Geophys. J. Int. (2000) **142,** 000–000

Computational Recipes for Electromagnetic Inverse Problems: I. Mathematical Framework

Gary D. Egbert[1⋆], Anna Kelbert[1]

[1] *College of Oceanic and Atmospheric Sciences, Oregon State University, 104 COAS Admin Bldg., Corvallis, OR* 97331-5503*, USA*

SUMMARY

We present a general mathematical framework for computations involving the Jacobian of the non-linear mapping from model parameters to observations in electromagnetic (EM) geophysical inverse problems. Such computations arise in all gradient based inversion methods, including variants on Gauss-Newton and non-linear conjugate gradients. Our analysis, which is based on the discrete formulation of the forward problem, divides computations into components (data functionals, forward and adjoint solvers, model parameter mappings), and clarifies dependencies among these elements within realistic numerical inversion codes. To be concrete we focus much of the specific discussion on 2D and 3D magnetotelluric (MT) inverse problems, but our analysis is applicable to a wide range of active and passive source EM methods. The general theory developed here provides the basis for development of a modular system of computer codes for inversion of EM geophysical data, which we discuss in a companion paper.

1 INTRODUCTION

Over the past decade or so regularized inversion codes have been developed for a range of three-dimensional (3D) frequency-domain electromagnetic (EM) induction problems, including magnetotellurics (MT; e.g., *Newman and Alumbaugh* 2000; *Siripunvaraporn et al.* 2004) global geomagnetic depth sounding (*Kelbert et al.* 2008), and controlled source methods including cross-well imaging (e.g., *Alumbaugh and Newman* 1997) and marine CSEM (e.g., *Commer and Newman* 2008). Generally these efforts have been based upon minimization of a penalty functional, a sum of data misfit and model norm terms. Several distinct algorithms have been applied to solve the minimization problem, including Gauss-Newton schemes (*Mackie and Madden* 1993; *Sasaki* 2001; *Siripunvaraporn et al.* 2004) and direct gradient-based minimization schemes such as non-linear conjugate gradients (NLCG; e.g., *Newman and Alumbaugh* 2000) or quasi-Newton schemes (e.g., *Newman and Boggs* 2004; *Avdeev and Avdeeva* 2009). All of these various applications, and the different inversions algorithms that have been used, share many common elements. Here we consider these commonalities explicitly, developing a general mathematical framework for frequency domain EM inverse problems. Through this framework we provide recipes for adapting previously developed inversion algorithms to new applications, for developing extensions to standard applications (e.g., new data types, model parametrizations, regularization approaches), and for developing and testing new inversion algorithms.

This framework also provides the foundation for our development of a modular system of computer codes for frequency domain EM inverse problems. Development of this modular system, which we describe in a companion paper (*Egbert et al.* 2010), has strongly influenced the formulation presented here. In particular, this development motivates us to clearly define the fundamental objects and methods required for the generic EM inverse problem, and to analyze the dependencies among these objects. These are the key results presented in this paper.

Recently, *Pankratov and Kuvshinov* (2010) have given a general formulation for calculation of derivatives for 3D frequency domain EM problems. A general formalism for derivative calculation is also central to our development,

so in principal there is considerable overlap between their development and what is presented here. However, in contrast to *Pankratov and Kuvshinov* (2010) we adopt a purely discrete approach, assuming from the outset that the forward problem has been discretized for numerical solution, so that all spaces (EM fields, model parameters, data) are finite dimensional. The penalty functional to be minimized is explicitly taken to be a discrete quadratic form, and derivatives, adjoints, etc., are all derived for this discrete problem. Similarly, we explicitly consider the need to use discrete interpolation operators to simulate the measurement process, and to represent dependence of the discrete model operator on the unknown parameters.

There has been considerable discussion in the ocean data assimilation literature concerning the virtues of discrete vs. continuous formulations of inverse problems (e.g., *Bennett* 2002). Certainly there are some issues in inverse problems (e.g., regularity and well-posedness) that can only be understood and discussed rigorously through consideration of the problem in continuous form (e.g., *Egbert and Bennett* 1996; *Zaron et al.* 2009). However, for development of actual practical inversion algorithms there are good reasons to focus on the discrete formulation. In particular, only through a direct treatment of the discrete problem can adjoint symmetry of the numerical implementation be guaranteed. Furthermore, in discrete form many derivations are trivial, and the steps actually required for computations are often more clearly and explicitly laid out.

This paper is organized as follows. In Section 2 we summarize some common linearized EM inversion methods based on gradient based minimization of a penalty functional, demonstrating at a coarse level the basic objects used for EM inversion methods. A key component in all methods is the Jacobian of the mapping from model parameters to data; we derive general expressions for this linear operator in Section 3. In Section 4 we consider more explicitly the discretization of the governing differential equations, and the dependence of the discrete equation coefficients on the model parameter. Here we introduce specific examples of EM inverse problems (2D and 3D MT) which we will follow throughout the remaining development, and in the companion paper (*Egbert et al.* 2010). As we discuss more fully there, these EM inverse problems are sufficiently different to motivate and illustrate much of the abstraction required of a flexible modular system. With these examples as motivation, we then show in Section 5 how operations with the Jacobian can be factored into reusable components, and we consider how these components depend on each other, and on details of the EM method (e.g., sources and receivers), model parametrization, and numerical discretization. In Section 6 we consider more explicitly the form of the Jacobian when there are multiple frequencies and multiple, possibly coupled, source geometries. The closing Section serves as a link between the general developments described here, and the more specific discussions of our implementation of the modular system as a set of Fortran 95 computer codes given in the companion paper (*Egbert et al.* 2010).

2 LINEARIZED EM INVERSION: OVERVIEW

We consider regularized inversion through gradient-based minimization of a penalty functional of the form

$$\mathcal{P}(\mathbf{m},\mathbf{d}) = (\mathbf{d}-\mathbf{f}(\mathbf{m}))^T\mathbf{C}_\mathbf{d}^{-1}(\mathbf{d}-\mathbf{f}(\mathbf{m})) + \nu(\mathbf{m}-\mathbf{m}_0)^T\mathbf{C}_\mathbf{m}^{-1}(\mathbf{m}-\mathbf{m}_0) \tag{1}$$

to recover, in a stable manner, $\mathbf{m}$, an M-dimensional Earth conductivity model parameter vector, which provides an adequate fit to a data vector $\mathbf{d}$ of dimension N_d. In (1), $\mathbf{C_d}$ is the covariance of data errors, $\mathbf{f}(\mathbf{m})$ defines the forward mapping, $\mathbf{m}_0$ is a prior or first guess model parameter, ν is a trade-off parameter, and $\mathbf{C_m}$ (or more properly $\nu^{-1}\mathbf{C_m}$) defines the model covariance or regularization term. In practice $\mathbf{C_d}$ is always taken to be diagonal, so by a simple rescaling of the data and forward mapping ($\mathbf{C_d}^{-1/2}\mathbf{d}$, $\mathbf{C_d}^{-1/2}\mathbf{f}$), we may eliminate $\mathbf{C_d}^{-1}$ from the definition of $\mathcal{P}$.

The prior model $\mathbf{m}_0$ and model covariance $\mathbf{C_m}$ can also be eliminated from (1) by the affine linear transformation of the model parameter $\tilde{\mathbf{m}} = \mathbf{C_m}^{-1/2}(\mathbf{m}-\mathbf{m}_0)$, and forward mapping $\tilde{\mathbf{f}}(\tilde{\mathbf{m}}) = \mathbf{f}(\mathbf{C_m}^{1/2}\tilde{\mathbf{m}}+\mathbf{m}_0)$, reducing (1) to

$$\mathcal{P}(\tilde{\mathbf{m}},\mathbf{d}) = (\mathbf{d}-\tilde{\mathbf{f}}(\tilde{\mathbf{m}}))^T(\mathbf{d}-\tilde{\mathbf{f}}(\tilde{\mathbf{m}})) + \nu\tilde{\mathbf{m}}^T\tilde{\mathbf{m}} = ||\mathbf{d}-\tilde{\mathbf{f}}(\tilde{\mathbf{m}})||^2 + \nu||\tilde{\mathbf{m}}||^2. \tag{2}$$

After minimizing (2) over $\tilde{\mathbf{m}}$, the model parameter in the untransformed space can be recovered as $\mathbf{m} = \mathbf{C_m}^{1/2}\tilde{\mathbf{m}}+\mathbf{m}_0$. Note that this model space transformation is in fact quite practical if instead of following the usual practice of defining $\mathbf{C_m}^{-1} = \mathbf{D}^T\mathbf{D}$, where $\mathbf{D}$ is a discrete representation of a gradient or higher order derivative operator, the regularization is formulated directly in terms of a smoothing operator (i.e., model covariance) $\mathbf{C_m}$. It is relatively easy to construct computationally efficient positive definite discrete symmetric smoothing operators for regularization (e.g., *Egbert* 1994; *Siripunvaraporn and Egbert* 2000; *Chua* 2001). Although the resulting covariance matrix $\mathbf{C_m}$ will not generally

be sparse and may not be practical to invert, all of the computations required for gradient evaluations and for minimization of the transformed penalty functions require only multiplication by the smoothing operator $\mathbf{C}_\mathbf{m}^{1/2}$ (i.e, half of the smoothing of $\mathbf{C}_\mathbf{m}$). It is also quite possible to define the covariance so that multiplication by either $\mathbf{C}_\mathbf{m}$ or $\mathbf{C}_\mathbf{m}^{-1}$ is practical. In the following we focus on the simplified "canonical" penalty functional (2), with tildes omitted.

We begin by summarizing some standard approaches to gradient-based minimization of (2) using a consistent notation. *Siripunvaraporn and Egbert* (2000); *Rodi and Mackie* (2001); *Newman and Boggs* (2004); *Avdeev* (2005) provide further details and discussion on these and related methods. A key component in all of these linearized search schemes is the $N_d \times M$ Jacobian, or sensitivity matrix, which we denote $\mathbf{J}$. This gives the derivative of $\mathbf{f}$ with respect to the model parameters, with $J_{ij} = \partial f_i / \partial m_j$. *Newman and Alumbaugh* (1997); *Spitzer* (1998); *Rodi and Mackie* (2001) provide detailed expressions for $\mathbf{J}$ for some specific EM inverse problems, and we will consider the general case extensively below (Section 3).

Search for a minimizer of (2) using $\mathbf{J}$ is iterative, as for example in the classical Gauss-Newton (GN) method. Let $\mathbf{m}_n$ be the model parameter at the nth iteration, $\mathbf{J}$ the sensitivity matrix evaluated at $\mathbf{m}_n$, and $\mathbf{r} = \mathbf{d} - \mathbf{f}(\mathbf{m}_n)$ the data residual. Then, linearizing the penalty functional in the vicinity of $\mathbf{m}_n$ for small perturbations $\delta\mathbf{m}$ leads to the $M \times M$ system of normal equations

$$(\mathbf{J}^T\mathbf{J} + \nu\mathbf{I})\delta\mathbf{m} = \mathbf{J}^T\mathbf{r} - \nu\mathbf{m}_n, \tag{3}$$

which can be solved for $\delta\mathbf{m}$ to yield a new trial solution $\mathbf{m}_{n+1} = \mathbf{m}_n + \delta\mathbf{m}$. As discussed in *Parker* (1994), this basic linearized scheme generally requires some form of step length damping for stability (e.g., a Levenberg-Marquardt approach; *Marquardt* 1963; *Rodi and Mackie* 2001).

There are many variants to this basic algorithm. For example, in the Occam approach (*Constable et al.* 1987; *Parker* 1994), (3) is rewritten as

$$(\mathbf{J}^T\mathbf{J} + \nu\mathbf{I})\mathbf{m} = \mathbf{J}^T\hat{\mathbf{d}}, \tag{4}$$

where $\hat{\mathbf{d}} = \mathbf{d} - \mathbf{f}(\mathbf{m}_n) + \mathbf{J}\mathbf{m}_n$. Although $\mathbf{m}_{n+1}$ is obtained directly as the solution to (4) the result is exactly equivalent to solving (3) for the change in the model at step $n+1$, and adding the result to $\mathbf{m}_n$. A more substantive difference is that in the Occam scheme step length control is achieved by varying ν, computing a series of trial solutions, and choosing the regularization parameter so that data misfit is minimized. An advantage of this approach is that ν is determined as part of the search process, and at convergence one is assured that the solution attains at least a local minimum of the model norm $||\mathbf{m}|| = (\mathbf{m}^T\mathbf{m})^{1/2}$, subject to the data fit attained (*Parker* 1994). The Occam scheme can also be implemented in the data space (*Siripunvaraporn and Egbert* 2000; *Siripunvaraporn et al.* 2005). The solution to (4) can be written as

$$\mathbf{m}_{n+1} = \mathbf{J}^T\mathbf{b}_n; \quad (\mathbf{J}\mathbf{J}^T + \nu\mathbf{I})\mathbf{b}_n = \hat{\mathbf{d}}, \tag{5}$$

as can be verified by substituting (5) into (4) and simplifying. This approach requires solving an $N_d \times N_d$ system of equations in the data space instead of the $M \times M$ model space system of equations (4), and can thus be more efficient if the model is heavily over-parametrized.

Computing the full Jacobian $\mathbf{J}$ required for any direct GN algorithm is a very demanding computational task for multi-dimensional EM problems, since (as we shall see in Section 3) the equivalent of one forward solution is required for each row (or column) of $\mathbf{J}$. An alternative is to solve the normal equations (4) or (5) with a memory efficient iterative Krylov-space solver such as conjugate gradients (CG). This requires computation of matrix-vector products such as $[\mathbf{J}^T\mathbf{J} + \nu\mathbf{I}]\mathbf{m}$, which can be accomplished without forming or storing $\mathbf{J}$ at the cost of two forward solutions (e.g., *Mackie and Madden* 1993). *Mackie and Madden* (1993), *Zhang et al.* (1995), *Newman and Alumbaugh* (1997), *Rodi and Mackie* (2001) and others have used CG to solve (3), while *Siripunvaraporn and Egbert* (2007) have applied the same approach to the corresponding data space equations of (5).

The GN schemes requires solving normal equations derived from a quadratic approximation to (1). Alternatively, the penalty functional can be directly minimized using a gradient-based optimization algorithm such as non-linear conjugate gradients (NLCG; e.g., *Rodi and Mackie* 2001; *Newman and Boggs* 2004; *Avdeev* 2005; *Kelbert et al.* 2008).

With this NLCG approach one must evaluate the gradient of (1) with respect to variations in model parameters $\mathbf{m}$:

$$\left.\frac{\partial \mathcal{P}}{\partial \mathbf{m}}\right|_{\mathbf{m}_n} = -2\,\mathbf{J}^T\mathbf{r} + 2\,\nu\mathbf{m}_n. \tag{6}$$

The gradient is then used to calculate a new "conjugate" search direction in the model space. After minimizing the penalty functional along this direction using a line search which requires a few evaluations of the forward operator, the gradient is recomputed. NLCG again utilizes essentially the same basic computational steps as required for solving the linearized Equations (3). In particular, the forward problem must be solved to evaluate $\mathbf{f}(\mathbf{m})$ and the residual $\mathbf{r}$ must be multiplied by $\mathbf{J}^T$. Quasi-Newton schemes (e.g., *Nocedal and Wright* 1999; *Newman and Boggs* 2004; *Haber* 2005; *Avdeev* 2005) provide an alternative approach to NLCG for direct minimization of (1), with similar advantages with regard to storage and computation of the Jacobian, and similar computational requirements.

All of these schemes for minimizing (1) can be abstractly expressed in terms of a small number of basic "objects" (data and model parameter vectors, $\mathbf{d}$ and $\mathbf{m}$), and operators (the forward mapping $\mathbf{f}(\mathbf{m})$, multiplication by the corresponding Jacobian $\mathbf{J}$ its transpose $\mathbf{J}^T$ (and, implicitly, the data and model covariances $\mathbf{C_m}$ and $\mathbf{C_d}$). Given modular computer codes which implement these basic objects, any of the inversion algorithms outlined here, as well as many variants, are readily implemented. In the next sections we analyze further the discrete forward problem, and provide a finer-grained general formulation of the modules required to implement essentially any linearized inverse scheme for any EM problem. In particular, we provide "recipes" for $\mathbf{J}$ in terms of more basic objects associated with the model parametrization, the forward solver, and the numerical simulation of the observation operators.

3 DATA SENSITIVITIES

The numerical discretization of the frequency domain EM partial differential equation (PDE) is written generically as

$$\mathbf{S_m e} = \mathbf{b} \tag{7}$$

where the vector $\mathbf{b}$ gives appropriate boundary and forcing terms for the particular EM problem, $\mathbf{e}$ is the N_e dimensional vector representing the discretized electric and/or magnetic fields (or perhaps potential functions), and $\mathbf{S_m}$ is an $N_\mathbf{e} \times N_\mathbf{e}$ coefficient matrix which depends on the M dimensional model parameter $\mathbf{m}$. We take $\mathbf{e}$ to represent both interior and boundary components of the discrete solution vector, so that any boundary conditions required for the problem are included in $\mathbf{b}$. Given a solution $\mathbf{e}$ to (7), data can always be written in the general form

$$d_j = f_j(\mathbf{m}) + \epsilon_j = \psi_j(\mathbf{e}(\mathbf{m}), \mathbf{m}) + \epsilon_j, \tag{8}$$

where ψ_j is some generally non-linear, but usually simple, function of the components of $\mathbf{e}$ (and possibly $\mathbf{m}$), and ϵ_j represents data error.

With this general setup we have, by the chain rule

$$J_{jk} \equiv \frac{\partial f_j}{\partial m_k} = \sum_l \frac{\partial \psi_j}{\partial e_l}\frac{\partial e_l}{\partial m_k} + \frac{\partial \psi_j}{\partial m_k} \quad . \tag{9}$$

Letting $\mathbf{F}, \mathbf{L}, \mathbf{Q}$ be the partial derivative matrices

$$F_{lk} = \left.\frac{\partial e_l}{\partial m_k}\right|_{\mathbf{m}_0} \qquad L_{jl} = \left.\frac{\partial \psi_j}{\partial e_l}\right|_{\mathbf{e}_0,\mathbf{m}_0} \qquad Q_{jk} = \left.\frac{\partial \psi_j}{\partial m_k}\right|_{\mathbf{e}_0,\mathbf{m}_0}, \tag{10}$$

where $\mathbf{e}_0$ is the solution to (7) for model parameter $\mathbf{m}_0$, the Jacobian at $\mathbf{m}_0$ can be written in matrix notation as

$$\mathbf{J} = \mathbf{LF} + \mathbf{Q}. \tag{11}$$

The jth row of $\mathbf{L}$ represents the linearized data functional, which is applied to the perturbation in the EM solution to compute the perturbation in d_j. These row vectors are generally very sparse, supported only on a few nodes surrounding the corresponding data site. When the observation functionals are independent of the model parameters (as they often are) $\mathbf{Q} \equiv \mathbf{0}$. When $\mathbf{Q}$ is required it is also typically sparse, but this depends on the specific nature of the model parametrization. Although, as we show below, derivation of expressions for $\mathbf{L}$ and $\mathbf{Q}$ can be quite involved for realistic EM data functionals, calculation of $\mathbf{F}$ presents the only real computational burden.

To derive a general expression for $\mathbf{F}$, differentiate (7) at $\mathbf{m}_0$ with respect to the model parameters $\mathbf{m}$. We assume that $\mathbf{b}$ is constant, independent of $\mathbf{m}$, although, as discussed in Appendix C (see also *Newman and Boggs* (2004))

some subtle issues related to this point may arise with specific solution approaches. Then letting $\mathbf{e}_0$ be the solution of (7) at $\mathbf{m}_0$, and noting that the EM solution $\mathbf{e}$ varies as $\mathbf{m}$ is varied, we obtain

$$\mathbf{S}_{\mathbf{m}_0}\left[\left.\frac{\partial \mathbf{e}}{\partial \mathbf{m}}\right|_{\mathbf{m}=\mathbf{m}_0}\right] = -\left.\frac{\partial}{\partial \mathbf{m}}\left(\mathbf{S}_{\mathbf{m}}\mathbf{e}_0\right)\right|_{\mathbf{m}=\mathbf{m}_0}, \tag{12}$$

or

$$\mathbf{S}_{\mathbf{m}_0}\mathbf{F} = \mathbf{P}. \tag{13}$$

The $N_e \times M$ matrix $\mathbf{P}$ depends on details of both the numerical model implementation and the conductivity parametrization (as discussed below), but is in general inexpensive to calculate, once the EM solution $\mathbf{e}_0$ has been computed. Putting together (11) and (13), we obtain an expression for the numerical Jacobian (or sensitivity matrix) $\mathbf{J}$ for general EM problems:

$$\mathbf{J} = \mathbf{L}\mathbf{S}_{\mathbf{m}_0}^{-1}\mathbf{P} + \mathbf{Q}. \tag{14}$$

Computing all of $\mathbf{J}$ would appear to require solving the induction equation once for each of the M columns of $\mathbf{P}$. However, simply taking the transpose of (14) we obtain

$$\mathbf{J}^T = \mathbf{P}^T[\mathbf{S}_{\mathbf{m}_0}^T]^{-1}\mathbf{L}^T + \mathbf{Q}^T, \tag{15}$$

so the sensitivity matrix can in fact be obtained by solving the transposed discrete EM system N_d times (once for each column of $\mathbf{L}^T$), the usual "reciprocity" trick for efficient calculation of sensitivities (e.g., *Rodi* 1976; *de Lugao et al.* 1997). It should also be emphasized that for many of the inversion algorithms described in Section 2, $\mathbf{J}$ is not explicitly calculated. Instead a series of multiplications of model space vectors by $\mathbf{J}$ and/or data space vectors by $\mathbf{J}^T$ are required. These in turn require implementation of the component operators $\mathbf{P}$, $\mathbf{L}$, $\mathbf{Q}$ and the solver $\mathbf{S}_{\mathbf{m}_0}^{-1}$, together with the adjoints (or transposes) of these operators.

The EM equations are self-adjoint (except for time reversal) with respect to the usual L^2 inner product (i.e., reciprocity holds). For now leaving aside complications regarding boundary conditions (these are discussed in Appendix B) this implies that on a uniform grid operator $\mathbf{S}_{\mathbf{m}}$ is symmetric. For more general grids the fact that the EM operator is self-adjoint implies

$$\mathbf{S}_{\mathbf{m}}^T = \mathbf{V}\mathbf{S}_{\mathbf{m}}\mathbf{V}^{-1}, \tag{16}$$

where $\mathbf{V}$ is a diagonal matrix of integration volume elements for the natural discrete representation of the L_2 integral inner product on the model domain (see Appendix B). Eq. (16) implies $\mathbf{S}_{\mathbf{m}}^T\mathbf{V} = \mathbf{V}\mathbf{S}_{\mathbf{m}}$ is a symmetric (though not Hermitian) matrix. It is easier to compute solutions to this symmetrized problem, so solutions to the forward problem are generally computed as $\mathbf{e} = (\mathbf{V}\mathbf{S}_{\mathbf{m}})^{-1}\mathbf{V}\mathbf{b}$ (e.g., see *Uyeshima and Schultz* 2000). The solution for the adjoint problem can also be written in terms of the symmetrized inverse operator as $\mathbf{e} = (\mathbf{S}_{\mathbf{m}}^T)^{-1}\mathbf{b} = \mathbf{V}(\mathbf{V}\mathbf{S}_{\mathbf{m}})^{-1}\mathbf{b}$; the principal difference from the forward case is thus the order in which multiplication by the diagonal matrix $\mathbf{V}$ and the symmetrized solver are called. In general the adjoint solver $(\mathbf{S}_{\mathbf{m}}^T)^{-1}$ for EM problems is trivial to implement, once a suitably general forward solver is available.

Before proceeding, two general points require discussion. First, we note that many of the computations in frequency domain EM problems are most efficiently implemented (and described) using complex arithmetic, but the model conductivity parameter $\mathbf{m}$ is real. Data might be complex (an impedance) or real (an apparent resistivity and/or phase). As already implicit in our formulation of the penalty functional (1), we formally assume that all data are real, i.e., real and imaginary parts of a complex observation are separate elements of the real data vector $\mathbf{d}$, and that the basic operators $\mathbf{f}$ and $\mathbf{J}$ have been recast as real mappings from model parameter to data vector. However, we will frequently use complex notation and we will often be somewhat casual in moving between real and complex variables. Thus, for example, the frequency domain forward problem (7) will generally be formulated and solved in terms of complex variables. $\mathbf{P}$ will then be a mapping from the real parameter space to the complex space of forcings for the forward problem, while $\mathbf{L}$ will be a mapping from a complex, back to a real space. Both $\mathbf{P}$ and $\mathbf{L}$ can be most conveniently represented by complex matrices, with the convention that for $\mathbf{L}$ only the real part of the matrix-vector product is retained. We discuss this more explicitly in Appendix A.

Second, in most cases EM data are obtained for a large number of distinct sources, i.e., different frequencies and/or different current source geometries. For example in the case of MT there are data for two source polarizations and a

wide range of frequencies, while for controlled source problems there may be a multiplicity of transmitter geometries or locations. Each of these distinct sources, which we refer to in general as "transmitters", requires solving a separate forward problem. In most, but not all, cases these forward problems are decoupled, so the data vector and forward modeling operator can be decomposed into $t = 1, \ldots, N_T$ blocks, one for each transmitter:

$$\mathbf{d} = \begin{pmatrix} \mathbf{d}_1 \\ \vdots \\ \mathbf{d}_{N_T} \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_{N_T} \end{pmatrix}. \tag{17}$$

Here $\mathbf{d}_t$ gives the data associated with a group of "receivers", consisting of possibly multiple components, at multiple locations, all making observations of fields generated by transmitter t. Thus, with multiple (decoupled) transmitters the Jacobian can be partitioned into N_T blocks in the obvious way, and each block can be represented as in (14), so that the full sensitivity matrix can be expressed as

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_1 \\ \vdots \\ \mathbf{J}_{N_T} \end{pmatrix} = \begin{pmatrix} \mathbf{L}_1\mathbf{S}_{1,\mathbf{m}}^{-1}\mathbf{P}_1 + \mathbf{Q}_1 \\ \vdots \\ \mathbf{L}_{N_T}\mathbf{S}_{N_T,\mathbf{m}}^{-1}\mathbf{P}_{N_T} + \mathbf{Q}_{N_T} \end{pmatrix}. \tag{18}$$

The matrices $\mathbf{P}_t$ and $\mathbf{Q}_t$ generally depend on the solution for transmitter t. If the transmitter t only specifies the source geometry, the differential operator for the PDE $\mathbf{S}_{t,\mathbf{m}}$ will be independent of the transmitter; however, in general the transmitter will also define the forward problem to solve. An obvious example is the MT case, where the operator depends on frequency. Only $\mathbf{L}_t$ and $\mathbf{Q}_t$ depend on the configuration of receivers; these also in general depend on the forward solution, and thus on transmitter t.

There is an important complication to the simple form of (18), most clearly illustrated by the case of 3D MT. Here evaluation of the forward operator for an impedance tensor element requires solutions for the pair of transmitters associated with two uniform source polarizations. Thus, for 3D MT the rows of the Jacobian corresponding to a single frequency are formed from components for two transmitters, corresponding to N-S and E-W polarized uniform sources, coupled through the linearized measurement operators $\mathbf{L}$ and $\mathbf{Q}$

$$\mathbf{J} = \sum_{t=1}^{2} \left[\mathbf{L}_t\mathbf{S}_{\mathbf{m}}^{-1}\mathbf{P}_t + \mathbf{Q}_t\right] \tag{19}$$

The same complication would arise for any plane-wave source transfer function (e.g., vertical, or inter-site magnetic). Other examples of multiple polarization EM inverse problems can be imagined, e.g., allowing for a combination of horizontal spatial gradients and uniform sources (e.g., *Egbert* 2002; *Schmucker* 2004; *Semenov and Shuman* 2009); would require allowing for 5 coupled sources. *Pankratov and Kuvshinov* (2010) discuss the general multiple polarization problem from a theoretical perspective, although to our knowledge, no actual applications of the theory to inversion of real data sets have yet been reported, beyond the standard two polarization uniform source case.

We return to the coupled multiple polarization case in Section 5.2, where we discuss measurement operators in more detail. Then, in Section 6, we consider the general multiple transmitter case further, and show more explicitly how the source and receiver configuration can result in special structure for the Jacobian, which can be exploited to improve computational efficiency. For the next few sections we focus on the simpler case of a single transmitter, as we develop the basic building blocks for more complex and realistic problems.

4 DISCRETIZATION OF THE FORWARD PROBLEM

To derive more explicit expressions for the operators $\mathbf{L}$, $\mathbf{P}$ and $\mathbf{Q}$, and hence for the full Jacobian, more specific assumptions about the numerical implementation of the forward problem (7) are required. To motivate the general development we consider as examples two specific cases in detail: inversion of 2D and 3D MT data. We discuss most explicitly a finite difference (FD) modeling approach, though most of the results obtained are more broadly applicable.

Numerical schemes for solving Maxwell's equations are often most elegantly formulated in terms of a pair of vector fields defined on conjugate grids. For example, the space of primary fields which we denote as $\mathcal{S}_P$ may represent the electric fields, while the space of dual fields, denoted $\mathcal{S}_D$, represents the magnetic fields. Even when the coupled first-order system (i.e., Maxwell's equations) is reduced to a second order equation involving only the primary field, it

is worthwhile to explicitly consider the dual field also. Most obviously, in many applications both electric and magnetic field components are required to evaluate the data functionals. Furthermore, depending on the model formulation, the dependence of the discrete PDE operator coefficients on the model parameter can generally be represented most explicitly through a mapping $\pi(\mathbf{m})$ from the model parameter space $\mathcal{M}$ – sometimes to $\mathcal{S}_P$, but in other cases to $\mathcal{S}_D$, and a general treatment should allow for both cases. Boundary conditions are of course a critical part of the formulation of the forward problem. These are included implicitly in our generic formula of the forward problem (7). In the following we omit technical details concerning implementation of boundary conditions, leaving discussion of these issues to Appendix B.

As a first illustration we consider FD modeling of the 3D quasi-static Maxwell's equations appropriate for MT. In the frequency domain (assuming a time dependence of $e^{i\omega t}$) the magnetic fields can be eliminated, resulting in a second order elliptic system of PDEs in terms of the electric fields alone

$$\nabla \times \nabla \times \mathbf{E} + i\omega\mu\sigma\mathbf{E} = 0, \tag{20}$$

with the tangential components of $\mathbf{E}$ specified on all boundaries. To solve (20) numerically in 3D, we consider a FD approximation on a staggered grid of dimension $N_x \times N_y \times N_z$, as illustrated in Figure 1 (e.g., *Yee* 1966; *Smith* 1996; *Siripunvaraporn et al.* 2002). In the staggered grid formulation the discretized electric field vector components are defined on cell edges (Figure 1). In our terminology, the primary field space $\mathcal{S}_P$ is the space of such finite dimensional cell edge vector fields. A typical element will be denoted by $\mathbf{e}$. As illustrated in Figure 1, the magnetic fields, which in continuous form satisfy $\mathbf{B} = (-i\omega)^{-1}\nabla \times \mathbf{E}$, are naturally defined on the discrete grid on cell faces. The dual field space $\mathcal{S}_D$ is thus the space of discrete vector fields defined on faces. A typical element of this space will be denoted by $\mathbf{h}$.

In the staggered grid FD discretization used for (20), the discrete magnetic and electric fields are related via

$$\mathbf{h} = (-i\omega)^{-1}\mathbf{C}\,\mathbf{e}, \tag{21}$$

where $\mathbf{C} : \mathcal{S}_P \mapsto \mathcal{S}_D$ is the discrete approximation of the curl of cell edge vectors, and (20) can be expressed in its discrete form as

$$\Big[\mathbf{C}^\dagger\mathbf{C} + \mathit{diag}(i\omega\mu\sigma(\mathbf{m}))\Big]\,\mathbf{e} = 0. \tag{22}$$

Here $\mathit{diag}(\mathbf{v})$ denotes a diagonal matrix with the components of the vector $\mathbf{v}$ on the diagonal, and $\mathbf{C}^\dagger : \mathcal{S}_D \mapsto \mathcal{S}_P$ is the discrete curl mapping interior cell face vectors to interior cell edges. As the notation indicates this operator is the adjoint of $\mathbf{C}$, relative to appropriate (i.e., volume weighted) inner products on the spaces $\mathcal{S}_D$ and $\mathcal{S}_P$. Although $\mathbf{e}$ is the full solution vector (including boundary components), (22) provides equations only for the interior nodes. Additional equations are required to constrain $\mathbf{e}$ on the boundary, and to complete specification of the discrete forward operator $\mathbf{S_m}$. These details, and further discussion of $\mathbf{C}$ and its adjoint, are provided in Appendix B. The key point here is that the dependence of the operator coefficients on the model parameter (which we take to be an element of some finite dimensional space $\mathcal{M}$) is made explicit through the mapping $\sigma : \mathcal{M} \mapsto \mathcal{S}_P$ in (22).

The 3D EM induction forward problem can also be formulated in terms of magnetic fields

$$\nabla \times \rho\nabla \times \mathbf{B} + i\omega\mu\mathbf{B} = 0, \tag{23}$$

now with the tangential component of the magnetic fields specified on boundaries. With this formulation (e.g., *Mackie et al.* 1994; *Uyeshima and Schultz* 2000), $\mathbf{B}$ would be the primary field, and the electric field $\mathbf{E} = \rho\nabla \times \mathbf{B}$ would be the dual field. Using an analogous staggered grid FD discretization, with magnetic field components defined on cell edges, and electric field components defined on cell faces, the discrete form of the induction equation now takes the form

$$\Big[\mathbf{C}^\dagger\,\mathit{diag}(\rho(\mathbf{m}))\mathbf{C} + \mathit{diag}(i\omega\mu)\Big]\,\mathbf{e} = 0. \tag{24}$$

In this case the dependence of the coefficients on model parameter $\mathbf{m}$ is made explicit through the mapping to the dual field space $\rho : \mathcal{M} \mapsto \mathcal{S}_D$.

It is also instructive to consider the 2D MT inverse problem. Now there are effectively two distinct modeling problems: for TE and TM modes, with electric and magnetic fields, respectively, parallel to the geologic strike. The TE mode case is essentially identical to the 3D electric field formulation of (20-22)). The TM mode case, which is solved in terms of the magnetic field instead of the electric field, is more instructive with regard to generalization. In the TM mode the magnetic field parallels the geological strike (x) and (23) can be reduced to a scalar PDE in the

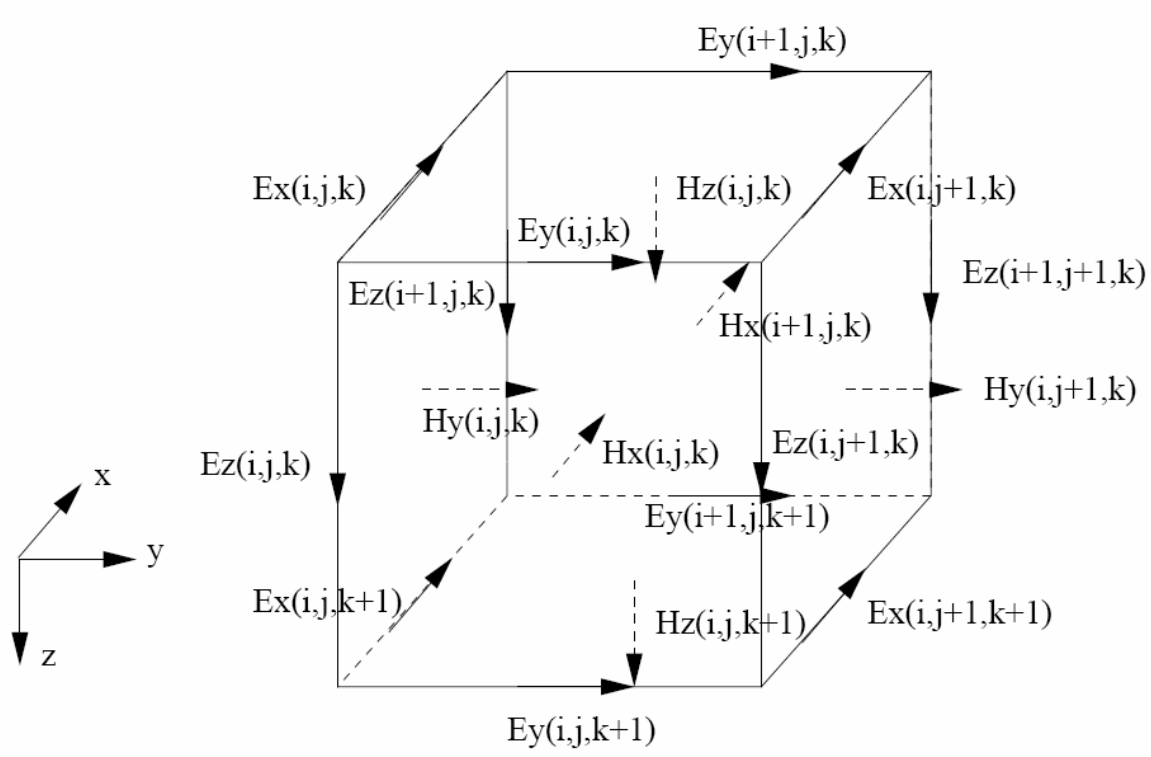

Figure 1. Staggered finite difference grid for the 3D MT forward problem. Electric field components defined on cell edges are the primary EM field component, which the PDE is formulated in terms of. The magnetic field components can be defined naturally on the cell faces; these are the secondary EM field in this numerical formulation.

$y-z$ plane

$$\partial_y \rho \partial_y B_x + \partial_z \rho \partial_z B_x + i\omega\mu B_x = 0, \tag{25}$$

with B_x specified on boundaries.

As for the 3D problems, for the discrete 2D problem we can define finite dimensional spaces of primary ($\mathcal{S}_P$) and dual ($\mathcal{S}_D$) EM fields. Now the primary field is B_x, defined on the nodes (corners) of the 2D grid, and the dual fields are the electric field components E_y and E_z defined on the vertical and horizontal cell edges (Figure 2). A natural centered finite difference approximation of (25) can be written in terms of a discrete 2D gradient operator $\mathbf{G} : \mathcal{S}_P \mapsto \mathcal{S}_D$ and a 2D divergence operator $\mathbf{D} : \mathcal{S}_D \mapsto \mathcal{S}_P$. Using $\mathbf{e} \in \mathcal{S}_P$ to denote the primary discrete EM field solution (B_x) we have a more explicit form of (7) for this discrete TM mode implementation,

$$[\mathbf{D}\, diag\,(\rho(\mathbf{m}))\, \mathbf{G} + i\omega\mu\mathbf{I}]\, \mathbf{e} = 0, \tag{26}$$

with additional equations again required to specify boundary data.

In most other finite difference or finite volume modeling approaches, e.g. with Maxwell's equations cast in terms of vector potentials, similar (although potentially more complicated) sets of conjugate spaces can be defined, the differential operator can be decomposed into discrete approximations to first order linear differential operators which map between conjugate grids, and the dependence of discrete operator coefficients on an abstract model parameter space can be described explicitly through a mapping $\pi : \mathcal{M} \mapsto \mathcal{S}_{P,D}$. Finite element approaches to EM modeling will result in similar structures.

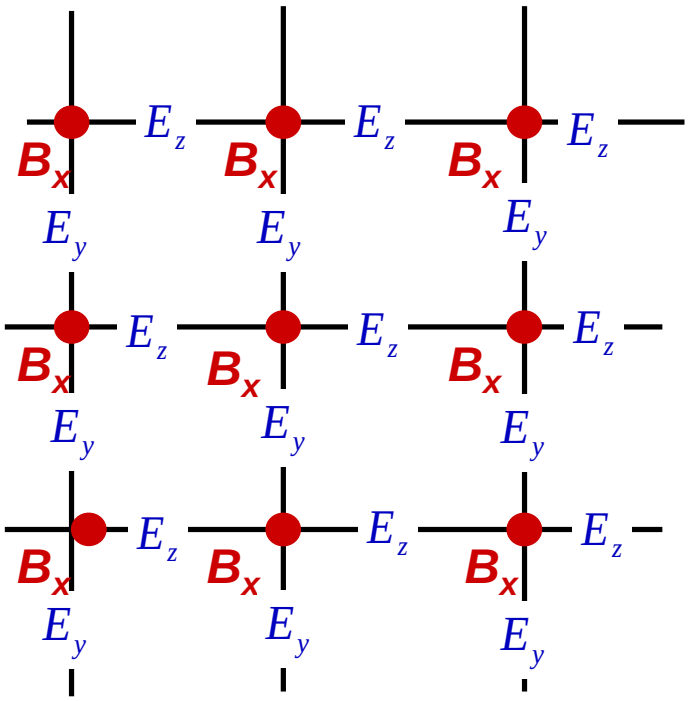

Figure 2. Finite difference grid for the 2D MT TM mode forward problem. The scalar B_x magnetic field, defined on 2D cell corners is the primary field. The secondary field components are E_y and E_z, defined on vertical and horizontal cell edges, respectively.

5 COMPONENTS OF THE JACOBIAN

5.1 Matrix P

We can give an explicit expression for the operator $\mathbf{P}$, assuming the forward operator $\mathbf{S_m}$ can be written in the general form

$$\mathbf{S_m e} \equiv \mathbf{S}_0\mathbf{e} + \mathbf{U}\,(\pi(\mathbf{m}) \circ \mathbf{Ve}), \tag{27}$$

where $\mathbf{S}_0$, $\mathbf{U}$ and $\mathbf{V}$ are some linear operators that do not depend on the model parameter vector $\mathbf{m}$, $\pi(\mathbf{m})$ is a (possibly non-linear) operator that maps the model parameter space $\mathcal{M}$ to the primary or dual grid, and ($\circ$) denotes the component-wise multiplication of the two vectors in $\mathcal{S}_{P,D}$ (also known as the Hadamard product). Note that on a finite difference grid, $\mathbf{S_m}$ (and hence $\mathbf{S}_0$ and $\mathbf{V}$) act on a full solution vector that includes both the interior and boundary edges (see Appendix B). All of the examples outlined above are special cases of (27), as we will discuss.

Assuming (27) and comparing (12) and (13) we find

$$\mathbf{P} = -\frac{\partial}{\partial \mathbf{m}} \left(\mathbf{S_m e}_0 \right) \Big|_{\mathbf{m}_0} = -\mathbf{U} \left(\frac{\partial \pi}{\partial \mathbf{m}} \Big|_{\mathbf{m}_0} \circ \mathbf{V e}_0 \right) \tag{28}$$

$$= -\mathbf{U} \left(\mathbf{V e}_0 \circ \frac{\partial \pi}{\partial \mathbf{m}} \Big|_{\mathbf{m}_0} \right) \tag{29}$$

$$= -\mathbf{U}\, diag(\mathbf{V e}_0)\, \frac{\partial \pi}{\partial \mathbf{m}} \Big|_{\mathbf{m}_0} \tag{30}$$

$$\tag{31}$$

Writing $\Pi_{\mathbf{m}_0}$ for the Jacobian of the (in general non-linear) model parameter mapping $\pi(\mathbf{m})$ evaluated at the background model parameter $\mathbf{m}_0$, we have

$$\mathbf{P} = -\mathbf{U}\, diag(\mathbf{V e}_0) \Pi_{\mathbf{m}_0} \tag{32}$$

$$\mathbf{P}^T = -\Pi_{\mathbf{m}_0}^T\, diag(\mathbf{V e}_0) \mathbf{U}^T. \tag{33}$$

Note that only the operator $\Pi_{\mathbf{m}}$ depends on the details of the model parametrization–the other terms depend only on the numerical discretization of the governing equations. Equation (32) and (33) provide broadly applicable recipes for implementation of the operators $\mathbf{P}$ and $\mathbf{P}^T$, as illustrated in the following examples. If the dependence of the forward operator on the model parameter cannot be cast as a special case of (27), similar formal steps could almost certainly be used to derive appropriate expressions for these operators.

5.1.1 Example: 2D MT

For the 2D TM problem (26), the PDE coefficients depend on the model parameters through $\rho : \mathcal{M} \mapsto \mathcal{S}_D$, i.e., the resistivity $\rho(\mathbf{m})$ defined on the dual grid, cell edges. To be specific, we consider the simplest model parametrizations, with conductivity or log conductivity for each cell in the numerical grid an independent parameter. From physical considerations, it is most reasonable to compute the required edge resistivities from cell conductivities by first transforming to resistivity, and then computing the area weighted average of resistivities of the two cells on either side of the edge. Representing the averaging operator from 2D cells to cell sides as $\mathbf{W}_{\rm TM}$, and letting $(\mathbf{m})^{-1}$ denote the component-wise inverse of the model parameter vector, we then have

$$\rho(\mathbf{m}) = \mathbf{W}_{\rm TM} (\mathbf{m})^{-1} \tag{34}$$

$$\rho(\mathbf{m}) = \mathbf{W}_{\rm TM} \mathbf{exp}(-\mathbf{m}), \tag{35}$$

for linear and log conductivity respectively. The model operator of (26) can be cast in the general form of (27) with the identifications $\mathbf{S}_0 \equiv -i\omega\mu\mathbf{I}$, $\mathbf{U} \equiv \mathbf{D}$, $\mathbf{V} \equiv \mathbf{G}$, and $\pi(\mathbf{m}) \equiv \rho(\mathbf{m})$. Thus we obtain the expressions for $\mathbf{P}$ and $\mathbf{P}^T$ in the 2D TM mode case:

$$\mathbf{P} = -\mathbf{D}\, diag(\mathbf{G e}_0) \Pi_{\mathbf{m}_0} \tag{36}$$

$$\mathbf{P}^T = -\Pi_{\mathbf{m}_0}^T\, diag(\mathbf{G e}_0) \mathbf{D}^T, \tag{37}$$

where $\Pi_{\mathbf{m}_0} = -\mathbf{W}_{\rm TM} [\, diag(\mathbf{m}_0)\,]^{-2}$ for the parametrization in terms of linear conductivity, or $\Pi_{\mathbf{m}_0} = -\mathbf{W}_{\rm TM} [\, diag(\mathbf{exp}(-\mathbf{m}_0))\,]$ for log conductivity.

5.1.2 Example: 3D MT

We again assume the simplest model parametrization, with conductivity, or the natural logarithm of conductivity, specified independently for each of the $M = N_x N_y N_z$ cells in the numerical grid. The discrete operator of (22) requires conductivity defined on cell edges, where the electric field components are defined. For physical consistency (current should be conserved) the edge conductivities should represent the volume weighted average of the surrounding four cells. Let $\mathbf{W}$ be the $N_e \times M$ matrix representing this weighted averaging operator, a mapping from $\mathcal{M}$ to $\mathcal{S}_P$. Then the conductivity parameter mapping is given by $\sigma(\mathbf{m}) = \mathbf{W m}$ or $\sigma(\mathbf{m}) = \mathbf{W exp}(\mathbf{m})$, for the cases of linear and log conductivity, respectively.

Eq. (22) can be seen to be a special case of (27) with the identifications $\mathbf{S}_0 \equiv \mathbf{C}^\dagger\mathbf{C}$, $\mathbf{U} \equiv i\omega\mu\mathbf{I}$, $\mathbf{V} \equiv \mathbf{I}$, and $\pi(\mathbf{m}) \equiv \sigma(\mathbf{m})$, and we have

$$\mathbf{P} = diag(-i\omega\mu\mathbf{e}_0)\,\Pi_{\mathbf{m}_0} \tag{38}$$

$$\mathbf{P}^T = \Pi_{\mathbf{m}_0}^T\, diag(i\omega\mu\mathbf{e}_0), \tag{39}$$

where $\Pi_{\mathbf{m}_0} = \mathbf{W}$ for linear conductivity, and $\Pi_{\mathbf{m}_0} = \mathbf{W}[\,diag(\mathbf{exp}(\mathbf{m}_0))\,]$ for logarithmic conductivity. Note that the transposes of the averaging operators $\mathbf{W}$ and $\mathbf{W}_{\rm TM}$ represent mappings from cell edges to cells, a weighted sum of contributions from all edges that bound a cell.

5.2 Matrices L and Q

We turn now to the matrices $\mathbf{L}$ and $\mathbf{Q}$, which represent the linearized observation process, as it is applied to the discrete numerical forward problem.

5.2.1 $\mathbf{L}$*: general case*

The very simplest sort of EM data is an observation of the primary field at a single location (e.g., $\epsilon = E_y(\mathbf{x})$), which can be represented as a local average of the modeled primary field

$$\epsilon = (\lambda^P)^T\mathbf{e}. \tag{40}$$

Here $\lambda^P \in \mathcal{S}_P$ is a sparse vector of interpolation coefficients, averaging from the discrete primary grid to the observation point $\mathbf{x}$. A point observation of the dual field (e.g., $\eta = B_x(\mathbf{x})$) is only slightly more complicated. Assuming, as will generally be the case, that the dual fields can be written as $\mathbf{h} = \mathbf{Te}$, where $\mathbf{T} : \mathcal{S}_P \mapsto \mathcal{S}_D$ is a discrete differential operator (e.g., see (21)), we have

$$\eta = (\lambda^D)^T\mathbf{Te}, \tag{41}$$

where $\lambda^D \in \mathcal{S}_D$ is again a sparse vector of interpolation coefficients, now representing averaging on the dual grid. For some problems $\mathbf{T} \equiv \mathbf{T}_{\pi(\mathbf{m})}$ will depend on the model parameter through $\pi(\mathbf{m})$ (see Section 5.2.4 for example). It is also possible for the interpolation coefficients λ^P and/or λ^D to depend on the model parameter $\mathbf{m}$. We will return to these complications, which are accounted for in the operator $\mathbf{Q}$, below.

Together (40) and (41) give the basic evaluation functionals for the fundamental observables (point measurements of magnetic and electric fields) in any EM problem. For controlled source problems, where the data are typically just point measurements of the primary or dual field, these evaluation functionals are already the rows of $\mathbf{L}$. More generally, EM data are functions of both electric and magnetic components, at one or more locations. The most obvious example is the impedance, the local ratio of electric and magnetic fields. Other examples include inter-station magnetic transfer functions, network MT accounting for the geometry of long dipoles (*Siripunvaraporn et al.* 2004), or horizontal spatial gradient methods based on array data (*Schmucker* 2003; *Semenov and Shuman* 2009). Inevitably real data must be based on a discrete set of observations of the magnetic and electric fields, so the general EM data functional can be represented as

$$\psi_j(\mathbf{e}(\mathbf{m}),\mathbf{m}) \equiv \gamma_j(\epsilon_1(\mathbf{m}),\ldots,\epsilon_{K_P}(\mathbf{m}),\eta_1(\mathbf{e}(\mathbf{m})),\ldots,\eta_{K_D}(\mathbf{e}(\mathbf{m}))), \tag{42}$$

where $\epsilon_k, k = 1,\ldots,K_P$ and $\eta_k, k = 1,\ldots,K_D$ are sets of primary and dual components computed at one or several points in the model domain, as $\epsilon_k = (\lambda_k^P)^T\mathbf{e}$ and $\eta_k = (\lambda_k^D)^T\mathbf{Te}$, respectively.

From (10), the jth row of $\mathbf{L}$ is then given by

$$\mathbf{l}_j = \left.\frac{\partial\psi_j}{\partial\mathbf{e}}\right|_{\mathbf{e}_0,\mathbf{m}_0} = \sum_{k=1}^{K_P}\left.\frac{\partial\gamma_j}{\partial\epsilon_k}\frac{\partial\epsilon_k}{\partial\mathbf{e}}\right|_{\mathbf{e}_0,\mathbf{m}_0} + \sum_{k=1}^{K_D}\left.\frac{\partial\gamma_j}{\partial\eta_k}\frac{\partial\eta_k}{\partial\mathbf{e}}\right|_{\mathbf{e}_0,\mathbf{m}_0} \tag{43}$$

$$= \sum_{k=1}^{K_P} a_{jk}^P(\lambda_k^P)^T + \sum_{k=1}^{K_D} a_{jk}^D(\lambda_k^D)^T\mathbf{T} \tag{44}$$

$$\tag{45}$$

where a_{jk}^P and a_{jk}^D are the partial derivatives of the jth data functional with respect to the kth local field components.

These coefficients depend only on the details of the data functional formulation, and the background EM solution $\mathbf{e}_0$. Eq. (43) thus implies that we can decompose $\mathbf{L}$ into two sparse matrices as

$$\mathbf{L} = \mathbf{A}^T \Lambda^T, \tag{46}$$

with

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_P \\ \mathbf{A}_D \end{bmatrix} \quad \text{and} \quad \Lambda = \begin{bmatrix} \Lambda_P & \mathbf{T}^T \Lambda_D \end{bmatrix}. \tag{47}$$

Here, $\mathbf{L}$ is a sparse $N_d \times N_e$ matrix that maps the EM solution to the data space, as in Eq. (14). $\mathbf{A}$ is a $K \times N_d$ sparse matrix ($K = K_P + K_D$), such that the non-zero elements in its jth column are the coefficients a_{jk}, the derivatives of the data functionals with respect to each of the relevant local magnetic or electric field components. Finally, Λ is an $N_e \times K$ sparse matrix, with columns $\lambda_k \in \mathcal{S}_P$ containing the field component evaluation functionals, i.e. the interpolation coefficients required to compute the kth electric/magnetic field component at a point from the primary EM field.

Thus, Λ depends only on the observation locations for each of the K local field components (and possibly on the model parameter $\mathbf{m}_0$). Observation functionals (non-linear or linearized) for any sort of EM data will be constructed from the same field component functionals, which are closely tied to the specific numerical discretization scheme used. $\mathbf{A}$, on the other hand, depends on details of the observation functionals (e.g., impedance vs. apparent resistivity), and will also depend in general on the background EM solution used for linearization, $\mathbf{e}_0$. However, $\mathbf{A}$ (which is essentially a linearization of γ) does not depend on the details of the numerical implementation of the forward problem.

5.2.2 $\mathbf{L}$*: Multivariate Transfer Functions*

Multivariate transfer functions (TFs) are an important special case of non-linear data functionals which deserve a closer look. Plane-wave source TFs provide the most important (and, in fact, only widely applied) example. In this case two independent sources are assumed, corresponding to spatially uniform sources of a fixed frequency polarized in the x and y directions. As a consequence of the linearity of the induction equations, under this assumption any point observation of the EM fields can be linearly related to two reference components, through a frequency dependent TF. Examples include the rows of the impedance tensor, such as

$$E_x = Z_{xx} B_x + Z_{xy} B_y, \tag{48}$$

vertical field TFs, and inter-site magnetic TFs. TF components such as Z_{xx} and Z_{xy}, which are estimated from time series of electric and magnetic fields observed at a single site, provide the basic input data for 3D MT inversion.

For completeness we consider the general case, where a generic predicted component, which we denote as Y, is related to N_p predicting variables $X_1, ..., X_{N_p}$ via the TF

$$Y = \theta_1 X_1 + ... + \theta_{N_p} X_{N_p}. \tag{49}$$

To evaluate the components of the complex TF vector $\mathbf{\Theta} = \left(\theta_1, ..., \theta_{N_p}\right)^T$ it is necessary to solve forward problems for each of the assumed source configurations–i.e., forward solutions $\mathbf{e}_1, ..., \mathbf{e}_{N_p}$ for N_p transmitters are required. To compute the TF we must evaluate Y and $X_j, j = 1, ..., N_p$ for each of these forward solutions. Here we represent this as

$$Y_i = \lambda_Y^T \mathbf{e}_i \qquad X_{ij} = \lambda_{X_j}^T \mathbf{e}_i \qquad i = 1, ..., N_p. \tag{50}$$

Then if $\mathbf{Y}$ denotes the vector of predicted components for the N_p transmitters, and $\mathbf{X}$ denotes the corresponding $N_p \times N_p$ matrix of predicting variables the TF can be computed as

$$\mathbf{\Theta} = \mathbf{X}^{-1} \mathbf{Y}. \tag{51}$$

Note that in general the evaluation functionals $\lambda_{X_j} \in \mathcal{S}_P$ might be more complicated than the simple interpolation operators considered previously – e.g., for the usual plane wave source case the predicting components are typically taken to be magnetic fields at the local site, which for the 3D MT example we have considered would require computation of the secondary field (multiplication by the operator $\mathbf{T}$) followed by interpolation. And for more exotic cases such as the generalized HSG TF (*Egbert* 2002; *Schmucker* 2003, 2004; *Semenov and Shuman* 2009; *Pankratov and Kuvshinov* 2010) the predicting components would involve magnetic fields measured at multiple sites, used to form some sort of estimate of uniform and gradient field components. We thus assume only that these are sparse vectors

representing linear functionals defined on $\mathcal{S}_P$.

Taking partial derivatives of $\mathbf{\Theta}$ with respect to $\mathbf{e}_i$ we find, after some simplification

$$\frac{\partial \mathbf{\Theta}}{\partial \mathbf{e}_i} = \mathbf{X}_0^{-1}\left[\frac{\partial \mathbf{Y}}{\partial \mathbf{e}_i} - \frac{\partial(\mathbf{X}\Theta_0)}{\partial \mathbf{e}_i}\right]. \tag{52}$$

In (52) the subscript zero denotes transfer functions and predicting components evaluated for the background forward solution. Note that the expression in brackets is a matrix of size $N_p \times N_e$ (N_e = dimension of $\mathbf{e}_i$), but only the i^{th} row is non-zero (only the i^{th} component of Y and row of $\mathbf{X}$ depend on solution $\mathbf{e}_i$). This row takes the form

$$\lambda_Y^T - \theta_1\lambda_{X_1}^T - ... - \theta_{N_p}\lambda_{X_{N_p}}^T, \tag{53}$$

which is independent of i.

As we noted at the end of Section 3, rows of $\mathbf{L}$ for TF components couple the terms $\mathbf{S}_\mathbf{m}^{-1}\mathbf{P}_i$ for multiple transmitters. We can now give an explicit form for this coupling, considering only a single predicted component Y, so that there are N_p complex rows of the matrix $\mathbf{L}$, one for each component of the TF. $\mathbf{L}$ can also be divided into N_p blocks of columns, one for each transmitter as in (19). From (52) and (53) $\mathbf{L}$ can be written in terms of $\mathbf{X}_0^{-1}$ and block diagonal matrices as

$$\mathbf{L} = \mathbf{X_0^{-1}}\begin{bmatrix} \mathbf{\Psi} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Psi} \end{bmatrix}\begin{bmatrix} \mathbf{\Lambda}^T & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Lambda}^T \end{bmatrix}, \tag{54}$$

where

$$\mathbf{\Psi} = \begin{bmatrix} 1 & -\mathbf{\Theta}^T \end{bmatrix} \qquad \mathbf{\Lambda} = \begin{bmatrix} \lambda_Y^T & \lambda_{X_1}^T & \dots & \lambda_{X_{N_p}}^T \end{bmatrix}. \tag{55}$$

The product of the first two matrices corresponds to $\mathbf{A}^T$ (and the rightmost of course to $\mathbf{\Lambda}$) in (46). The more explicit form here more clearly defines the coupling between transmitters, and has important implications for efficient calculation of the full Jacobian, as we discuss further in Section 6.

5.2.3 Matrix $\mathbf{Q}$

When either the evaluation functionals or the field transformation operator $\mathbf{T}$ have an explicit dependence on the model parameter (denoted in the latter case by $\mathbf{T}_{\pi(\mathbf{m})}$) there is an additional term in the sensitivity matrix, which we have denoted $\mathbf{Q}$. The jth row of this matrix is given by

$$\mathbf{q}_j = \left.\frac{\partial \psi_j}{\partial \mathbf{m}}\right|_{\mathbf{e}_0,\mathbf{m}_0} = \left[\sum_{k=1}^{K_P} a_{jk}^P \left.\frac{\partial(\lambda_k^P)^T\mathbf{e}_0}{\partial \pi}\right|_{\pi(\mathbf{m}_0)}\right.$$
$$\left. + \sum_{k=1}^{K_D} a_{jk}^D \left(\left.\frac{\partial(\lambda_k^D)^T\mathbf{T}_{\pi(\mathbf{m}_0)}\mathbf{e}_0}{\partial \pi}\right|_{\pi(\mathbf{m}_0)} + (\lambda_k^D)^T \left.\frac{\partial \mathbf{T}_{\pi(\mathbf{m})}\mathbf{e}_0}{\partial \pi}\right|_{\pi(\mathbf{m}_0)}\right)\right] \Pi_{\mathbf{m}_0}. \tag{56}$$

where $\pi(\mathbf{m})$ is the (possibly non-linear) model parameter mapping to the dual or primary grid, and $\Pi_{\mathbf{m}_0}$ is the Jacobian of this mapping. Defining

$$\tilde{\mathbf{T}}_{\pi(\mathbf{m}_0),\mathbf{e}_0} = \left.\frac{\partial}{\partial \pi}\left[\mathbf{T}_{\pi(\mathbf{m})}\mathbf{e}_0\right]\right|_{\pi(\mathbf{m}_0)} \tag{57}$$

$$\tilde{\Lambda}_P^T = \left.\frac{\partial}{\partial \pi}\left[(\Lambda_P)^T\mathbf{e}_0\right]\right|_{\pi(\mathbf{m}_0)} \tag{58}$$

$$\text{and} \quad \tilde{\Lambda}_D^T = \left.\frac{\partial}{\partial \pi}\left[(\Lambda_D)^T\mathbf{T}_{\pi(\mathbf{m}_0)}\mathbf{e}_0\right]\right|_{\pi(\mathbf{m}_0)} \tag{59}$$

Eq. (56) can be given in matrix notation

$$\mathbf{Q} = \begin{bmatrix} \mathbf{A}_P^T & \mathbf{A}_D^T \end{bmatrix}\begin{bmatrix} \tilde{\Lambda}_P^T \\ \tilde{\Lambda}_D^T + \Lambda_D^T\tilde{\mathbf{T}} \end{bmatrix}\Pi_{\mathbf{m}_0}. \tag{60}$$

If the interpolation coefficients are independent of the model parameters (as will be most often the case) this reduces to

$$\mathbf{Q} = \mathbf{A}_D^T \Lambda_D^T \tilde{\mathbf{T}}_{\pi(\mathbf{m}_0),\mathbf{e}_0} \Pi_{\mathbf{m}_0}. \tag{61}$$

5.2.4 Example: 2D MT

For 2D MT the fundamental observation is an impedance, the ratio E/B of orthogonal components of the electric and magnetic fields. For the TE mode E_x corresponds to the primary (modeled) field $\mathbf{e}$, while B_y is the secondary field, which is computed as $\mathbf{h} = \mathbf{T}_E\mathbf{e}; \mathbf{T}_E = (-i\omega)^{-1}\mathbf{OG}$. $\mathbf{O}$ is a diagonal matrix with entries $+1$ and -1 for components corresponding to y and z edges respectively. Columns of Λ_P and Λ_D now represent bi-linear spline interpolation from the 2D grid nodes and edges, respectively, to the data sites. These are independent of the model parameter, so $\mathbf{Q} \equiv 0$.

The impedance can be written explicitly as

$$Z \equiv \gamma_j(\mathbf{e}) = \frac{\lambda_E^T \mathbf{e}}{\lambda_B^T \mathbf{T}_E \mathbf{e}},$$

where $\mathbf{e}$ is the (primary) electric field, and λ_E and λ_B are, respectively, columns of Λ_P and Λ_D, and represent bi-linear spline interpolation functionals on node (primary) and edge (dual) spaces. From (43), the row of $\mathbf{L}$ corresponding to an impedance is found to be

$$\mathbf{l}_j \equiv \mathbf{l}_Z = (\lambda_B^T \mathbf{T}_E \mathbf{e}_0)^{-1} \lambda_E^T - [\lambda_E^T \mathbf{e}_0 / (\lambda_B^T \mathbf{T}_E \mathbf{e}_0)^2] \lambda_B^T \mathbf{T}_E. \tag{62}$$

For the TM mode, Λ_P and Λ_D are the same as in the TE case, but the roles of primary and dual fields are reversed, so that

$$Z \equiv \gamma_j(\mathbf{e}) = \frac{\lambda_E^T \mathbf{T}_B \mathbf{e}}{\lambda_B^T \mathbf{e}},$$

$\mathbf{e}$ now denoting the (primary) magnetic field. Also the field transformation operator is now $\mathbf{T}_B = diag[\rho(\mathbf{m})]\mathbf{OG}$, and thus depends on the model parameter, so $\mathbf{Q}$ will be non-zero. Row j of $\mathbf{L}$ is now

$$\mathbf{l}_j \equiv \mathbf{l}_Z = -[\lambda_E^T \mathbf{T}_B \mathbf{e}_0 / (\lambda_B^T \mathbf{e}_0)^2] \lambda_B^T + (\lambda_B^T \mathbf{e}_0)^{-1} \lambda_E^T \mathbf{T}_B, \tag{63}$$

while the corresponding row of $\mathbf{Q}$ is found to be

$$\mathbf{q}_j \equiv \mathbf{q}_Z = (\lambda_B^T \mathbf{e}_0)^{-1} \lambda_E^T \, diag[\mathbf{OG}\mathbf{e}_0] \Pi_{\mathbf{m}_0}. \tag{64}$$

Note that the expressions for the scalar impedance for 2D MT can also be derived as a special (degenerate) case of the multivariate transfer functions considered above.

Linearized data functionals for apparent resistivity and phase are discussed in Appendix A.

5.2.5 Example: 3D MT

For the 3D magnetotelluric problem formulated in terms of the electric fields (Section 5.1.2), the discrete operator $\mathbf{T} = (i\omega)^{-1}\mathbf{C}$ maps from edges to faces, computing magnetic fields through application of the discrete curl operator. Interpolation from edges and faces to an arbitrary location within the 3D staggered grid model domain can be based on something simple such as tri-linear splines. In this case both Λ and $\mathbf{T}$ are independent of $\mathbf{m}$, and so $\mathbf{Q} \equiv 0$.

$\mathbf{L}$ can be readily derived as a special case of the multivariate TF with $N_p = 2$. Each row of the 2×2 impedance tensor is a separate TF–i.e., Y in the general development of Section 5.2.2 corresponds to E_x for the first row, and E_y for the second. The predictor variables X_1, X_2 correspond to the local horizontal magnetic field. Thus, $\lambda_{X_i}^T = \lambda_{Bi}^T \mathbf{T}, i = 1, 2$ are functionals for computing the two magnetic field components, and $\lambda_Y^T = \lambda_{Ek}$ for rows $k = 1, 2$ of the impedance. The 2×2 matrix $\mathbf{X}$ thus has elements $X_{ij} = \lambda_{Bi}^T \mathbf{T}\mathbf{e}_j$ (same for both rows of the impedance). From (54) and (55) the row of the (complex) $\mathbf{L}$ corresponding to impedance element ki is

$$\left[\begin{array}{cc} \mathbf{X}_{i1}^{-1} \left(\lambda_{Ek}^T - Z_{k1} \lambda_{B1}^T \mathbf{T} - Z_{k2} \lambda_{B2}^T \mathbf{T} \right) & \mathbf{X}_{i2}^{-1} \left(\lambda_{Ek}^T - Z_{k1} \lambda_{B1}^T \mathbf{T} - Z_{k2} \lambda_{B2}^T \mathbf{T} \right), \end{array} \right] \tag{65}$$

where the components of $\mathbf{X}$, and the impedance $\mathbf{Z}$ are calculated from the background solution. Note that this row of $\mathbf{L}$ has two blocks, (each of length N_e) which multiply perturbations to the two polarizations $\delta\mathbf{e}_j, j = 1, 2$, and are summed to compute the total perturbation δZ_{ki} to the impedance element. Rows of $\mathbf{L}$ for vertical field TFs, which

relate B_z to the two local horizontal components of the magnetic field, have the same form, with λ_{Bz} replacing λ_{Ek} and the two components of the vertical field TF replacing $Z_{kj}, j = 1, 2$.

6 MULTIPLE TRANSMITTERS

We now give a more explicit discussion of how all of the pieces of $\mathbf{J}$ fit together in the case of multiple transmitters, allowing for the sort of coupling that occurs with multivariate TFs. In general there will be N_T transmitters, corresponding to different source geometries and/or different frequencies. There will also be a total of N_R measured components of the EM field at some location. Note that these would correspond to the actual field observations. Some or all of the observations actually used for the inversion would be constructed from these, e.g., through TFs, with possible further transformation to apparent resistivity and phase. In general, subsets of receiver locations may be used for each transmitter. The full Jacobian for all data can then be written

$$\mathbf{J} = \mathbf{A}^T \begin{bmatrix} \Lambda^T & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Lambda^T \end{bmatrix} \begin{bmatrix} \mathbf{S}_1^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_{N_T}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{P}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_{N_T} \end{bmatrix} + \begin{bmatrix} \mathbf{Q}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q}_{N_T} \end{bmatrix} . \tag{66}$$

where $\mathbf{\Lambda}$ is the measurement operator, evaluating solutions for each transmitter for all of the N_R receivers. The matrices $\mathbf{P}_t$ and $\mathbf{Q}_t$ are generally different for each transmitter t, as they depend on the forward solution $\mathbf{e}_t$, computed for the reference model parameter used for the Jacobian calculation (see (12)). Two transmitters, indexed by t_1, t_2, which only differ in the geometry of the source will typically have identical forward operators, i.e., $\mathbf{S}_{t_1} = \mathbf{S}_{t_2}$ (though this is not true for the 2D MT case, where the two source polarizations are decoupled, and different forward problems are solved for TE and TM modes). The solution operators will always be different for transmitters corresponding to different frequencies. Complications such as the possibility that not all receiver/transmitter pairs are observed, coupling between transmitters through TFs, and further non-linear transformations of data are embedded in the matrix $\mathbf{A}$. This matrix will generally be very sparse, with diagonal blocks coupling at most a few transmitters.

Perhaps the simplest specific example of (66) is the controlled source cross-well imaging problem (e.g., *Alumbaugh and Newman* 1997). In this case transmitters are point magnetic dipoles in one well, and observations are point measurements of the magnetic field in another well. Assuming all transmitter-receiver pairs are observed, the total number of data is $N_d = N_T N_R$, and we may take $\mathbf{A} = \mathbf{I}$. Assuming further that all data are taken at a single frequency, the forward operators are all identical, $\mathbf{S}_t \equiv \mathbf{S}$. Then (assume $\mathbf{Q} \equiv 0$) the transpose of the full Jacobian can be computed as

$$\mathbf{J}^T = \begin{bmatrix} \mathbf{P}_1(\mathbf{S}^T)^{-1}\mathbf{\Lambda} & \mathbf{P}_2(\mathbf{S}^T)^{-1}\mathbf{\Lambda} & \dots \mathbf{P}_{N_T}(\mathbf{S}^T)^{-1}\mathbf{\Lambda} \end{bmatrix} . \tag{67}$$

This requires N_T forward solutions, and N_R adjoint solutions (one for each column λ_r of $\mathbf{\Lambda}$) to evaluate all $N_T N_R$ rows of $\mathbf{J}^T$. Storing the full Jacobian (of size $N_T N_R \times M$) might be prohibitively expensive in terms of memory. However, the actual number of forward and adjoint solutions required for calculation of the full Jacobian is in fact only N_R more than that required for a single evaluation of the gradient, used for a single step in an NLCG or quasi-Newton scheme. Considering that each line search in NLCG would then require N_T forward solutions to evaluate the penalty functional, one might question whether these direct optimization schemes are more practical and efficient than a Gauss-Newton scheme for a problem such as this, certainly if N_R and N_T were of comparable magnitude. Extensions of this simple case to allow for multiple frequencies, or more complex sampling patterns with only some transmitter-receiver pairs, would be straightforward.

In the simple cross-well example the Jacobian "factors" into components dependent on the transmitter and receiver with the sensitivity for data d_{tr} (where t and r are, respectively, the transmitter and receiver indices) is $\mathbf{P}_t(\mathbf{S}^T)^{-1}\lambda_r$. A similar factorization will apply to any problem where there are transmitters with a single frequency (more precisely, with identical forward solvers), but multiple source geometries. Many active source problems, in particular marine CSEM, would fall into this category.

This source-receiver factorization also applies to the case of multivariate TFs, and more complicated data derived from them. Consider the N_p rows of $\mathbf{J}$ associated with the components of a single TF $\mathbf{\Theta}$. These rows of $\mathbf{J}$ can be

represented in the general form (66), with a single forward operator $\mathbf{S}_t \equiv \mathbf{S}$. From (54) we thus have

$$\mathbf{J}_\Theta = \mathbf{X}^{-1} \begin{bmatrix} \mathbf{\Psi\Lambda}^T & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Psi\Lambda}^T \end{bmatrix} \begin{bmatrix} (\mathbf{S}^T)^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (\mathbf{S}^T)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{P}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_{N_p} \end{bmatrix} \tag{68}$$

or, for the transpose

$$\mathbf{J}_\Theta^T = \begin{bmatrix} \mathbf{P}_1^T (\mathbf{S}^T)^{-1} \mathbf{\Lambda\Psi}^T & \dots & \mathbf{P}_{N_p}^T (\mathbf{S}^T)^{-1} \mathbf{\Lambda\Psi}^T \end{bmatrix} \left(\mathbf{X}^{-1}\right)^T . \tag{69}$$

Thus, all N_p rows of $\mathbf{J}$ require only a single adjoint solution, which must then be multiplied by each of the matrices $\mathbf{P}_t, t = 1, \dots, N_p$. The resulting model space vectors are then coupled, through the $N_p \times N_p$ matrix $(\mathbf{X}^{-1})^T$ to form the N_p rows of $\mathbf{J}_\Theta$. For multivariate transfer function problems there will generally be several predicted components at a single site, each associated with an N_p component TF. Each of these TFs will require a separate adjoint solution, $(\mathbf{S}^T)^{-1}\mathbf{\Lambda_j \Psi_j}^T$ since Λ and Ψ will be different for each TF, but all share the same transmitter dependent matrices $\mathbf{P}_t$, and the same coupling matrix.

In the context of the 3D MT problem, one has two TFs, corresponding to the two rows of the impedance tensor, and hence two adjoint solutions are required to compute sensitivities for the full impedance tensor. If vertical field TFs are also included, there would be a third TF, and a third adjoint solution would be required. A direct calculation of sensitivities through the transposed equation (15), taking account of (19) but ignoring the factorization of $\mathbf{L}$ used to derive (68) and (69), would suggest that two adjoint solutions are required for each of the four components of the impedance tensor. This would imply a total of eight adjoint solutions to evaluate the full sensitivity for an impedance tensor at one location/frequency. Thus, the more careful analysis given here suggests substantial efficiencies, reducing the total number of adjoint solutions for a full calculation of the Jacobian by a factor of 4.

Sensitivities for any data derived from impedance tensor components can obviously be constructed from the adjoint solutions $(\mathbf{S}^T)^{-1}\mathbf{\Lambda_j \Psi_j}^T, j = 1, 2$ essentially as in (69) but with a modified coupling matrix, analogous to $(\mathbf{X}^{-1})^T$. An example would be the 4 components of the phase tensor (e.g., *Caldwell et al.* 2004), which is a non-linear function of the full impedance.

7 CONCLUSIONS

We have developed general recipes for the Jacobian calculations that are central to a wide range of EM inversion algorithms. Our analysis is based on the discrete formulation of the forward problem, including explicit (though often fairly abstract) treatment of parameter mappings and data functionals in the numerical implementation. Through this analysis, we show how the Jacobian can be decomposed into simpler operators, and we analyze the dependence of these operators on specific aspects of the EM problem, such as the configuration and type of transmitters and receivers, or on implementation specific details, such as the model parametrization or the nature of the numerical discretization. We have used the general formulation to provide fairly explicit expressions for Jacobian calculations for several example problems, including 2D and 3D MT, and 3D controlled source problems with multiple transmitters and simple data functionals. To the extent that we have discussed numerical and discretization details of the forward solver, we have focused on finite difference methods. However, much of our theory is more generally applicable – e.g., the division of the Jacobian into components, and the dependencies of these components on details of the EM problem and model parametrization will hold in general. The framework we have developed should thus be useful to guide development of inversion algorithms for a wide class of EM problems, with a broad range of numerical implementations of the forward problem.

The theory developed here provides the basis for our development of a modular system of computer codes for EM inversion, which we discuss in a companion paper (*Egbert et al.* 2010). With a modular approach inversion codes developed for one purpose can be rapidly adapted to other EM problems, and development of new capabilities can be considerably simplified. For example, we have discussed the 3D MT inversion in detail, but have not specifically considered inter-site magnetic transfer functions. Adding the capability to invert these would require minor modifications to the data functionals (essentially modifying matrix $\mathbf{A}$ in (46)). More generally, with a properly designed system, any 3D problem can be developed with fairly minor modifications to the 3D MT system, essentially by modifying

the sources for the forward problem, and the measurement functionals (assuming the forward code is already accurate enough – for controlled source a secondary field formulation will be useful). The modular system also provides great flexibility with regard to changing model parametrizations. Our analysis here shows explicitly how the model parametrization interacts with the components of the Jacobian – through the mapping $\pi(\mathbf{m})$, and the Jacobian of this mapping ($\Pi_{\mathbf{m}_0}$) – and guides our development of a system which supports interchangeable model parametrizations. Another key result of our analysis is the "factorization" of the Jacobian into components dependent only on transmitters, and on receivers. This would appear to have important implications for efficient implementation of inversion algorithms, and also for comparing efficiency of different schemes. These issues will be explored more thoroughly in the companion paper.

This last point provides a good illustration of one of the challenges in developing a flexible modular system. On the one hand the greatest degree of generality and code reuse is attained by hiding instance specific details; on the other hand these details can sometimes be exploited to develop more efficient implementations for specific cases. For example, one approach to the multivariate TF problem with coupled transmitters would be to consider the fundamental EM solution object as $\mathbf{e} \equiv (\mathbf{e}_1, ..., \mathbf{e}_{N_p})$, and take $\mathbf{P}$ and $\mathbf{S}$ to represent the full block diagonal operators of (68). Thus, (14) will be valid for multiple polarization problems also, and indeed for almost anything one could imagine, provided we interpret the component operators with enough generality. In the context of a modular system, if basic objects such as EM solutions and data functionals are constructed so that details such as the multiplicity of source polarizations are hidden, simple, but still very general routines can be written to implement Jacobian and gradient calculations for all cases. However, as our analysis of the structure of sensitivities for multivariate TFs illustrates, hiding instance specific details can also obscure opportunities for computational efficiency. In general there will always be some tension between the abstraction required for a flexible general purpose inversion system, and the optimally efficient schemes which might be developed for a specific case.

APPENDIX A: TRANSFORMATION OF JACOBIAN TO REAL FORM

To allow for the fact that the model parameter $\mathbf{m}$ is typically real, and in some cases data are also real, we have assumed that $\mathbf{d}$ and $\mathbf{J}$ are real, with any complex observations (e.g., an impedance) represented as two real elements of the data vector. However throughout the text we have used complex notation for $\mathbf{L}, \mathbf{S}_{\mathbf{m_0}}^{-\mathbf{1}}, \mathbf{P}$ and $\mathbf{Q}$, so $\mathbf{J}$ computed from (14) would also be complex. In fact, for complex observations it is readily verified that the real and imaginary parts of a row of the complex expression for the Jacobian give the sensitivity (a vector in the real model parameter space) for the corresponding real and imaginary parts of one observation. Thus, to keep the Jacobian and the data vector strictly real, we can set

$$\bar{\mathbf{d}} = \begin{bmatrix} \Re(\mathbf{d}) \\ \Im(\mathbf{d}) \end{bmatrix} \qquad \bar{\mathbf{J}} = \begin{bmatrix} \Re(\mathbf{J}) \\ \Im(\mathbf{J}) \end{bmatrix} = \Re\left[\begin{bmatrix} \mathbf{L} \\ -i\mathbf{L} \end{bmatrix} \mathbf{S}^{-1}\mathbf{P} + \begin{bmatrix} \mathbf{Q} \\ -i\mathbf{Q} \end{bmatrix} \right] \tag{A1}$$

with the convention that for any observations that are intrinsically real the rows corresponding to the imaginary component are omitted. From (A1), $\bar{\mathbf{J}}^T\bar{\mathbf{d}} = \Re(\mathbf{J}^T)\Re(\mathbf{d}) + \Im(\mathbf{J}^T)\Im(\mathbf{d})$. It is easily seen that

$$\bar{\mathbf{J}}^T\bar{\mathbf{d}} = \Re\left[\mathbf{J}^T\mathbf{d}^*\right] = \Re\left[\mathbf{P}^T{\mathbf{S}^T}^{-1}\mathbf{L}^T\mathbf{d}^* + \mathbf{Q}\mathbf{d}^*\right], \tag{A2}$$

where the superscript asterisk denotes the complex conjugate. Thus, the complex component matrices can be used to construct the real Jacobian $\bar{\mathbf{J}}$, and to implement multiplication by this matrix and its transpose. Note also that while we assume the data vector is real, real and imaginary parts of sensitivities for a complex observation are computed (e.g., via (15)) with a single adjoint solution.

Apparent resistivity and phase provide examples of observations that are intrinsically real. In terms of the impedance the apparent resistivity is defined as

$$\rho_a = T\left|Z\right|^2/5 = T\left[Z_r^2 + Z_i^2\right]/5 \tag{A3}$$

where Z_r and Z_i are real and imaginary parts of the impedance Z. Applying the chain rule

$$\frac{\partial \rho_a}{\partial \mathbf{m}} = \frac{\partial \rho_a}{\partial Z_r}\frac{\partial Z_r}{\partial \mathbf{m}} + \frac{\partial \rho_a}{\partial Z_i}\frac{\partial Z_i}{\partial \mathbf{m}} = \frac{2T}{5}\left[Z_r\frac{\partial Z_r}{\partial \mathbf{m}} + Z_i\frac{\partial Z_i}{\partial \mathbf{m}}\right] \tag{A4}$$

$$= \frac{2T}{5}\left[Z_r\Re\frac{\partial Z}{\partial \mathbf{m}} + Z_i\Im\frac{\partial Z}{\partial \mathbf{m}}\right] = \Re\left[\frac{2TZ^*}{5}\frac{\partial Z}{\partial \mathbf{m}}\right] = \Re\left[\frac{2TZ^*\mathbf{l}_Z^T}{5}\frac{\partial \mathbf{e}}{\partial \mathbf{m}}\right]. \tag{A5}$$

Thus, $\mathbf{l}_\rho = 2TZ^*\mathbf{l}_Z^T/5$ gives the (complex) row of $\mathbf{L}$ for an apparent resistivity, again with the convention that the real part of the product in (14) is taken for the corresponding row of the real Jacobian. Similarly for the phase $\phi = \tan^{-1}\left(Z_r/Z_i\right)$, we find that the row of $\mathbf{L}$ takes the form $\mathbf{l}_\phi = iZ^*\mathbf{l}_Z^T/\left|Z\right|^2$.

APPENDIX B: 3D STAGGERED GRID DETAILS

Here we give a more precise definition of the discrete finite difference operator corresponding to $\nabla\times\nabla\times + i\omega\mu\sigma$ and its adjoint, and clarify implementation of boundary conditions for the 3D MT problem. Similar considerations apply to other cases considered in the text. To do this we need to distinguish more precisely between interior and boundary nodes in the grid. In the main text $\mathcal{S}_P$ ($\mathcal{S}_D$) have been used to denote the space of discrete complex vector fields defined on all edges (faces) of the staggered grid. Here we use the same symbols with tildes ($\tilde{\mathcal{S}}_P$, $\tilde{\mathcal{S}}_D$) to indicate the restriction to interior edges or faces. The discrete curl operator is naturally defined as a mapping from all edges to all faces, but we need only consider the partial mapping which computes the curl for interior faces (see e.g., *Kelbert* (2006) for details). Denote this as

$$\mathbf{C} : \mathcal{S}_P \mapsto \tilde{\mathcal{S}}_D \tag{B1}$$

and partition $\mathbf{e} \in \mathcal{S}_P$ and $\mathbf{C}$ into interior and boundary edge components

$$\mathbf{e} = \begin{bmatrix} \tilde{\mathbf{e}} \\ \mathbf{e}_b \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} \tilde{\mathbf{C}} & \mathbf{C}_b \end{bmatrix} \tag{B2}$$

so that $\tilde{\mathbf{C}} : \tilde{\mathcal{S}}_P \mapsto \tilde{\mathcal{S}}_D$ and $\mathbf{C}\mathbf{e} = \tilde{\mathbf{C}}\tilde{\mathbf{e}} + \mathbf{C}_b\mathbf{e}_b$.

To define adjoints precisely we need to specify inner products. The natural inner products for the primary and

dual spaces (interior nodes only) are

$$\langle \tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2 \rangle_P = \tilde{\mathbf{e}}_1^* \mathbf{V}_E \tilde{\mathbf{e}}_2 \qquad \left\langle \tilde{\mathbf{h}}_1, \tilde{\mathbf{h}}_2 \right\rangle_D = \tilde{\mathbf{h}}_1^* \mathbf{V}_F \tilde{\mathbf{h}}_2. \tag{B3}$$

In (B3) $\mathbf{V}_E$ and $\mathbf{V}_F$ are real diagonal matrices of edge and face volume elements. Edge volumes, for example, are defined as one fourth of the total volume of the four cells sharing the edge, so that the first discrete inner product in (B3) approximates the integral L_2 inner product for vector fields $\int\int\int \mathbf{E}_1^*(\mathbf{x}) \cdot \mathbf{E}_2(\mathbf{x}) dV$. The adjoint of the interior curl operator $\tilde{\mathbf{C}}^\dagger : \tilde{\mathcal{S}}_D \mapsto \tilde{\mathcal{S}}_P$ satisfies, by definition,

$$\left\langle \tilde{\mathbf{h}}, \tilde{\mathbf{C}}\tilde{\mathbf{e}} \right\rangle_D = \left\langle \tilde{\mathbf{C}}^\dagger \tilde{\mathbf{h}}, \tilde{\mathbf{e}} \right\rangle_P \qquad \forall \tilde{\mathbf{e}} \in \tilde{\mathcal{S}}_P, \tilde{\mathbf{h}} \in \tilde{\mathcal{S}}_D. \tag{B4}$$

Noting that that $\tilde{\mathbf{C}}$ is real, one then readily derives

$$\tilde{\mathbf{C}}^\dagger = \mathbf{V}_E^{-1} \tilde{\mathbf{C}}^T \mathbf{V}_F. \tag{B5}$$

From the definitions of $\mathbf{V}_E$ and $\mathbf{V}_F$ one can verify that $\tilde{\mathbf{C}}^\dagger$ indeed corresponds to the appropriate geometric definition of the curl operator defined on cell faces. Thus the electric field equation (22) with source $\mathbf{j}_s$

$$\nabla \times \nabla \times \mathbf{E} + i\omega\mu\sigma\mathbf{E} = \mathbf{j}_s \tag{B6}$$

can be approximated on the discrete grid as

$$\tilde{\mathbf{C}}^\dagger \mathbf{C}\mathbf{e} + i\omega\mu\sigma\tilde{\mathbf{e}} = [\tilde{\mathbf{C}}^\dagger \tilde{\mathbf{C}} + i\omega\mu\sigma]\tilde{\mathbf{e}} + \tilde{\mathbf{C}}^\dagger \mathbf{C}_b \mathbf{e}_b = \tilde{\mathbf{b}}, \tag{B7}$$

where $\tilde{\mathbf{b}} \in \tilde{\mathcal{S}}_P$ gives the discrete approximation for the source current $\mathbf{j}_s$ inside the domain; these currents (and hence $\tilde{\mathbf{b}}$) vanish for the 3D MT example we have focused on. The discrete system (B7) has one equation for each of the $\tilde{N}_\mathbf{e}$ interior edges, but $N_\mathbf{e}$ (= total number of edges) unknowns. Boundary conditions are thus required, most simply specification of tangential electric field components on the boundary edges. Then the full system of equations ($\mathbf{S}\mathbf{e} = \mathbf{b}$) can be decomposed into interior and boundary components as

$$\begin{bmatrix} \tilde{\mathbf{C}}^\dagger \tilde{\mathbf{C}} + i\omega\mu\sigma & \tilde{\mathbf{C}}^\dagger \mathbf{C}_b \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{e}} \\ \mathbf{e}_b \end{bmatrix} = \begin{bmatrix} \mathbf{S}_{ii} & \mathbf{S}_{ib} \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{e}} \\ \mathbf{e}_b \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{b}} \\ \mathbf{b}_b \end{bmatrix}, \tag{B8}$$

where $\mathbf{b}_b$ represents the specified boundary data. Eliminating the boundary edges results in a well-posed $\tilde{N}_\mathbf{e} \times \tilde{N}_\mathbf{e}$ problem for electric fields restricted to interior edges

$$\left[\tilde{\mathbf{C}}^\dagger \tilde{\mathbf{C}} + i\omega\mu\sigma\mathbf{I}\right] \tilde{\mathbf{e}} = \mathbf{S}_{ii}\tilde{\mathbf{e}} = \tilde{\mathbf{b}} - \tilde{\mathbf{C}}^\dagger \mathbf{C}_b \mathbf{b}_b, \tag{B9}$$

with the RHS determined from the boundary data, and any source terms in the domain. Using (B5) we see that the discrete operator in (B9) can be written as $\mathbf{S}_{ii} = \mathbf{V}_E^{-1} \tilde{\mathbf{C}}^T \mathbf{V}_F \tilde{\mathbf{C}} + i\omega\mu\sigma\mathbf{I}$. Thus, as sketched in Section 3, the system $\mathbf{S}\mathbf{e} = \mathbf{b}$ can be reduced to symmetric form by eliminating the boundary nodes, and then multiplying both sides of the resulting equation (B9) by $\mathbf{V}_E$.

We emphasize that in our treatment of the discrete forward problem we take $\mathbf{e}$, $\mathbf{S}$, and $\mathbf{b}$ to include both interior and boundary nodes. Thus to be precise in our application of (27) to the 3D finite difference equations considered here, we should take

$$\mathbf{S}_0 = \begin{bmatrix} \tilde{\mathbf{C}}^\dagger \tilde{\mathbf{C}} & \tilde{\mathbf{C}}^\dagger \mathbf{C}_b \\ 0 & \mathbf{I} \end{bmatrix}, \tag{B10}$$

and we should define $\pi(\mathbf{m}) \equiv \sigma(\mathbf{m}) \equiv 0$ on boundary edges. This is a general property of $\pi(\mathbf{m})$, since the boundary conditions do not depend on the model parameter. This implies that the columns of $\mathbf{P}$ corresponding to boundary nodes will all vanish. Also, accounting for the boundary conditions in the transpose of $\mathbf{S}$ we have, in the notation of (B8),

$$\mathbf{S}^T\mathbf{e} = \begin{bmatrix} \mathbf{S}_{ii} & 0 \\ \mathbf{S}_{ib}^T & \mathbf{I} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{e}} \\ \mathbf{e}_b \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{b}} \\ \mathbf{b}_b \end{bmatrix}. \tag{B11}$$

The transposed solution operator $(\mathbf{S}^T)^{-1}\mathbf{b}$, which appears extensively throughout the main text, can thus be interpreted as solution of the homogeneous problem (for interior nodes)

$$\mathbf{S}_{ii}^T \tilde{\mathbf{e}} = \tilde{\mathbf{b}} \tag{B12}$$

followed by computation of the boundary terms

$$\mathbf{e}_b = \mathbf{b}_b - \mathbf{S}_{ib}^T \tilde{\mathbf{e}}. \tag{B13}$$

In fact, solutions to the adjoint problem $(\mathbf{S}^T)^{-1}\mathbf{b}$ are always multiplied by $\mathbf{P}^T$, and because the rows of $\mathbf{P}^T$ corresponding to boundary components are zero, the boundary terms in (B13) are never actually required for our purposes.

APPENDIX C: DEPENDENCE OF SOURCE TERMS ON MODEL PARAMETERS

In some cases (in particular for active source problems) it is appropriate to use a so-called "secondary field" approach to solve the forward problem (e.g., *Alumbaugh et al.* 1996). In this case a background (typically 1D) conductivity is assumed, allowing quasi-analytic computation of a background solution, with the "secondary" field due to deviation from the background conductivity then computed numerically. More precisely, the total field solution is represented as $\mathbf{e} = \hat{\mathbf{e}} + \delta\mathbf{e}$, where the background field $\hat{\mathbf{e}}$ satisfies the 1D equation defined by conductivity parameter $\hat{\mathbf{m}}$. It is readily verified that the secondary field $\delta\mathbf{e}$ satisfies the induction equation with a modified source. Assuming the 3D operator can be expressed as in (27) this takes the form

$$\mathbf{S}_\mathbf{m}\delta\mathbf{e} = -\mathbf{U}\left[(\pi(\mathbf{m}) - \pi(\hat{\mathbf{m}})) \circ \mathbf{V}\hat{\mathbf{e}}\right]. \tag{C1}$$

The RHS in (C1) depends on the model parameter $\mathbf{m}$, suggesting that an additional term should be included in equation (12).

However, if we differentiate both sides of (C1) with respect to $\mathbf{m}$ and use (27) again we find

$$\frac{\partial}{\partial\mathbf{m}}\left[\mathbf{S}_0\delta\mathbf{e} + \mathbf{U}(\pi(\mathbf{m}) \circ \mathbf{V}\delta\mathbf{e})\right] = -\frac{\partial}{\partial\mathbf{m}}\left[\mathbf{U}(\pi(\mathbf{m}) \circ \mathbf{V}\hat{\mathbf{e}}\right], \tag{C2}$$

implying

$$0 = \frac{\partial}{\partial\mathbf{m}}\left[\mathbf{S}_0\delta\mathbf{e} + \mathbf{U}(\pi(\mathbf{m}) \circ \mathbf{V}\mathbf{e})\right] = \frac{\partial}{\partial\mathbf{m}}\left[\mathbf{S}_\mathbf{m}\mathbf{e}\right], \tag{C3}$$

the last equality following from the fact that $\mathbf{S}_0\hat{\mathbf{e}}$ does not depend on $\mathbf{m}$. Thus, as long as the RHS of the original problem is independent of the model parameter, $\partial\mathbf{e}/\partial\mathbf{m} = \partial[\delta\mathbf{e}]/\partial\mathbf{m}$ satisfies (12) without any additional terms, even if the equation for the secondary field does depend on $\mathbf{m}$. Note also that the correct derivative for the discrete problem solved with a secondary field approach does not itself entail this solution approach (except indirectly through the dependence of the derivative on the forward solution).

REFERENCES

Alumbaugh, D., G. A. Newman, L. Prevost, and J. N. Shadid (1996), Three-dimensional wide band electromagnetic modeling on massively parallel computers, *Radio Science*, *33*, 1–23.

Alumbaugh, D. L., and G. A. Newman (1997), Three-dimensional massively parallel electromagnetic inversion: II. Analysis of a crosswell electromagnetic experiment, *Geophysical Journal International*, *128*, 355–363, doi:10.1111/j.1365-246X.1997.tb01560.x.

Avdeev, D. (2005), Three-dimensional electromagnetic modelling and inversion from theory to application, *Surveys in Geophysics*, *26*(6), 767–799.

Avdeev, D., and A. Avdeeva (2009), 3d magnetotelluric inversion using a limited-memory quasi-newton optimization, *Geophysics*, *74*(3), F45–F57, doi:10.1190/1.3114023.

Bennett, A. F. (2002), *Inverse Modeling of the Ocean and Atmosphere*, Cambridge University Press.

Caldwell, T. G., H. M. Bibby, and C. Brown (2004), The magnetotelluric phase tensor, *Geophysical Journal International*, *158*(2), 457–469.

Chua, B. (2001), An inverse ocean modeling system, *Ocean Modelling*, *3*, 137–165, doi:10.1016/S1463-5003(01)00006-3.

Commer, M., and G. A. Newman (2008), New advances in three-dimensional controlled-source electromagnetic inversion, *Geophysical Journal International*, *172*, 513–535, doi:10.1111/j.1365-246X.2007.03663.x.

Constable, S. C., R. L. Parker, and C. G. Constable (1987), Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics*, *52*(3), 289–300, http://mahi.ucsd.edu/Steve/Occam/home.html.

de Lugao, P., O. Portniaguine, , and M. Zhdanov (1997), Fast and stable two- dimensional inversion of magnetotelluric data, *J. Geomag. Geoelec.*, *49*(11-12), 1437–1454.

Egbert, G., A. Kelbert, N. Meqbel, K. Tandon, and A. Weng (2010), Computational recipes for EM inverse problems: II. Modular Implementation, *submitted to Geophys. J. Int.*

Egbert, G. D. (1994), A new stochastic process on the sphere: application to characterization of long-period global scale external sources, in *14th workshop on electromagnetic induction in the Earth and moon*, Brest, France.

Egbert, G. D. (2002), Processing and interpretation of electromagnetic induction array data, *Surveys in Geophysics*, *23*(2-3), 207–249.

Egbert, G. D., and A. F. Bennett (1996), *Modern approaches to data assimilation in ocean modeling*, chap. Data assimilation methods for ocean tides, p. 147, Elsevier Science B.V.

Haber, E. (2005), Quasi-Newton methods for large-scale electromagnetic inverse problems, *Inverse Problems*, *21*(1), 305–323.

Kelbert, A. (2006), Geophysical inverse theory applied to reconstruction of large-scale heterogeneities in electrical conductivity of Earth's mantle, Ph.D. thesis, Cardiff University.

Kelbert, A., G. D. Egbert, and A. Schultz (2008), Non-linear conjugate gradient inversion for global em induction: resolution studies, *Geophysical Journal International*, *173*(2), 365–381, doi:10.1111/j.1365-246X.2008.03717.x.

Mackie, R. L., and T. R. Madden (1993), 3-Dimensional magnetotelluric inversion using conjugate gradients, *Geophys. J. Int.*, *115*(1), 215–229.

Mackie, R. L., J. T. Smith, and T. R. Madden (1994), 3-Dimensional electromagnetic modeling using finite-difference equations - the magnetotelluric example, *Radio Sci.*, *29*(4), 923–935.

Marquardt, D. W. (1963), An algorithm for least-squares estimation of non-linear parameters, *J. Soc. Ind. Appl. Math.*, *11*, 431–441.

Newman, G. A., and D. L. Alumbaugh (1997), Three-dimensional massively parallel electromagnetic inversion – i. theory, *Geophysical Journal International*, *128*, 345–354, doi:10.1111/j.1365-246X.1997.tb01559.x.

Newman, G. A., and D. L. Alumbaugh (2000), Three-dimensional magnetotelluric inversion using non-linear conjugate gradients, *Geophysical Journal International*, *140*, 410–424.

Newman, G. A., and P. T. Boggs (2004), Solution accelerators for large-scale three-dimensional electromagnetic inverse problems, *Inverse Problems*, *20*, 151–170, doi:10.1088/0266-5611/20/6/S10.

Nocedal, J., and S. Wright (1999), *Numerical Optimization*, New York: Springer-Verlag.

Pankratov, O., and A. Kuvshinov (2010), General formalism for the efficient calculation of derivatives of EM frequency-domain responses and derivatives of the misfit, *Geophysical Journal International*, *181*(1), 229–249.

Parker, R. L. (1994), *Geophysical Inverse Theory*, Princeton University Press, Princeton, New Jersey.

Rodi, W., and R. L. Mackie (2001), Nonlinear conjugate gradients algorithm for 2-D magnetotelluric inversion, *Geophysics*, *66*(1), 174–187.

Rodi, W. L. (1976), A technique for improving the accuracy of finite element solutions for magnetotelluric data, *Geophysical Journal International*, *44*, 483–506, doi:10.1111/j.1365-246X.1976.tb03669.x.

Sasaki, Y. (2001), Full 3-D inversion of electromagnetic data on PC, *Journal of Applied Geophysics*, *46*, 45–54, doi:10.1016/S0926-9851(00)00038-0.

Schmucker, U. (2003), Horizontal spatial gradient sounding and geomagnetic depth sounding in the period range of daily variation, in *Protokoll ber das Kolloquium elektromagnetische Tiefenforschung, Kolloquium: Konigstein*, pp. 228–237.

Schmucker, U. (2004), Multivariate magneto-variational soundings (MVS), in *Proceedings of the 17th EM Induction Workshop, Hyderabad, India*.

Semenov, V. Y., and V. N. Shuman (2009), Impedances for induction soundings of the earth's mantle, *Acta Geophysica*, *58*(4), 527–542.

Siripunvaraporn, W., and G. Egbert (2000), An efficient data-subspace inversion method for 2-D magnetotelluric data, *Geophysics*, *65*(3), 791–803.

Siripunvaraporn, W., and G. Egbert (2007), Data space conjugate gradient inversion for 2-d magnetotelluric data, *Geophysical Journal International*, *170*, 986–994, doi:10.1111/j.1365-246X.2007.03478.x.

Siripunvaraporn, W., G. Egbert, and Y. Lenbury (2002), Numerical accuracy of magnetotelluric modeling: A comparison of finite difference approximations, *Earth Planets and Space*, *54*(6), 721–725.

Siripunvaraporn, W., M. Uyeshima, and G. Egbert (2004), Three-dimensional inversion for network-magnetotelluric data, *Earth Planets and Space*, *56*(9), 893–902.

Siripunvaraporn, W., G. Egbert, Y. Lenbury, and M. Uyeshima (2005), Three-dimensional magnetotelluric inversion: data-space method, *Physics of the Earth and Planetary Interiors*, *150*(1-3), 3–14.

Smith, J. T. (1996), Conservative modeling of 3-D electromagnetic fields: 1. Properties and error analysis, *Geophysics*, *61*(5), 1308–1318.

Spitzer, K. (1998), The three-dimensional dc sensitivity for surface and subsurface sources, *Geophysical Journal International*, *134*, 736–746, doi:10.1046/j.1365-246x.1998.00592.x.

Uyeshima, M., and A. Schultz (2000), Geomagnetic induction in a heterogeneous sphere: a new three-dimensional forward solver using a conservative staggered-grid finite difference method, *Geophys. J. Int.*, *140*(3), 636–650.

Yee, K. (1966), Numerical solution of inital boundary value problems involving Maxwell's equations in isotropic media, *IEEE Transactions on Antennas and Propagation*, *14*, 302–307, doi:10.1109/TAP.1966.1138693.

Zaron, E. D., B. S. Chua, and L. Xu (2009), Recent advances in representer-based variational data assimilation, *submitted to Monthly Weather Review*.

Zhang, J., R. L. Mackie, and T. R. Madden (1995), 3-D resistivity forward modeling and inversion using conjugate gradients, *Geophysics*, *60*(5), 1313–1325.