22 \text{ km}$ . This body is identified as the NACP anomaly. The data are fit to 10 per cent RMS. Plot B is the result from Occam2-D with biasing of the upper 2.2 km to a particular 1-D structure. Again, the NACP anomaly is most likely the conductive body between  $-50 < x < 50 \text{ km}$ ; this time the body is centred deeper at  $z \approx 40 \text{ km}$ . The data are fit to 15 per cent RMS. Occam2-D inverts models in a brick mesh so the structure in these models contain discontinuous jumps while our solutions, based on adaptive triangle meshing, do not. It should be noted that our structure fits the data much more accurately, yet is smoother at depth.

The model from Romo *et al.* (2005) is from an inversion of all 40 frequencies of the COPROD data set, instead of from only the reduced COPROD2 data set. Therefore we expect their model to possess more detail near the surface than ours; our model relies on the lower frequencies with larger skin depths. They use all 35 data sites as we do. They find a solution with the NACP anomaly and a smaller anomaly to the east, positive  $x$ -direction. This is identified as the TOBE (Thompson nickel belt) anomaly. They do not use TE and TM modes for inversion, instead using a 'series and parallel' rotation for pseudo TE and TM modes of data. See Romo *et al.* (2005) for more information on their inversion method. Consequently this model is derived from slightly different data as well as a completely different method of inversion. Nevertheless, we believe it is worthwhile to compare the models, albeit with caution. They fit the data to 5 per cent RMS.

Our solution recovers two conductive anomalies, NACP and TOBE, in similar regions as deGroot-Hedlin & Constable (1993) and Romo *et al.* (2005). The overall resistivity is less extreme in its maximum and minimum values and looks generally smoother than the models in Fig. 5, but this could be due to our choice of regularization. We note that if we regularize the solution by tightening the bounds on  $\sigma$  and  $\rho$ , instead of using the smoothing regularization function (29), we can also get sharper anomalies. In any case, we are encouraged by the fact that our code is both fitting the data and



**Figure 4.** The joint TE-TM COPROD2 data inversion resistivity solution,  $\Omega_m$ . The data fit is plotted above, with each period of data (real admittances, imaginary admittances) in the following colors: 14 s is (blue,red), 57 s is (green,brown), 114 s is (orange,yellow), 228 s is (red,blue), 455 s (purple,gray), 910 s is (brown,green). Where the line appears to not come close to intersect with the data points (notably in the longest period of TM), the fields are creating close to delta functions which do not plot well.



**Figure 5.** Joint inversions of COPROD2 data by other researchers for comparison with our result in Fig. 4. A and B are from deGroot-Hedlin & Constable (1993) and C is from Romo *et al.* (2005). The horizontal and vertical scale of the three models is the same as our model in Fig. 4. The models are placed between the dashed lines in respect to where they lie in  $-243 \leq x \leq 297$  km, the horizontal extent of our model.

reproducing the general features in the conductivity models seen by others.

## 5.2 COPROD2 inversion with bounds

Here we demonstrate regularization with bounds only, so that  $R(\sigma, \rho) = 0$  in the objective equation (13). To get good convergence, we

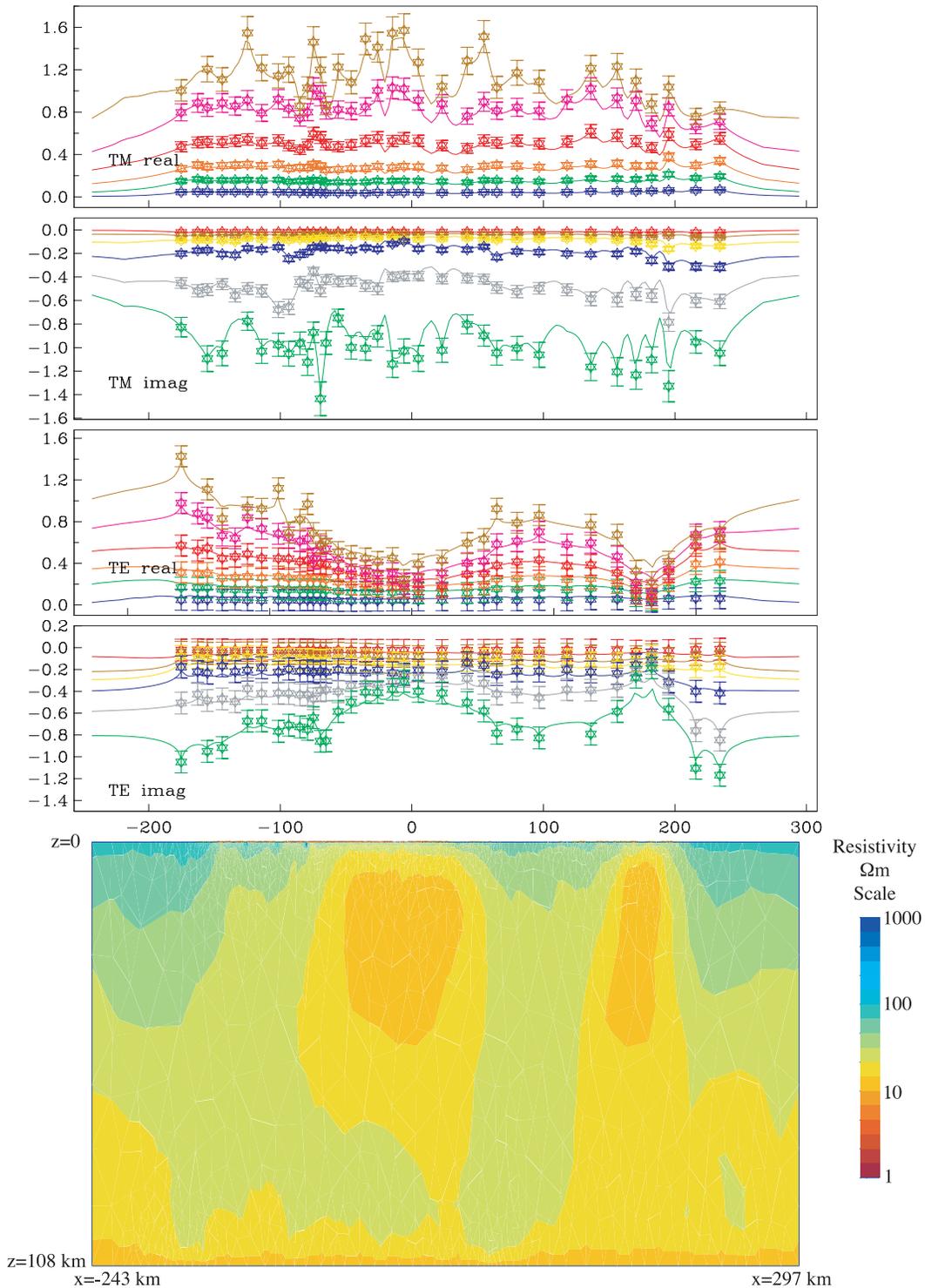
have to choose rather narrow bounds if we insist on absolutely no smoothing regularization terms. However, we want to make these narrow bounds as wide as possible so that we are not influencing the solution too much and/or driving up the data misfit. For the COPROD2 data, we desire the resistivity to be  $12.5 \Omega\text{m} \leq \rho \leq 1000 \Omega\text{m}$ . Recall (11), the qualitative description of the subproblem solved in the primal interior method, repeated here with the

two-sided bounds:

$$\mathcal{L}_{\gamma\eta} = \{\text{objective}\} + \gamma \{\text{violation of PDE and other constraints}\} - \eta \{\log(\sigma - 0.001) + \log(-\sigma + 10) + \log(\rho - 0.001) + \log(-\rho + 10)\}, \quad (31)$$

where in successive subproblems  $\gamma \rightarrow \infty$ ,  $\eta \rightarrow 0$ . In the early cycles of solving with  $\gamma$  small, (26), the constraint  $\sigma - 1/\rho = 0$ , is not strongly enforced. Hence the variables  $\sigma$  and  $\rho$  each need to be controlled, or regularized and we choose to do this with bounds.

The bounded solution along with the data fit is shown in Fig. 6. The RMS misfit is 1.1 per cent. The constraints are the same as the



**Figure 6.** The joint TE–TM COPROD2 data inversion regularized by bounds only; we constrain  $10 \leq \rho \leq 1000 \Omega\text{m}$ . The data fit is plotted above, with each period of data (real admittances, imaginary admittances) in the following colours: 14 s is (blue,red), 57 s is (green,brown), 114 s is (orange,yellow), 228 s is (red,blue), 455 s (purple,gray), 910 s is (brown,green). Where the line appears to not come close to intersect with the data points (notably in the longest period of TM), the fields are creating close to delta functions which do not plot well.

previous problem, except that the Dirichlet boundary condition on the structure at large-depth (17) and (19), is slightly more resistive with  $\rho_0 = 12.5 \Omega\text{m}$  so it is within the bounds. This problem took approximately 12.5 hr to run on a 2 processor 3 GHz Intel Xeon desktop with 32 GB RAM. The solution in Fig. 6 is more resistive than the solution in Fig. 4, as we demanded with the bound constraints. Because we only used bounds and not smoothing regularization, the structure shown in Fig. 6 does not fade slowly into the Dirichlet boundary condition on the structure at large-depth. However, we are encouraged in the validity of our previous assumptions, as again we see that the value of the Dirichlet boundary condition on the structure at large-depth does not appear to affect the near-surface features of the solution since we have made the boundary deep enough.

Being able to use bounds for regularization is a great advantage over other inverse codes (e.g. deGroot-Hedlin & Constable 1990; Rodi & Mackie 2001; Oldenburg & Ellis 1993; Smith & Booker 1991; Romo *et al.* 2005). We can place the same bound on the entire region, as we have done here. Or if we know *a priori* the location of structure, we can place bounds only on that location and smooth the rest of the domain.

## 6 CONCLUSIONS

We have solved the 2-D MT inverse problem with modern optimization technology. Our approach can find solutions similar to those found by other codes based on quite different principles. Inversion of the COPROD2 demonstrates the versatility of our regularization options: for the first problem we penalized structure size and roughness and with the second problem conductivity bounds stabilized the solution. We are able to satisfy the requirements of observation to a higher accuracy than earlier codes, without introducing extreme roughness.

Our program improves on previous codes in several ways. For example, inclusion of large numbers of measurement sites does not increase computational work significantly, whereas with other approaches the work increases proportionally to the number of sites. Moreover, based on our experiments with artificial data, we have reason to believe that the addition of sites will improve the convergence behaviour and reduce the data misfit in the inversion. Another advantage is the use of triangle mesh refinement: this process allocates the computational effort efficiently by only refining heavily in the regions where it is necessary. Arguably the most important improvement our code offers is the ability to include specific constraints such as bounds on conductivity, or to regularize the structure in the more traditional way. The optimization format allows these changes to be easily added to the problem setup.

## ACKNOWLEDGMENTS

We thank Alan Jones for making the processed COPROD2 data set publicly available and also P.J. Savage and PanCanadian by making the raw data available. The work of Ashley Van Beusekom was partially supported by the Scripps Institution of Oceanography Seafloor Electromagnetic Methods Consortium. The work of Randolph Bank and Philip Gill was supported by National Science Foundation Grant DMS-0511766. Suggestions by Thomas Kalscheuer and an anonymous reviewer led to improvements in the manuscript.

## REFERENCES

- Bank, R., 2007. PLTMG: A software package for solving elliptic partial differential equations, User's guide 10.0, Tech. rep., Dept. of Math., University of California, San Diego.
- Bank, R. & Holst, M., 2003. A new paradigm for parallel adaptive meshing algorithms, *SIAM Rev.*, **45**, 292–323.
- Bank, R. & Smith, R., 2002. An algebraic multilevel multigraph algorithm, *SIAM J. Sci. Comput.*, **23**(5), 1572–1592.
- Bank, R., Gill, P. & Marcia, R., 2003. Interior methods for a class of elliptic variational inequalities, in *Large-scale PDE-Constrained Optimization*, vol. 30 of Lecture Notes in Computational Science and Engineering, pp. 218–235, Springer-Verlag, Berlin.
- Berdichevsky, M.N., C., Dimitriev, V. & Pozdnjakova, E., 1998. On two-dimensional interpretation of magnetotelluric soundings, *Geophys. J. Int.*, **133**, 585–606.
- Cai, Z., Lazarov, R., Manteuffel, T. & McCormick, S., 1994. First-order system least squares for second-order partial differential equations: Part I, *SIAM J. Numer. Anal.*, **31**, 1785–1802.
- Cai, Z., Lazarov, R., Manteuffel, T. & McCormick, S., 1997. First-order system least squares for second-order partial differential equations: Part II, *SIAM J. Numer. Anal.*, **34**, 425–454.
- deGroot-Hedlin, C. & Constable, S., 1990. Occam's inversion to generate smooth, two-dimensional models from magnetotelluric data, *Geophysics*, **55**(12), 1613–1624.
- deGroot-Hedlin, C. & Constable, S., 1993. Occam's inversion and the North American Central Plains electrical anomaly, *J. Geomag. Geoelectr.*, **45**, 985–999.
- Everett, M., 1996. Homotopy, polynomial equations, gross boundary data, and small Helholtz systems, *J. Comp. Phys.*, **124**, 431–441.
- Gill, P., Murray, W. & Wright, M., 1981. *Practical Optimization*, Academic Press, New York.
- Gill, P., Murray, W. & Saunders, M., 2005. SNOPT: An SQP algorithm for large-scale constrained optimization, *SIAM Rev.*, **47**(1), 99–131.
- Haber, E., Ascher, U. & Oldenburg, D., 2000. On optimization techniques for solving nonlinear inverse problems, *Inverse Problems*, **16**(5), 1263–1280.
- Haber, E., Ascher, U. & Oldenburg, D., 2004. Inversion of 3D electromagnetic data in frequency and time domain using an inexact all-at-once approach, *Geophysics*, **69**(5), 1216–1228.
- Jones, A., 1988. Static shift of magnetotelluric data and its removal in a sedimentary basin environment, *Geophysics*, **53**, 967–978.
- Jones, A., 1993. The COPROD2 dataset—tectonic setting, recorded MT data, and comparison of models, *J. Geomag. Geoelectr.*, **45**(9), 933–955.
- Jones, A. & Savage, P., 1986. North American Central Plains conductivity anomaly goes east, *Geophys. Res. Lett.*, **13**, 685–688.
- Lelièvre, P., Oldenburg, D. & Williams, N., 2008. Constraining geophysical inversion with geologic information, *SEG*.
- Medin, A., 2008. The Magnetotelluric Inverse Problem, *PhD thesis*, University of California, San Diego.
- Medin, A., Parker, R. & Constable, S., 2007. Making sound inferences from geomagnetic sounding, *Phys. Earth planet. Int.*, **160**, 51–59.
- Oldenburg, D., 2004. Summary of UBC-GIF project: Inversion and modeling of applied geophysical electromagnetic data (IMAGE), *Tech. rep.*, Dept. of Earth Sci. University of British Columbia.
- Oldenburg, D. & Ellis, G., 1993. Efficient inversion of magnetotelluric data in two dimensions, *Phys. Earth planet. Inter.*, **81**, 177–200.
- Parker, R., 1980. The inverse problem of electromagnetic induction: existence and construction of solutions based upon incomplete data, *J. geophys. Res.*, **85**, 4421–4428.
- Parker, R., 1982. The existence of a region inaccessible to magnetotelluric sounding, *Geophys. J. R. astr. Soc.*, **68**, 165–170.
- Parker, R., 1994. *Geophysical Inverse Theory*, Princeton University Press, Princeton.
- Rodi, W. & Mackie, R., 2001. Nonlinear conjugate gradients algorithm for 2-D magnetotelluric inversion, *Geophysics*, **66**(1), 174–187.
- Romo, J., Grómez-Treviño, E. & Esparza, F., 2005. Series and parallel transformations of the magnetotelluric impedance tensor: theory and applications, *Phys. Earth planet. Inter.*, **150**, 63–83.

- Schmucker, U., 1971. Interpretation of induction anomalies above non-uniform surface layers, *Geophysics*, **36**(1), 156–165.
- Smith, J. & Booker, J., 1991. Rapid inversion of two- and three dimensional magnetotelluric data, *J. geophys. Res.*, **96**(B3), 3905–3922.
- Weidelt, P., 1972. The inverse problem in geomagnetic induction, *Z. Geophys.*, **38**, 257–289.
- Weidelt, P., 1975. Inversion of two-dimensional conductivity structures, *Phys. Earth planet. Inter.*, **10**, 282–291.
- Weidelt, P. & Kaikkonen, P., 1994. Local 1-D interpretation of magnetotelluric B-polarization impedances, *Geophys. J. Int.*, **117**, 733–748.

## APPENDIX A: A BRIEF DESCRIPTION OF FOSLS

We illustrate the FOSLS method with the setup for solving the TE mode forward equations at a single frequency. Inclusion of an objective, other frequencies and constraints, etc, simply adds more terms in the Lagrangian but does not change the essential procedure. For induction in the TE mode, the equation governing the along-strike ( $y$  component) of the complex electric field  $E(x, z)$  is (10) which we write

$$-\frac{i}{\omega\mu_0}\nabla\nabla E - \sigma E = 0. \quad (\text{A1})$$

For FOSLS format it is necessary to write  $E$  in terms of its real and imaginary parts

$$E = u + iv \Leftrightarrow \begin{pmatrix} u \\ v \end{pmatrix}$$

and define a matrix of first derivatives  $\mathcal{U}$

$$\nabla E = \mathcal{U} = \begin{pmatrix} \partial_x u & \partial_x v \\ \partial_z u & \partial_z v \end{pmatrix} = \begin{pmatrix} U & V \\ W & Y \end{pmatrix},$$

where  $\partial/\partial x$  and  $\partial/\partial z$  are denoted by  $\partial_x$  and  $\partial_z$ . The second-order differential eq. (A1) becomes a set of six first-order equations in the six real variables  $u, v, U, V, W, Y$ :

$$\partial_x u - U = 0 \quad (\text{A2})$$

$$\partial_z u - W = 0 \quad (\text{A3})$$

$$\partial_x v - V = 0 \quad (\text{A4})$$

$$\partial_z v - Y = 0 \quad (\text{A5})$$

$$\frac{1}{\omega\mu_0}(\partial_x V + \partial_z Y) - \sigma u = 0 \quad (\text{A6})$$

$$-\frac{1}{\omega\mu_0}(\partial_x U + \partial_z W) - \sigma v = 0. \quad (\text{A7})$$

Eqs (A2)–(A5) are the components of  $\nabla E = \mathcal{U}$  and (A6) and (A7) are the real and imaginary parts of the TE-mode eq. (A1).

Next we consider the boundary conditions. Dirichlet boundary conditions in PLTMG can be set directly. Only the Neumann conditions require attention for FOSLS. Neumann conditions apply on the top boundary in TE mode (air or ground surface) to assure continuity of the horizontal magnetic field. In general, this means that  $\partial_n E - \Phi = 0$  on  $\Gamma_n$ . Usually, the boundary  $\Gamma_n$  is the air surface  $z = -\tau$ , so the normal derivative simplifies to  $\partial_z$ . As the driving magnetic field is uniform and horizontal, the general function  $\Phi$

is taken to be the constant function  $-1$ . It is also necessary to include consistency conditions that link the derivatives to their scalar variables. Thus the component form of the Neumann condition is

$$\partial_z u - W = 0 \quad (\text{A8})$$

$$\partial_z v - Y = 0 \quad (\text{A9})$$

$$\partial_z u + 1 = 0 \quad (\text{A10})$$

$$\partial_z v + 0 = 0. \quad (\text{A11})$$

The FOSLS formulation involves solving these equations by minimizing the sum of the squares of the  $L^2$  norms of the each of the left-hand sides of the domain eqs (A2)–(A7) and  $H^{\frac{1}{2}}$  norms of the left-hand sides of the boundary eqs (A8)–(A11). If this sum of squares is forced to be zero, then all the equations will be satisfied exactly. This procedure is called minimizing the least squares functional. The  $L^2$  norm of a function  $f$  on a domain  $\Omega$  is the standard

$$\|f\|_{2,\Omega} = \left( \int_{\Omega} |f|^2 dx dz \right)^{\frac{1}{2}}.$$

The  $H^{\frac{1}{2}}$  norm of a function on the boundary  $\Gamma$  of the domain  $\Omega$  is a more technical construct, which we will not go into here because, in practice, the numerical implementation of the boundary terms uses  $L^2$  norms scaled by the inverse of the element size  $h^{-1}$  instead of exact  $H^{\frac{1}{2}}$  norms. In vector notation the least squares functional for the current system is

$$\begin{aligned} G_0(E, \mathcal{U}, \sigma, \Phi) = & \|\nabla E - \mathcal{U}\|_{2,\Omega}^2 + \left\| -\frac{i}{\omega\mu_0}\nabla\mathcal{U} - \sigma E \right\|_{2,\Omega}^2 \\ & + \frac{1}{h}\|\partial_z E - \mathcal{U}_z\|_{2,\Gamma}^2 \\ & + \frac{1}{h}\|\partial_z E - \Phi\|_{2,\Gamma}^2, \end{aligned} \quad (\text{A12})$$

where  $\mathcal{U}_z = (W \ Y)$  and the vector and matrix norms involve the sum of squares of the entries (as in the Frobenius matrix norm). Again, the numerical strategy for solving the forward problem is to minimize this functional. This gives an exact solution when  $G_0$  is zero.

However for FOSLS, simply minimizing the least-squares functional of the six domain equations and four boundary equations is not enough, since equation (A12) is not a functional whose homogeneous ( $\Phi = 0$ ) form is elliptic. In other words, we do not have a functional equivalent to a diagonal form of scalar norms, which implies that the system variables  $E$  and  $\mathcal{U}$  are still partially coupled. The FOSLS technique drops off-diagonal blocks in the Jacobian, which has no serious effects for elliptic systems. As the equations stand now, if we drop those coupling terms, the finite element discretization may become unstable and inaccurate due to the build up of errors in  $\mathcal{U}$  with vanishing divergence under the coarser grids created during the multigrid refinement process (Cai *et al.* 1997; Bank & Smith 2002). The divergence-free errors can be suppressed if we augment the system of equations. Because  $\mathcal{U} = \nabla E$ , it is given that  $\nabla \times \mathcal{U} = 0$ . Suppose that we add to eqs (A2)–(A11) the constraints:

$$\partial_z U - \partial_x W = 0 \quad (\text{A13})$$

$$\partial_z V - \partial_x Y = 0. \quad (\text{A14})$$

Note this supplies apparently redundant conditions, but does so without coupling the variables  $E$  and  $\mathcal{U}$ . It can be shown that including them by penalizing  $\|\nabla \times \mathcal{U}\|_{2,\Omega}^2$  explicitly in the functional removes divergence free, oscillatory errors (Cai *et al.* 1997). Furthermore, adding such curl equations imposes desirable continuity requirements on  $\mathcal{U}$ . The new FOSLS functional for the TE mode becomes

$$\begin{aligned} G(E, \mathcal{U}, \sigma, \Phi) = & \|\nabla E - \mathcal{U}\|_{2,\Omega}^2 + \left\| -\frac{i}{\omega\mu_0} \nabla \mathcal{U} - \sigma E \right\|_{2,\Omega}^2 \\ & + \|\nabla \times \mathcal{U}\|_{2,\Omega}^2 + \frac{1}{h} \|\partial_z E - \mathcal{U}_z\|_{2,\Gamma}^2 \\ & + \frac{1}{h} \|\partial_z E - \Phi\|_{2,\Gamma}^2. \end{aligned} \quad (\text{A15})$$

To apply PLTMG to this functional we need to express it in a weak form similar to (5). First, we express the functional (A15) as a sum of just two terms, the squared norms of the conditions on the domain and those on the boundary:

$$\begin{aligned} F(\mathbf{u}, \sigma, \Phi) = & \|\mathcal{F}\mathbf{u} - 0\|_{2,\Omega}^2 + \frac{1}{h} \|\mathcal{B}\mathbf{u} - \Phi\|_{2,\Gamma}^2 \\ = & \langle \mathcal{F}\mathbf{u}, \mathcal{F}\mathbf{u} \rangle_{2,\Omega} + \frac{1}{h} \langle \mathcal{B}\mathbf{u} - \Phi, \mathcal{B}\mathbf{u} - \Phi \rangle_{2,\Gamma}, \end{aligned}$$

where  $\mathbf{u} = (u \ U \ W \ v \ V \ Y)^T$  and the matrices  $\mathcal{F}$  and  $\mathcal{B}$  are  $8 \times 6$  and  $4 \times 6$ , respectively. In explicit form, the matrices are

$$\mathcal{F} = \begin{pmatrix} \partial_x & -1 & 0 & 0 & 0 & 0 \\ \partial_z & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \partial_x & -1 & 0 \\ 0 & 0 & 0 & \partial_z & 0 & -1 \\ -\sigma & 0 & 0 & 0 & \xi \partial_x & \xi \partial_z \\ 0 & -\xi \partial_x & -\xi \partial_z & -\sigma & 0 & 0 \\ 0 & \partial_z & -\partial_x & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \partial_z & -\partial_x \end{pmatrix}$$

and

$$\mathcal{B} = \begin{pmatrix} \partial_z & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \partial_z & 0 & -1 \\ \partial_z & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \partial_z & 0 & 0 \end{pmatrix},$$

where  $\xi = 1/(\omega\mu_0)$ . Next, consider the first-order variation of  $F$  with respect to  $\mathbf{u}$ , that is,

$$\delta F = 2\langle \mathcal{F}, \mathcal{F}\delta\mathbf{u} \rangle_{2,\Omega} + \frac{2}{h} \langle \mathcal{B}\mathbf{u} - \Phi, \mathcal{B}\delta\mathbf{u} \rangle_{2,\Gamma}.$$

The weak condition for stationarity is obtained by setting this variation to zero for a large class  $\Lambda$  of test functions. It follows that solving (A2)–(A11) is equivalent to enforcing the condition

$$A(\mathbf{u}, \Lambda) = 0 \text{ for all } \Lambda \in H_c^1, \quad (\text{A16})$$

where

$$A(\mathbf{u}, \Lambda) = 2\langle \mathcal{F}, \mathcal{F}\Lambda \rangle_{2,\Omega} + \frac{2}{h} \langle \mathcal{B}\mathbf{u} - \Phi, \mathcal{B}\Lambda \rangle_{2,\Gamma}$$

for a sufficiently large set of test functions  $\Lambda$ . Thus  $A(\mathbf{u}, \Lambda)$  may be used in place of  $A^*$  in (8) when PLTMG is applied to the vector equations. The functional  $A(\mathbf{u}, \Lambda)$  is discretized in terms of a triangular mesh analogous to the approximation of  $A^*(\mathbf{u}, \Lambda)$  for the scalar case.

In principle, when the functional  $A(\mathbf{u}, \Lambda)$  is elliptic, FOSLS can achieve optimal error and optimal multigrid convergence. A two-stage algorithm is used to solve eq. (A16). First, all equations involving  $E$  are deleted and the remaining equations are solved for  $\mathcal{U}$ . Then, the current approximation of  $\mathcal{U}$  held fixed and the equations involving  $E$  are used to solve for  $E$ . Cai *et al.* (1994, 1997) prove that the least-squares functional of these 10 equations is elliptic in a technical sense called product  $H^1$  ellipticity. This property allows the application of PLTMG to the 2-D MT inverse problem without loss of efficiency.