Background for Program "EM1DFM"

Version 1.0

Developed under the consortium research project

INVERSION AND MODELLING OF APPLIED GEOPHYSICAL ELECTROMAGNETIC DATA

UBC – Geophysical Inversion Facility,
Department of Earth & Ocean Sciences,
University of British Columbia,
Vancouver, CANADA.

26 July 2000

Contents

Acknowledgments

This project was made possible by funding from the "IMAGE" consortium, of which the following companies are participants: AGIP, Anglo American, Billiton, Cominco, Falconbridge, INCO, MIM, Muskox Minerals, Newmont, Placer Dome and Rio Tinto, and from the Natural Sciences and Engineering Research Council of Canada.

Contact Information

Author: Colin Farquharson,
Outreach Coordinator: Francis Jones,

UBC – Geophysical Inversion Facility,
Department of Earth & Ocean Sciences,
University of British Columbia,
Geophysics & Astronomy Building,
2219 Main Mall,
Vancouver, B.C., V6T 1Z4,
CANADA.

`http://www.eos.ubc.ca/ubcgif`

`ubcgif@eos.ubc.ca`

Abstract

The name: EM1DFM – electromagnetics ("EM"), one-dimensional models ("1D"), frequency-domain observations ("F"), and magnetic (dipole) sources and receivers ("M").

The observations are the inphase and/or quadrature components of:
- the secondary H-field normalized by the primary (i.e., free-space) field in ppm, or
- the secondary H-field normalized by the primary field in %, or
- the secondary H-field in A/m, or
- the total H-field in A/m.

Receiver coils can be oriented in the x-, y- or z-directions, and they can be at any position relative to their respective magnetic dipole transmitter. Transmitters can be at any height, and oriented in the x-, y- or z-directions. Observations can be for any frequency or set of frequencies. All the observations (in any combination) to be used to construct the one-dimensional model at a particular horizontal location are grouped together as one "sounding". Measurement uncertainties can be in the same units as the observations or as relative uncertainties in percent.

The product:
- an electrical conductivity model (with magnetic susceptibility fixed), or
- a strictly-positive magnetic susceptibility model (with conductivity fixed), or
- both conductivity and strictly-positive susceptibility models, or
- both conductivity and susceptibility models (with no positivity constraint).

The Earth models are composed of layers of uniform conductivity/susceptibility with fixed interface depths. The value of the conductivity/susceptibility in each layer is sought by the inversion. Multiple soundings can be handled in a single run of the program. Each sounding is inverted independently for a one-dimensional model under the sounding location, and a composite two-dimensional image of the subsurface beneath the survey area written out.

Four possible methods for determining the degree of regularization:
- constant (user-supplied) trade-off parameter in the objective function being minimized, or
- the trade-off parameter is automatically chosen to achieve a user-supplied target misfit, or
- the trade-off parameter is automatically chosen using the GCV criterion, or
- the trade-off parameter is automatically chosen using the L-curve criterion.

Full flexibility of the sum-of-squares measure of model structure is provided. The balance between the conductivity and susceptibility components is adjustable (if both are active in the inversion), the balance between the "smallest" and "flattest" parts of both conductivity and susceptibility components is adjustable, reference models can be included in either or both the "smallest" and "flattest" parts, and additional user-supplied weighting of the layers can be incorporated.

Documentation is provided in three parts:
- the description herein of the forward modelling and inversion algorithms used by program EM1DFM,
- the HTML document "Manual for Program EM1DFM" containing detailed descriptions of the inputs to and outputs from program EM1DFM (and descriptions of the related independent forward-modelling program and utility programs), and
- a frames-based HTML document describing the graphical user interface.

1. Program EM1DFM: An Overview

This first of two main sections in this document provides an overview of the forward modelling and inversion algorithms that are implemented in program EM1DFM. Detailed descriptions of the mathematics of these algorithms can be found in the second section.

Electromagnetic loop-loop surveys

The type of survey data which program EM1DFM is designed to interpret is frequency-domain, small loop, electromagnetic data acquired to determine the spatial variation of the electrical conductivity and/or magnetic susceptibility of the subsurface. This variation can be of interest for many reasons. For example, it might indicate the location and size of an ore deposit in its surrounding host rocks, or it might reveal the thickness of overburden that has to be removed during a mining operation, or it could give the depth to an aquifer and its extent. The acquired data are measurements of the magnetic field associated with electric currents and magnetization induced in the subsurface by a time-varying current in a transmitter loop. Information about how the conductivity and susceptibility vary with depth is obtained by making measurements for different frequencies of the transmitter current, and for different separations, heights and orientations of the transmitters and receivers. Making such measurements at different locations gives information about the lateral variation in the subsurface. For a review of electromagnetic loop-loop methods in geophysics see Spies & Frischknecht (1987).

One-dimensional models of the Earth

Program EM1DFM is a one-dimensional inversion program: the mathematical representation that it uses to model the Earth varies only with depth. In particular, the representation comprises many uniform, horizontal layers, and it is the conductivity and/or susceptibility within each layer that is computed. Four options are available: just the conductivity in each layer can be constructed, or just the susceptibility (constrained to be positive) in each layer, or both the conductivity and susceptibility with the positivity constraint on the susceptibility, or both the conductivity and susceptibility without the positivity constraint. When measurements from different locations are being interpreted, the one-dimensional models from all the locations are juxtaposed to create a two-dimensional image of the subsurface.

The forward-modelling and sensitivity-computing procedures

For a given layered model, the values of the magnetic field that would be observed for particular arrangements of sources and receivers, and for particular frequencies, are computed using a matrix propagation, spectral expansion approach. This is described in detail in Section 2.3 of this document, and in Farquharson et al. (2000).

Because of the inductive nature of a loop source, the magnetic field can be derived from a scalar potential – the vertical component of the Schelkunoff F-potential. And because of the lateral invariance of the model, the potential can be represented as a spectral expansion in terms of horizontal wavenumbers, thus reducing the problem to that of solving an ordinary differential equation rather than a partial differential equation. The solution for the potential in each uniform layer in the wavenumber domain is a combination of an upward-decaying and a downward-decaying exponential. The continuity of the tangential E- and H-fields across the interfaces between the layers relates the solutions in neighbouring layers to one another. This relationship is propagated through the complete stack of layers as a product of matrices resulting in a pair of simultaneous equations whose solution determines the spectral coefficients of the potential in the air. Because of the axial symmetry of the mathematics for the potential, the transformation back to the spatial domain can be expressed as a Hankel transform, and is computed using the digital filtering routine of Anderson (1982). The derivatives required to calculate the three components of the magnetic field from the potential are absorbed within the Hankel transforms, and thus numerical differentiation is avoided.

The inverse problem of determining conductivity and/or susceptibility from electromagnetic measurements is nonlinear. It is solved by program EM1DFM using an iterative procedure at each iteration of which a linearized approximation of the full nonlinear problem is solved. This process requires knowledge of the

Jacobian matrix of sensitivities, or Fréchet derivatives. These are computed within program EM1DFM using essentially a sensitivity-equation approach. (See Section 2.4 of this document, and Farquharson et al., 2000, for a full description of the mathematics of the sensitivity computations.) Expressions for the sensitivities can be obtained by differentiating directly the expressions for the magnetic fields. The necessary computations for the sensitivities are then equivalent to performing the same number of forward modellings as there are layers. These forward modellings require the same individual layer matrices, and the same partial matrix products, as the forward modelling to compute the magnetic fields. These matrices are saved within program EM1DFM during a forward modelling for a particular model, and are then re-used in any subsequent sensitivity computations. This is therefore a rapid means of generating the Jacobian matrix. Because of its derivation from the direct differentiation of the forward-modelling procedure, it is also stable.

The inversion algorithms

There are four different variations of the inversion algorithm within program EM1DFM. Detailed descriptions of each one are given in Section 2.5, and in Farquharson & Oldenburg (2000). Please refer also to the review of linear inverse theory at http://www.geop.ubc.ca/ubcgif/tutorials/lininv/index.htm. This site contains the majority of the material relevant to the methods used in program EM1DFM, and there is an interactive JAVA applet that is designed to aid understanding of many of the features which have been incorporated into program EM1DFM.

The goal of all four variants of the inversion algorithm is to construct the simplest Earth model that fits the observations to an acceptable degree. To do this, the inverse problem is formulated as an optimization problem in which the model that minimizes the objective function:

$$\Phi \;=\; \phi_d \;+\; \beta\,\phi_m \;-\; \gamma\,\phi_{LB}$$

is sought. The first term, ϕ_d, is the traditional sum-of-squares measure of misfit between the observations and the data computed for a particular Earth model. The second term involves ϕ_m, a measure of the simplicity of a model. This measure can be tailored to accomplish several different effects. It can be chosen to be a measure of the amount of structure in a model, and thus cause the inversion process to construct models containing enough features to adequately reproduce the observations, but no more. Or ϕ_m can be chosen as a measure of how different a model is from a preconceived image of the subsurface, should one be available. This form of ϕ_m will cause the inversion process to construct a model as similar as possible to the reference model while still fitting the observations. The coefficient, β, of the second term in the objective function is the trade-off parameter. This balances the usually opposing influences of minimizing the misfit and simplifying the model. This trade-off parameter can be determined in various ways, and this leads to four variants of the inversion algorithm. Finally, the third term, $-\gamma\,\phi_{LB}$, in the objective function is a logarithmic barrier term that is used to enforce positivity of the susceptibility, if this is required.

The relationship between the measured magnetic fields and the conductivities and susceptibilities of the subsurface is nonlinear. The inverse problem of determining these properties of the subsurface from magnetic field values is therefore nonlinear. As mentioned above in conjunction with the need for the Jacobian matrix of sensitivities, program EM1DFM uses an iterative procedure to solve this nonlinear problem. At each iteration, the full nonlinear problem is linearized about the current model, and the subsequent linear inverse problem solved. As the iterations proceed, convergence to the model that solves the nonlinear inverse problem will generally occur.

The four different ways in which the trade-off parameter is dealt with in the four variations of the inversion algorithm are as follows. Which is the most appropriate for a particular data-set depends to a large extent on how well the noise in the observations is known. Detailed descriptions of the relevant mathematics can be found in Sections 2.5.2 to 2.5.5.

1. Fixed user-supplied trade-off parameter

This is the simplest variant of the inversion algorithm since no procedure for automatically determining a suitable value of the trade-off parameter is involved. This means it is the fastest at performing a single inversion. However, the value of the trade-off parameter that will give the best results is almost never known, and can only be determined using this option by a trial-and-error procedure.

2. Discrepancy principle

This approach involves adjusting the trade-off parameter at each iteration so that the misfit approaches a pre-defined target value. If the noise in the observations is uncorrelated, Gaussian and has zero mean, and reliable estimates of the standard deviation of the noise in each observation are available, an appropriate target value is the number of observations. If the standard deviations of the noise have been underestimated and the target misfit cannot therefore be attained, this version of the inversion algorithm will choose the value of the trade-off parameter that gives the minimum possible misfit.

3. GCV criterion

The generalized cross-validation (GCV) function provides a means for estimating the most appropriate value of the trade-off parameter when only the relative amounts of noise in the observations are known. The GCV function is a measure of how well the result of an inversion can predict additional observations. The value of the trade-off parameter that minimizes the GCV function therefore gives the model that best predicts each observation should it be omitted from the inversion.

4. L-curve criterion

If, for a particular iteration of the inversion procedure, the logarithm of the linearized misfit is plotted as a function of the logarithm of the model norm for a wide range of values of the trade-off parameter, the resulting curve tends to have a characteristic "L" shape. The corner of this curve corresponds to equal emphasis being placed on fitting the observations and constructing a simple model. The value of the trade-off parameter corresponding to the point of maximum curvature on the L-curve is therefore the one that is used in this variant of the inversion algorithm.

2. Theoretical Background

2.1 Electromagnetic loop-loop data

Program EM1DFM is designed to interpret frequency-domain, small loop, electromagnetic data. These data are measurements of the magnetic field due to currents and magnetization induced in the Earth by a sinusoidal time-varying current in a small transmitter loop. The receiver loop is assumed to be sufficiently small that the measurements can be considered as point measurements of the magnetic field, and the transmitter loop is assumed to be sufficiently small that it can be represented by a magnetic dipole. Program EM1DFM can handle any combination of measurements made at different frequencies of the transmitter current, and for any separations and heights above the Earth's surface of the transmitter and receiver loops. The loops can be oriented in the x-, y- or z-directions. Any combination of inphase and/or quadrature parts of the fields can be dealt with. Program EM1DFM accepts observations in four different forms: values of the secondary magnetic field (total minus free-space) normalized by the free-space field and given in parts-per-million, values of the secondary field normalized by the free-space field and given in percent, values of the secondary H-field in A/m, and values of the total H-field in A/m. If the transmitter and receiver have the same orientation, the x-, y- and z-components of the secondary field are normalized by the x-, y- and z-components of the free-space field respectively. If the transmitter and receiver have different orientations, the secondary field is normalized by the magnitude of the free-space field.

2.2 The Earth model

Program EM1DFM models the Earth beneath a measurement location as a stack of uniform, horizontal layers. It is the physical properties of the layers that are obtained during the inversion of all measurements gathered at a single location, with the depths to the layer interfaces remaining fixed. If measurements made at multiple locations are being interpreted, the corresponding one-dimensional models are juxtaposed to create a two-dimensional image of the subsurface.

Because horizontal position is meaningless in a model that varies only vertically, all measurements that are to be inverted for a single one-dimensional model must be grouped together as a "sounding". Each different sounding is inverted for a separate one-dimensional model. The horizontal location at which a sounding is centred is not used within the program, but is written out to distinguish results for different soundings.

The electrical conductivity of Earth materials varies over several orders of magnitude, making it more natural to invert for the logarithms of the layer conductivities rather than the conductivities themselves. This also ensures that conductivities in the constructed model are positive. It is not appropriate, however, to invert for the logarithms of the layer susceptibilities. Near-zero values of susceptibility would become overly important and large values would be overestimated (Zhang & Oldenburg, 1997). Therefore susceptibilities are found directly by the inversion, with the option of imposing a positivity constraint.

The coordinate system used for describing the Earth models has z as positive downwards, with the surface of the Earth at $z = 0$. The z-coordinates of the source and receiver locations (which must be above the surface) are therefore negative.

2.3 Forward modelling

The method used to compute the magnetic field values for a particular source-receiver arrangement over a layered Earth model is the matrix propagation approach described in Farquharson et al. (2000). The method uses the z-component of the Schelkunoff $\mathbf{F}$-potential (Ward & Hohmann, 1987):

$$\mathbf{E} = -\nabla \times \mathbf{F}, \tag{1}$$

$$\mathbf{H} = -\sigma \mathbf{F} + \frac{1}{i\omega\mu} \nabla\big(\nabla \cdot \mathbf{F}\big), \tag{2}$$

where $\mathbf{E}$ & $\mathbf{H}$ are the electric and magnetic fields, σ & μ are the conductivity and permeability of the uniform region to which the above equations refer, and a time-dependence of $e^{i\omega t}$ has been assumed.

In the jth layer ($j > 0$), with conductivity σ_j and permeability μ_j, the z-component of the Schelkunoff potential satisfies the equation (assuming the quasi-static approximation):

$$\nabla^2 F_j \; - \; i\omega\mu_j\sigma_j F_j \; = \; 0. \tag{3}$$

The permeability is related to the susceptibility κ_j via the equation: $\mu_j = \mu_0(1+\kappa_j)$, where μ_0 is the permeability of free space. Applying the two-dimensional Fourier transform to eq. (3) gives:

$$\frac{d^2\tilde{F}_j}{dz^2} \; - \; u_j^2\,\tilde{F}_j \; = \; 0, \tag{4}$$

where $u_j^2 = k_x^2 + k_y^2 + i\omega\mu_j\sigma_j$, and k_x and k_y are the horizontal wavenumbers. The solution to this equation is:

$$\tilde{F}_j(k_x,k_y,z,\omega) \; = D_j(k_x,k_y,\omega)\,e^{-u_j(z-z_j)} \; + \; U_j(k_x,k_y,\omega)\,e^{u_j(z-z_j)}, \tag{5}$$

where D_j and U_j are the coefficients of the downward- and upward-decaying parts of the solution.

At the interface between layer $j-1$ and layer j, which is at a depth of z_j, the conditions on $\tilde{F}$ are:

$$\tilde{F}_{j-1}\big|_{z=z_j} \; = \; \tilde{F}_j\big|_{z=z_j}\,, \tag{6}$$

$$\frac{1}{\mu_{j-1}}\,\frac{d\tilde{F}_{j-1}}{dz}\bigg|_{z=z_j} \; = \; \frac{1}{\mu_j}\,\frac{d\tilde{F}_j}{dz}\bigg|_{z=z_j}. \tag{7}$$

Applying these conditions to the solutions in layer j and layer $j-1$ ($j \geq 2$) gives:

$$\begin{pmatrix} e^{-u_{j-1}t_{j-1}} & e^{u_{j-1}t_{j-1}} \\ -\frac{u_{j-1}}{\mu_{j-1}}e^{-u_{j-1}t_{j-1}} & \frac{u_{j-1}}{\mu_{j-1}}e^{u_{j-1}t_{j-1}} \end{pmatrix} \begin{pmatrix} D_{j-1} \\ U_{j-1} \end{pmatrix} \; = \; \begin{pmatrix} 1 & 1 \\ -\frac{u_j}{\mu_j} & \frac{u_j}{\mu_j} \end{pmatrix} \begin{pmatrix} D_j \\ U_j \end{pmatrix}, \tag{8}$$

where $t_{j-1} = z_j - z_{j-1}$ is the thickness of layer $j-1$. Rearranging, explicitly factoring out from the matrix the exponential term whose argument has a positive real part, gives:

$$\begin{pmatrix} D_{j-1} \\ U_{j-1} \end{pmatrix} \; = \; e^{u_{j-1}t_{j-1}}\;\underline{\mathbf{M}}_j\;\begin{pmatrix} D_j \\ U_j \end{pmatrix}, \tag{9}$$

where

$$\underline{\mathbf{M}}_j \; = \; \begin{pmatrix} \frac{1}{2}\left(1+\frac{\mu_{j-1}u_j}{\mu_j u_{j-1}}\right) & \frac{1}{2}\left(1-\frac{\mu_{j-1}u_j}{\mu_j u_{j-1}}\right) \\ \frac{1}{2}\left(1-\frac{\mu_{j-1}u_j}{\mu_j u_{j-1}}\right)e^{-2u_{j-1}t_{j-1}} & \frac{1}{2}\left(1+\frac{\mu_{j-1}u_j}{\mu_j u_{j-1}}\right)e^{-2u_{j-1}t_{j-1}} \end{pmatrix}, \tag{10}$$

for $j \geq 2$.

In layer 0 (the air halfspace), $\tilde{F}$ is given by

$$\tilde{F}_0 \; = D_0\,e^{-u_0 z} \; + \; U_0\,e^{u_0 z}, \tag{11}$$

which leads to

$$\begin{pmatrix} D_0 \\ U_0 \end{pmatrix} \; = \; \underline{\mathbf{M}}_1\;\begin{pmatrix} D_1 \\ U_1 \end{pmatrix} \tag{12}$$

where

$$\underline{\mathbf{M}}_1 \; = \; \begin{pmatrix} \frac{1}{2}\left(1+\frac{\mu_0 u_1}{\mu_1 u_0}\right) & \frac{1}{2}\left(1-\frac{\mu_0 u_1}{\mu_1 u_0}\right) \\ \frac{1}{2}\left(1-\frac{\mu_0 u_1}{\mu_1 u_0}\right) & \frac{1}{2}\left(1+\frac{\mu_0 u_1}{\mu_1 u_0}\right) \end{pmatrix}. \tag{13}$$

Application of eqs. (12) and (8) relates the coefficients, U_0 and D_0, of the solution in the air to those, U_M and D_M, of the solution in the basement halfspace:

$$\begin{pmatrix} D_0 \\ U_0 \end{pmatrix} \;=\; \underline{\mathbf{M}}_1 \;\exp\!\Big(\sum_{j=2}^{M} u_{j-1}t_{j-1}\Big) \;\prod_{j=2}^{M}\underline{\mathbf{M}}_j \;\begin{pmatrix} D_M \\ U_M \end{pmatrix}. \tag{14}$$

There is no upward-decaying part of the solution in the basement halfspace: $U_M = 0$. In the air, the downward-decaying part is due to the source: $D_0 = D_0^S$. Equation (14) can therefore be rewritten as

$$\begin{pmatrix} D_0^S \\ U_0 \end{pmatrix} \;=\; E\,\underline{\mathbf{P}} \;\begin{pmatrix} D_M \\ 0 \end{pmatrix}, \tag{15}$$

where the matrix $\underline{\mathbf{P}}$ is given by

$$\underline{\mathbf{P}} \;=\; \underline{\mathbf{M}}_1 \;\prod_{j=2}^{M}\underline{\mathbf{M}}_j, \tag{16}$$

and the factor E by

$$E \;=\; \exp\!\Big(\sum_{j=2}^{M} u_{j-1}t_{j-1}\Big). \tag{17}$$

The first of the implied pair of equations in eq. (15) gives

$$D_M \;=\; \frac{1}{E}\frac{1}{P_{11}}\,D_0^S, \tag{18}$$

and the second gives

$$U_0 \;=\; E\,P_{21}\,D_M. \tag{19}$$

Substituting eq. (18) into eq. (19) gives

$$U_0 \;=\; \frac{P_{21}}{P_{11}}\,D_0^S, \tag{20}$$

which does not involve any exponential terms whose arguments have positive real parts, making this formulation inherently stable. The solution for $\tilde{F}$ in the air halfspace is therefore given by

$$\tilde{F}_0 \;= D_0^S\Big(e^{-u_0 z} \;+\; \frac{P_{21}}{P_{11}}e^{u_0 z}\Big). \tag{21}$$

For a unit vertical magnetic dipole source at a height h (i.e., $z = -h$ for $h > 0$) above the surface of the Earth:

$$D_0^S \;=\; \frac{i\omega\mu_0}{2u_0}\,e^{-u_0 h} \tag{22}$$

(Ward & Hohmann, 1987, eq. 4.40), and for a unit x-directed magnetic dipole source at $z = -h$:

$$D_0^S \;=\; -\frac{i\omega\mu_0}{2}\,\frac{ik_x}{k_x^2+k_y^2}\,e^{-u_0 h} \tag{23}$$

(Ward & Hohmann, 1987, eq. 4.106). Once whichever of these terms is appropriate is substituted into eq. (21), the solution is completed by converting the required inverse two-dimensional Fourier transform to a Hankel transform, and using eq. (2) to obtain the three components of the H-field above the Earth model ($z < 0$):

$$H_x(x,y,z,\omega) \;=\; \frac{1}{4\pi}\,\frac{x}{r}\int_0^\infty \Big(e^{-\lambda|z+h|} \;-\; \frac{P_{21}}{P_{11}}e^{\lambda(z-h)}\Big)\,\lambda^2\,J_1(\lambda r)\,d\lambda, \tag{24}$$

$$H_y(x,y,z,\omega) \;=\; \frac{1}{4\pi}\,\frac{y}{r}\int_0^\infty \Big(e^{-\lambda|z+h|} \;-\; \frac{P_{21}}{P_{11}}e^{\lambda(z-h)}\Big)\,\lambda^2\,J_1(\lambda r)\,d\lambda, \tag{25}$$

$$H_z(x,y,z,\omega) \;=\; \frac{1}{4\pi}\int_0^\infty \Big(e^{-\lambda|z+h|} \;+\; \frac{P_{21}}{P_{11}}e^{\lambda(z-h)}\Big)\,\lambda^2\,J_0(\lambda r)\,d\lambda \tag{26}$$

for a z-directed magnetic dipole source at $(0, 0, -h)$, $h > 0$, and:

$$H_x(x,y,z,\omega) = -\frac{1}{4\pi}\left(\frac{1}{r} - \frac{2x^2}{r^3}\right)\int_0^\infty \left(e^{-\lambda|z+h|} - \frac{P_{21}}{P_{11}}e^{\lambda(z-h)}\right)\lambda\, J_1(\lambda r)\, d\lambda$$
$$- \frac{1}{4\pi}\frac{x^2}{r^2}\int_0^\infty \left(e^{-\lambda|z+h|} - \frac{P_{21}}{P_{11}}e^{\lambda(z-h)}\right)\lambda^2\, J_0(\lambda r)\, d\lambda, \quad (27)$$

$$H_y(x,y,z,\omega) = \frac{1}{2\pi}\frac{xy}{r^3}\int_0^\infty \left(e^{-\lambda|z+h|} - \frac{P_{21}}{P_{11}}e^{\lambda(z-h)}\right)\lambda\, J_1(\lambda r)\, d\lambda$$
$$- \frac{1}{4\pi}\frac{xy}{r^2}\int_0^\infty \left(e^{-\lambda|z+h|} - \frac{P_{21}}{P_{11}}e^{\lambda(z-h)}\right)\lambda^2\, J_0(\lambda r)\, d\lambda, \quad (28)$$

$$H_z(x,y,z,\omega) = \frac{1}{4\pi}\frac{x}{r}\int_0^\infty \left(e^{-\lambda|z+h|} + \frac{P_{21}}{P_{11}}e^{\lambda(z-h)}\right)\lambda^2\, J_1(\lambda r)\, d\lambda \quad (29)$$

for an x-directed magnetic dipole source at $(0, 0, -h)$, $h > 0$.

The Hankel transforms in eqs.(24) to (29) are computed using the digital filtering routine of Anderson (1982). The kernels of these equations are pre-computed at a certain number of logarithmically-spaced values of λ. Anderson's routine then extracts the values of the kernels at the values of λ it requires by cubic spline interpolation. The number of values of λ at which the kernels are pre-computed (50 minimum) can be specified in the input file "`em1dfm.in`"; see "line 11" in the input file description (Section 3.1.1 of the Manual).

There are three places where previously-computed components of eqs. (24) to (29) can be re-used. The propagation of the matrices through the layers depends on frequency, and must be re-done for each different value. However, the propagated matrix $\underline{\mathbf{P}}$, and hence the ratio P_{21}/P_{11}, does not depend on the relative location and orientation of the transmitter and receiver, and so can be re-used for all transmitters and receivers for the same frequency. Furthermore, if there are multiple transmitter-receiver pairs with the same height (and the same frequency), there is no difference in the kernels of their Hankel transforms, and so the values of the kernels computed for one pair can be re-used for all the others. It is to ensure this grouping of the survey parameters that the observations file is structured the way it is (see Section 3.1.2 of the Manual).

The individual propagation matrices $\underline{\mathbf{M}}_j$, and each matrix computed in the construction of the propagation matrix $\underline{\mathbf{P}}$, are saved in the forward-modelling routine. These are then re-used in the computation of the sensitivities.

2.4 Computing the sensitivities

The inverse problem of determining the conductivity and/or susceptibility of the Earth from electromagnetic measurements is nonlinear. Program EM1DFM uses an iterative procedure to solve this problem. At each iteration the linearized approximation of the full nonlinear problem is solved. This requires the Jacobian matrix of sensitivities, $\underline{\mathbf{J}} = \left(\underline{\mathbf{J}}^\sigma\ \underline{\mathbf{J}}^\kappa\right)$, where:

$$J^\sigma_{ij} = \frac{\partial d_i}{\partial \log \sigma_j}, \quad (30)$$

$$J^\kappa_{ij} = \frac{\partial d_i}{\partial \kappa_j}, \quad (31)$$

in which d_i is the ith observation, and σ_j & κ_j are the conductivity and susceptibility of the jth layer.

The algorithm for computing the sensitivities is obtained by differentiating the expressions for the H-fields (see Section 2.3) with respect to the model parameters (Farquharson et al., 2000). For example, the sensitivity with respect to m_j (either the conductivity or susceptibility of the jth layer) of the z-component of the H-field for a z-directed magnetic dipole source is given by differentiating eq. (26):

$$\frac{\partial H_z}{\partial m_j}(x,y,z,\omega) = \frac{1}{4\pi}\int_0^\infty \left(e^{-\lambda|z+h|} + \frac{\partial}{\partial m_j}\left[\frac{P_{21}}{P_{11}}\right]e^{\lambda(z-h)}\right)\lambda^2\, J_0(\lambda r)\, d\lambda. \quad (32)$$

The derivative of the coefficient is simply

$$\frac{\partial}{\partial m_j}\left[\frac{P_{21}}{P_{11}}\right] \;=\; \frac{\partial P_{21}}{\partial m_j}\,\frac{1}{P_{11}} \;-\; \frac{P_{21}}{P_{11}^2}\,\frac{\partial P_{11}}{\partial m_j}, \tag{33}$$

where P_{11} and P_{21} are elements of the propagation matrix $\underline{\mathbf{P}}$ given by eq. (16). The derivative of $\underline{\mathbf{P}}$ with respect to m_j $(1 \le j \le M-1)$ is

$$\frac{\partial \underline{\mathbf{P}}}{\partial m_j} \;=\; \underline{\mathbf{M}}_1\,\underline{\mathbf{M}}_2 \ldots \underline{\mathbf{M}}_{j-1}\left(\frac{\partial \underline{\mathbf{M}}_j}{\partial m_j}\,\underline{\mathbf{M}}_{j+1} \;+\; \underline{\mathbf{M}}_j\,\frac{\partial \underline{\mathbf{M}}_{j+1}}{\partial m_j}\right)\underline{\mathbf{M}}_{j+2} \ldots \underline{\mathbf{M}}_M. \tag{34}$$

The sensitivities with respect to the conductivity and susceptibility of the basement halfspace are given by

$$\frac{\partial \underline{\mathbf{P}}}{\partial m_M} \;=\; \underline{\mathbf{M}}_1\,\underline{\mathbf{M}}_2 \ldots \underline{\mathbf{M}}_{M-1}\,\frac{\partial \underline{\mathbf{M}}_M}{\partial m_M}. \tag{35}$$

The derivatives of the individual layer matrices with respect to the conductivities and susceptibilities are straightforward to derive, and are not given here.

Just as for the forward modelling, the Hankel transform in eq. (32), and those in the corresponding expressions for the sensitivities of the other observations, are computed using the digital filtering routine of Anderson (1982).

The partial propagation matrices

$$\underline{\mathbf{P}}_k \;=\; \underline{\mathbf{M}}_1 \prod_{j=2}^{k} \underline{\mathbf{M}}_j, \qquad k \;=\; 2, \ldots, M \tag{36}$$

are computed during the forward modelling, and saved for re-use during the sensitivity computations. This sensitivity-equation approach therefore has the efficiency of an adjoint-equation approach.

2.5 The inversion algorithms

In program EM1DFM, there are four different inversion algorithms. They all have the same general formulation (described in Section 2.5.1), but differ in their treatment of the trade-off parameter (see Sections 2.5.2 to 2.5.5). In addition, there are four possibilities for the Earth model constructed by the inversion: (a) just conductivity, (b) just susceptibility (with positivity enforced), (c) both conductivity and susceptibility (with positivity of the susceptibilities enforced), and (d) both conductivity and susceptibility (without the positivity constraint).

2.5.1 General formulation

The aim of each inversion algorithm is to construct the simplest model that adequately reproduces the observations. This is achieved by posing the inverse problem as an optimization problem in which the model is sought that minimizes the objective function:

$$\Phi \;=\; \phi_d \;+\; \beta\,\phi_m \;-\; \gamma\,\phi_{LB}. \tag{37}$$

The three components of this objective function are as follows. ϕ_d is the data misfit:

$$\phi_d \;=\; \left\|\underline{\mathbf{W}}_d\left(\mathbf{d} \;-\; \mathbf{d}^{\mathrm{obs}}\right)\right\|^2, \tag{38}$$

where $\|\cdot\|$ represents the l_2-norm, $\mathbf{d}^{\mathrm{obs}}$ is the vector containing the N observations, and $\mathbf{d}$ is the forward-modelled data. It is assumed that the noise in the observations is Gaussian and uncorrelated, and that the

estimated standard deviation of the noise in the ith observation is of the form $s_0\hat{s}_i$, where $\hat{s}_i$ indicates the amount of noise in the ith observation relative to that in the others, and s_0 is a scale factor that specifies the total amount of noise in the set of observations. The matrix $\underline{\mathbf{W}}_d$ is therefore given by:

$$\underline{\mathbf{W}}_d \;=\; \mathrm{diag}\big\{1/(s_0\hat{s}_1),\,\ldots\,,1/(s_0\hat{s}_N)\big\}. \tag{39}$$

The model-structure component of the objective function is ϕ_m. In its most general form it contains four terms:

$$\begin{aligned}\phi_m \;=\; \alpha_s^\sigma \left\|\underline{\mathbf{W}}_s^\sigma\big(\mathbf{m}^\sigma-\mathbf{m}_s^{\sigma,\mathrm{ref}}\big)\right\|^2 \;+\; \alpha_z^\sigma \left\|\underline{\mathbf{W}}_z^\sigma\big(\mathbf{m}^\sigma-\mathbf{m}_z^{\sigma,\mathrm{ref}}\big)\right\|^2 \;+\\ \alpha_s^\kappa \left\|\underline{\mathbf{W}}_s^\kappa\big(\mathbf{m}^\kappa-\mathbf{m}_s^{\kappa,\mathrm{ref}}\big)\right\|^2 \;+\; \alpha_z^\kappa \left\|\underline{\mathbf{W}}_z^\kappa\big(\mathbf{m}^\kappa-\mathbf{m}_z^{\kappa,\mathrm{ref}}\big)\right\|^2,\end{aligned} \tag{40}$$

where $\mathbf{m}^\sigma$ is the vector containing the logarithms of the layer conductivities, and $\mathbf{m}^\kappa$ is the vector containing the layer susceptibilities. The matrices $\underline{\mathbf{W}}_s^\sigma$ and $\underline{\mathbf{W}}_s^\kappa$ are:

$$\underline{\mathbf{W}}_s^\sigma \;=\; \underline{\mathbf{W}}_s^\kappa \;=\; \mathrm{diag}\big\{\,\sqrt{t_1},\,\ldots\,,\sqrt{t_{M-1}},\,\sqrt{t_{M-1}}\,\big\}, \tag{41}$$

(t_j is the thickness of the jth layer). The matrices $\underline{\mathbf{W}}_z^\sigma$ and $\underline{\mathbf{W}}_z^\kappa$ are:

$$\underline{\mathbf{W}}_z^\sigma \;=\; \underline{\mathbf{W}}_z^\kappa \;=\; \begin{pmatrix} -\sqrt{\frac{2}{t_1+t_2}} & \sqrt{\frac{2}{t_1+t_2}} & & & & \\ & -\sqrt{\frac{2}{t_2+t_3}} & \sqrt{\frac{2}{t_2+t_3}} & & & \\ & & \ddots & & & \\ & & & -\sqrt{\frac{2}{t_{M-2}+t_{M-1}}} & \sqrt{\frac{2}{t_{M-2}+t_{M-1}}} & \\ & & & & -\sqrt{\frac{1}{t_{M-1}}} & \sqrt{\frac{1}{t_{M-1}}} \\ & & & & & 0 \end{pmatrix}. \tag{42}$$

The rows of any of these four weighting matrices can be scaled if desired (see Section 3.1.9 of the Manual). The vectors $\mathbf{m}_s^{\sigma,\mathrm{ref}}$, $\mathbf{m}_z^{\sigma,\mathrm{ref}}$, $\mathbf{m}_s^{\kappa,\mathrm{ref}}$ and $\mathbf{m}_z^{\kappa,\mathrm{ref}}$ contain the layer conductivities/susceptibilities for the four possible reference models. The four terms in ϕ_m therefore correspond to the "smallest" and "flattest" terms for the conductivity and susceptibility parts of the model. The relative importance of the four terms is governed by the coefficients α_s^σ, α_z^σ, α_s^κ and α_z^κ, which are discussed in Section 2.5.6. β is the trade-off parameter that balances the opposing effects of minimizing the misfit and minimizing the amount of structure in the model. It is the different ways in which β is determined that distinguish the four inversion algorithms in program EM1DFM from one another. They are described in the next sections.

Finally, the third component of the objective function is a logarithmic barrier term:

$$\phi_{LB} \;=\; \sum_{j=1}^{M} \log\, c\kappa_j, \tag{43}$$

where c is a constant, usually equal to 1. This term is how the positivity constraint on the layer susceptibilities is enforced. It, and its coefficient γ, are described in Section 2.5.7.

As mentioned in Section 2.4, the inverse problem considered here is nonlinear. It is solved using an iterative procedure. At the nth iteration, the actual objective function being minimized is:

$$\Phi^n \;=\; \phi_d^n \;+\; \beta^n\,\phi_m^n \;-\; \gamma^n\,\phi_{LB}^n. \tag{44}$$

In the misfit ϕ_d^n, the forward-modelled data $\mathbf{d}^n$ are those for the model that is sought at this iteration. They are approximated by:

$$\mathbf{d}^n \;=\; \mathbf{d}^{n-1} \;+\; \underline{\mathbf{J}}^{\sigma,n-1}\,\delta\mathbf{m}^\sigma \;+\; \underline{\mathbf{J}}^{\kappa,n-1}\,\delta\mathbf{m}^\kappa, \tag{45}$$

where $\delta\mathbf{m}^{\sigma} = \mathbf{m}^{\sigma,n} - \mathbf{m}^{\sigma,n-1}$ & $\delta\mathbf{m}^{\kappa} = \mathbf{m}^{\kappa,n} - \mathbf{m}^{\kappa,n-1}$, and $\underline{\mathbf{J}}^{\sigma,n-1}$ & $\underline{\mathbf{J}}^{\kappa,n-1}$ are the two halves of the Jacobian matrix given by eqs (30) & (31) and evaluated for the model from the previous iteration. At the nth iteration, the problem to be solved is that of finding the change, $(\delta\mathbf{m}^{\sigma}, \delta\mathbf{m}^{\kappa})$, to the model which minimizes the objective function Φ^n. Differentiating eq. (44) with respect to the components of $\delta\mathbf{m}^{\sigma}$ & $\delta\mathbf{m}^{\kappa}$, and equating the resulting expressions to zero, gives the system of equations to be solved. The derivatives of ϕ_d^n (incorporating the approximation of eq. 45) and ϕ_m are straightforward to calculate. However, a further approximation must be made to linearize the derivatives of the logarithmic barrier term:

$$\frac{\partial\,\phi_{LB}^n}{\partial\,\delta m_k^{\kappa}} \;=\; \frac{\partial}{\partial\,\delta\kappa_k}\sum_{j=1}^{M}\log\left(\kappa_j^{n-1}+\delta\kappa_j\right), \tag{46}$$

$$\;=\; \frac{1}{\kappa_k^{n-1}+\delta\kappa_j}, \tag{47}$$

$$\;\approx\; \frac{1}{\kappa_k^{n-1}}\left(1\;-\;\frac{\delta\kappa_k}{\kappa_k^{n-1}}\right). \tag{48}$$

The linear system of equations to be solved for $(\delta\mathbf{m}^{\sigma}, \delta\mathbf{m}^{\kappa})$ is therefore:

$$\left[\,\underline{\mathbf{J}}^{n-1T}\underline{\mathbf{W}}_d^T\underline{\mathbf{W}}_d\underline{\mathbf{J}}^{n-1} \;+\; \beta^n\sum_{i=1}^{2}\underline{\mathbf{W}}_i^T\underline{\mathbf{W}}_i \;+\; \frac{\gamma^n}{2}\underline{\hat{\mathbf{X}}}^{n-1T}\underline{\hat{\mathbf{X}}}^{n-1}\,\right]\delta\mathbf{m} \;=\; \underline{\mathbf{J}}^{n-1T}\underline{\mathbf{W}}_d^T\underline{\mathbf{W}}_d\left(\mathbf{d}^{\mathrm{obs}}-\mathbf{d}^{n-1}\right)$$

$$+\;\beta^n\sum_{i=1}^{2}\underline{\mathbf{W}}_i^T\underline{\mathbf{W}}_i\left(\mathbf{m}_i^{\mathrm{ref}}-\mathbf{m}^{n-1}\right) \;+\; \frac{\gamma^n}{2}\underline{\hat{\mathbf{X}}}^{n-1T}\underline{\hat{\mathbf{X}}}^{n-1}\mathbf{m}^{n-1}, \tag{49}$$

where:

$$\underline{\mathbf{J}}^{n-1} \;=\; \left(\underline{\mathbf{J}}^{\sigma,n-1}\;\underline{\mathbf{J}}^{\kappa,n-1}\right) \tag{50}$$

$$\underline{\mathbf{W}}_1 \;=\; \begin{pmatrix}\sqrt{\alpha_s^{\sigma}}\,\underline{\mathbf{W}}_s^{\sigma} & 0\\ 0 & \sqrt{\alpha_s^{\kappa}}\,\underline{\mathbf{W}}_s^{\kappa}\end{pmatrix}, \tag{51}$$

$$\underline{\mathbf{W}}_2 \;=\; \begin{pmatrix}\sqrt{\alpha_z^{\sigma}}\,\underline{\mathbf{W}}_z^{\sigma} & 0\\ 0 & \sqrt{\alpha_z^{\kappa}}\,\underline{\mathbf{W}}_z^{\kappa}\end{pmatrix}, \tag{52}$$

$\mathbf{m}_1^{\mathrm{ref}} = (\mathbf{m}_s^{\sigma,\mathrm{ref}\,T}\;\mathbf{m}_s^{\kappa,\mathrm{ref}\,T})^T$, $\mathbf{m}_2^{\mathrm{ref}} = (\mathbf{m}_z^{\sigma,\mathrm{ref}\,T}\;\mathbf{m}_z^{\kappa,\mathrm{ref}\,T})^T$, and

$$\underline{\hat{\mathbf{X}}}^{n-1} \;=\; \begin{pmatrix}0 & (\underline{\mathbf{X}}^{n-1})^{-1}\end{pmatrix}, \tag{53}$$

where $\underline{\hat{\mathbf{X}}}^{n-1} = \mathrm{diag}\{m_1^{\kappa,n-1},\ldots,m_M^{\kappa,n-1}\}$. The solution to equation (49) is equivalent to the least-squares solution of:

$$\begin{pmatrix}\underline{\mathbf{W}}_d\underline{\mathbf{J}}^{n-1}\\ \sqrt{\beta^n}\,\underline{\mathbf{W}}_1\\ \sqrt{\beta^n}\,\underline{\mathbf{W}}_2\\ \sqrt{\gamma^n/2}\,\underline{\hat{\mathbf{X}}}^{n-1}\end{pmatrix}\delta\mathbf{m} \;=\; \begin{pmatrix}\underline{\mathbf{W}}_d\left(\mathbf{d}^{\mathrm{obs}}-\mathbf{d}^{n-1}\right)\\ \sqrt{\beta^n}\,\underline{\mathbf{W}}_1\left(\mathbf{m}_1^{\mathrm{ref}}-\mathbf{m}^{n-1}\right)\\ \sqrt{\beta^n}\,\underline{\mathbf{W}}_2\left(\mathbf{m}_2^{\mathrm{ref}}-\mathbf{m}^{n-1}\right)\\ \sqrt{\gamma^n/2}\,\underline{\hat{\mathbf{X}}}^{n-1}\,\mathbf{m}^{n-1}\end{pmatrix}. \tag{54}$$

Once the step $\delta\mathbf{m}$ has been determined by the solution of eq. (49) or eq. (54), the new model is given by:

$$\mathbf{m}^n \;=\; \mathbf{m}^{n-1} \;+\; \nu\,\delta\mathbf{m}. \tag{55}$$

There are two conditions on the step length ν. First, if positivity of the layer susceptibilities is being enforced:

$$\nu\,\delta\kappa_j \;>\; -\kappa_j^{n-1} \tag{56}$$

must hold for all $j = 1,\ldots,M$. Secondly, the objective function must be decreased by the addition of the step to the model:

$$\phi_d^n \;+\; \beta^n\,\phi_m^n \;-\; \gamma^n\,\phi_{LB}^n \;<\; \phi_d^{n-1} \;+\; \beta^n\,\phi_m^{n-1} \;-\; \gamma^n\,\phi_{LB}^{n-1} \tag{57}$$

where ϕ_d^n is now the misfit computed using the full forward modelling for the new model $\mathbf{m}^n$. To determine $\mathbf{m}^n$, a step length, ν, of either 1, or the maximum value for which eq. (56) is true, whichever is greater, is tried. If eq. (57) is true for this step length, it is accepted. If eq. (57) is not true, ν is decreased by factors of 2 until it is true.

2.5.2 Algorithm 1: fixed trade-off parameter

The trade-off parameter, β, remains fixed at its user-supplied value throughout the inversion. The least-squares solution of eq. (54) is used. This is computed using the subroutine "`LSQR`" of Paige & Saunders (1982). If the desired value of β is known, this is the fastest of the four inversion algorithms as it does not involve a line search over trial values of β at each iteration. If the appropriate value of β is not known, it can be found using this algorithm by trail-and-error. This may or may not be time-consuming.

2.5.3 Algorithm 2: discrepancy principle

If a complete description of the noise in a set of observations is available, that is, both s_0 and $\hat{s}_i$ $(i = 1, \ldots, N)$ are known, the expectation of the misfit, $E(\phi_d)$, is equal to the number of observations N. Algorithm 2 therefore attempts to choose the trade-off parameter so that the misfit for the final model is equal to a target value of `chifac` $\times N$. If the noise in the observations is well known, `chifac` should equal 1. However, `chifac` can be adjusted by the user to give a target misfit appropriate for a particular data-set. If a misfit as small as the target value cannot be achieved, the algorithm searches for the smallest possible misfit.

Experience has shown that choosing the trade-off parameter at early iterations in this way can lead to excessive structure in the model, and that removing this structure once the target (or minimum) misfit has been attained can require a significant number of additional iterations. A restriction is therefore placed on the greatest-allowed decrease in the misfit at any iteration, thus allowing structure to be slowly but steadily introduced into the model. In program EM1DFM, the target misfit at the nth iteration is given by

$$\phi_d^{n,\mathrm{tar}} \;=\; \max\bigl(\,\texttt{mfac} \times \phi_d^{n-1}, \texttt{chifac} \times N\,\bigr), \tag{58}$$

where the user-supplied factor `mfac` is such that $0.1 \leq$ `mfac` ≤ 0.5.

The step $\delta\mathbf{m}$ is found from the solution of eq. (54) using subroutine `LSQR` of Paige & Saunders (1982). The line search at each iteration moves along the ϕ_d versus $\log\beta$ curve until either the target misfit, $\phi_d^{n,\mathrm{tar}}$, is bracketed, in which case a bisection search is used to converge to the target, or the minimum misfit $(> \phi_d^{n,\mathrm{tar}})$ is bracketed, in which case a golden section search (for example, Press et al., 1986) is used to converge to the minimum. The starting value of β for each line search is β^{n-1}. For the first iteration, the starting value of β $(= \beta^0)$ for the line search is given by $N/\phi_m(\mathbf{m}^\dagger)$, where $\mathbf{m}^\dagger$ contains typical values of conductivity and/or susceptibility. (Specifically, $\mathbf{m}^\dagger$ is a model whose top $M/5$ layers have a conductivity of 0.02 S/m and susceptibility of 0.02 SI units, and whose remaining layers have a conductivity of 0.01 S/m and susceptibility of 0 SI units. Also, the reference models used in the computation of $\phi_m(\mathbf{m}^\dagger)$ are homogeneous halfspaces of 0.01 S/m and 0 SI units.) The line search is efficient, but does involve the full forward modelling to compute the misfit for each trial value of β.

2.5.4 Algorithm 3: GCV criterion

If only the relative amount of noise in the observations is known, that is, $\hat{s}_i$ $(i = 1, \ldots, N)$ is known but not s_0, the appropriate target value for the misfit cannot be determined, and hence Algorithm 2 is not the most suitable. The generalized cross-validation (GCV) method provides a means of estimating, during the course of an inversion, a value of the trade-off parameter that results in an appropriate fit to the observations, and in so doing, effectively estimating the level of noise, s_0, in the observations (see, for example, Wahba, 1990; Hansen, 1998).

The GCV method is based on the following argument (Wahba, 1990; Haber, 1997; Haber & Oldenburg, 2000). Consider inverting all but the first observation using a trial value of β, and then computing the

individual misfit between the first observation and the first forward-modelled datum for the model produced by the inversion. This can be repeated leaving out all the other observations in turn, inverting the retained observations using the same value of β, and computing the misfit between the observation left out and the corresponding forward-modelled datum. The best value of β can then be defined as the one which gives the smallest sum of all the individual misfits. For a linear problem, this corresponds to minimizing the GCV function. For a nonlinear problem, the GCV method can be applied to the linearized problem being solved at each iteration (Haber, 1997; Haber & Oldenburg, 2000; Li & Oldenburg, 2000; Farquharson & Oldenburg, 2000). From eq. (49), the GCV function for the nth iteration is given by:

$$GCV(\beta) \;=\; \frac{\left\| \underline{\mathbf{W}}_d \hat{\mathbf{d}} \;-\; \underline{\mathbf{W}}_d \underline{\mathbf{J}}^{n-1} \, \underline{\mathbf{M}}^{-1} \big(\underline{\mathbf{J}}^{n-1T} \underline{\mathbf{W}}_d^T \underline{\mathbf{W}}_d \, \hat{\mathbf{d}} + \mathbf{r} \big) \right\|^2}{\big[\mathrm{trace}\big(\underline{\mathbf{I}} \;-\; \underline{\mathbf{W}}_d \underline{\mathbf{J}}^{n-1} \, \underline{\mathbf{M}}^{-1} \underline{\mathbf{J}}^{n-1T} \underline{\mathbf{W}}_d^T \big) \big]^2}, \tag{59}$$

where:

$$\underline{\mathbf{M}}(\beta) \;=\; \Big[\underline{\mathbf{J}}^{n-1T} \underline{\mathbf{W}}_d^T \underline{\mathbf{W}}_d \underline{\mathbf{J}}^{n-1} \;+\; \beta \sum_{i=1}^{2} \underline{\mathbf{W}}_i^T \underline{\mathbf{W}}_i \;+\; \frac{\gamma^n}{2} \hat{\underline{\mathbf{X}}}^{n-1T} \hat{\underline{\mathbf{X}}}^{n-1} \Big], \tag{60}$$

$$\mathbf{r} \;=\; \beta \sum_{i=1}^{2} \underline{\mathbf{W}}_i^T \underline{\mathbf{W}}_i \big(\mathbf{m}_i^{\mathrm{ref}} - \mathbf{m}^{n-1} \big) \;+\; \frac{\gamma^n}{2} \hat{\underline{\mathbf{X}}}^{n-1T} \hat{\underline{\mathbf{X}}}^{n-1} \mathbf{m}^{n-1}, \tag{61}$$

and $\hat{\mathbf{d}} = \mathbf{d}^{\mathrm{obs}} - \mathbf{d}^{n-1}$. If $\beta^\star$ is the value of the trade-off parameter that minimizes eq. (59) at the nth iteration, the actual value of β used to compute the new model is given by:

$$\beta^n \;=\; \max\big(\beta^\star, \mathtt{bfac} \times \beta^{n-1} \big), \tag{62}$$

where the user-supplied factor `bfac` is such that $0.01 < \mathtt{bfac} < 0.5$. As for Algorithm 2, this limit on the allowed decrease in the trade-off parameter prevents unnecessary structure being introduced into the model at early iterations.

The inverse of the matrix $\underline{\mathbf{M}}$ required in eq. (59), and the solution to eq. (49) given this inverse, is computed using the Cholesky factorization routines from LAPACK (Anderson et al., 1999). The line search at each iteration moves along the curve of the GCV function versus the logarithm of the trade-off parameter until the minimum is bracketed (or $\mathtt{bfac} \times \beta^{n-1}$ reached), and then a golden section search (e.g., Press et al., 1986) is used to converge to the minimum. The starting value of β in the line search is β^{n-1} (β^0 is estimated in the same way as for Algorithm 2). This is an efficient search, even with the inversion of the matrix $\underline{\mathbf{M}}$.

2.5.5 Algorithm 4: L-curve criterion

As for the GCV-based method described in Section 2.5.4, the L-curve method provides a means of estimating an appropriate value of the trade-off parameter if only $\hat{s}_i$, $i = 1, \ldots, N$, are known and not s_0. For a linear inverse problem, if the misfit, ϕ_d, is plotted against the model norm, ϕ_m, for all reasonable values of the trade-off parameter, β, the resulting curve tends to have a characteristic "L"-shape, especially when plotted on logarithmic axes (see, for example, Hansen, 1998). The corner of this L-curve corresponds to roughly equal emphasis on the misfit and model norm during the inversion. Moving along the L-curve away from the corner is associated with a progressively smaller decrease in the misfit for large increases in the model norm, or a progressively smaller decrease in the model norm for large increases in the misfit. The value of β at the point of maximum curvature on the L-curve is therefore the most appropriate, according to this criterion.

For a nonlinear problem, the L-curve criterion can be applied to the linearized inverse problem at each iteration (Li & Oldenburg, 1999; Farquharson & Oldenburg, 2000). In this situation, the L-curve is defined using the linearized misfit, which uses the approximation given in eq. (45) for the forward-modelled data. The curvature of the L-curve is computed using the formula (Hansen, 1998):

$$C(\beta) \;=\; \frac{\zeta' \eta'' - \zeta'' \eta'}{\big((\zeta')^2 + (\eta')^2 \big)^{3/2}}, \tag{63}$$

where $\zeta = \log \phi_d^{\rm lin}$ and $\eta = \log \phi_m$. The prime denotes differentiation with respect to $\log \beta$. As for both Algorithms 2 & 3, a restriction is imposed on how quickly the trade-off parameter can be decreased from one iteration to the next. The actual value of β chosen for use at the nth iteration is given by eq. (62), where $\beta^\star$ now corresponds to the value of β at the point of maximum curvature on the L-curve.

Experience has shown that the L-curve for the inverse problem considered here does not always have a sharp, distinct corner. The associated slow variation of the curvature with β can make the numerical differentiation required to evaluate eq. (63) prone to numerical noise. The line search along the L-curve used in program EM1DFM to find the point of maximum curvature is therefore designed to be robust (rather than efficient). The L-curve is sampled at equally-spaced values of $\log \beta$, and long differences are used in the evaluation of eq. (63) to introduce some smoothing. A parabola is fit through the point from the equally-spaced sampling with the maximum value of curvature and its two nearest neighbours. The value of β at the maximum of this parabola is taken as $\beta^\star$. In addition, it is sometimes found that, for the range of values of β that are tried, the maximum value of the curvature of the L-curve on logarithmic axes is negative. In this case, the curvature of the L-curve on linear axes is investigated to find a maximum. As for Algorithms 1 & 2, the least-squares solution to eq. (54) is used, and is computed using subroutine `LSQR` of Paige & Saunders (1982).

2.5.6 Relative weighting within the model norm

The four coefficients in the model norm (see eq. 40) are ultimately the responsibility of the user. Larger values of α_s^σ relative to α_z^σ result in constructed conductivity models that are closer to the supplied reference model. Smaller values of α_s^σ relative to α_z^σ result in flatter conductivity models. Likewise for the coefficients related to susceptibilities.

If both conductivity and susceptibility are active in the inversion, the relative size of α_s^σ & α_z^σ to α_s^κ & α_z^κ is also required. Program EM1DFM includes a simple means of calculating a default value for this relative balance. Using the layer thicknesses, weighting matrices $\underline{\mathbf{W}}_s^\sigma$, $\underline{\mathbf{W}}_z^\sigma$, $\underline{\mathbf{W}}_s^\kappa$ & $\underline{\mathbf{W}}_z^\kappa$, and user-supplied weighting of the smallest and flattest parts of the conductivity and susceptibility components of the model norm (see *acs*, *acz*, *ass* & *asz* in the input file description, line 5, Section 3.1.1), the following two quantities are computed for a test model $\mathbf{m}^\star$:

$$\phi_m^\sigma \;=\; acs \left\| \underline{\mathbf{W}}_s^\sigma \left(\mathbf{m}^\star - \mathbf{m}_s^{\sigma,\rm ref} \right) \right\|^2 \;+\; acz \left\| \underline{\mathbf{W}}_z^\sigma \left(\mathbf{m}^\star - \mathbf{m}_z^{\sigma,\rm ref} \right) \right\|^2, \tag{64}$$

$$\phi_m^\kappa \;=\; ass \left\| \underline{\mathbf{W}}_s^\kappa \left(\mathbf{m}^\star - \mathbf{m}_s^{\kappa,\rm ref} \right) \right\|^2 \;+\; asz \left\| \underline{\mathbf{W}}_z^\kappa \left(\mathbf{m}^\star - \mathbf{m}_z^{\kappa,\rm ref} \right) \right\|^2. \tag{65}$$

The conductivity and susceptibility of the top $N/5$ layers in the test model are 0.02 S/m and 0.02 SI units respectively, and the conductivity and susceptibility of the remaining layers are 0.01 S/m and 0 SI units. The coefficients of the model norm used in the inversion are then $\alpha_s^\sigma = acs$, $\alpha_z^\sigma = acz$, $\alpha_s^\kappa = A^s \times ass$ & $\alpha_z^\kappa = A^s \times asz$ where $A^s = \phi_m^\sigma / \phi_m^\kappa$. It has been found that a balance between the conductivity and susceptibility portions of the model norm computed in this way is adequate as an initial guess. However, the balance usually requires modification by the user to obtain the best susceptibility model. (The conductivity model tends to be insensitive to this balance.) If anything, the default balance will suppress the constructed susceptibility model.

2.5.7 Positivity of susceptibility

Program EM1DFM can perform an unconstrained inversion for susceptibilities (along with the conductivities) as well as invert for values of susceptibility that are constrained to be positive. Following Li & Oldenburg (2000), the positivity constraint is implemented by incorporating a logarithmic barrier term in the objective function (see eqs. 37 & 43). For the initial iteration, the coefficient of the logarithmic barrier term is chosen so that this term is of equal important to the rest of the objective function:

$$\gamma^0 \;=\; \frac{\phi_d^0 + \beta^0 \phi_m^0}{-\phi_{LB}^0}. \tag{66}$$

At subsequent iterations, the coefficient is reduced according to the formula:

$$\gamma^n \;=\; \big(1 \;-\min(\nu^{n-1}, 0.925)\big)\,\gamma^{n-1}, \tag{67}$$

where ν^{n-1} is the step length used at the previous iteration. As mentioned at the end of Section 2.5.1, when positivity is being enforced, the step length at any particular iteration must satisfy eq. (56).

2.5.8 Convergence criteria

To determine when an inversion algorithm has converged, the following criteria are used (Gill et al., 1981):

$$\Phi^{n-1} \;-\; \Phi^n \;<\; \tau\,(1+\Phi^n), \tag{68}$$

$$\|\,\mathbf{m}^{n-1} - \mathbf{m}^n\| \;<\; \sqrt{\tau}\,(1+\|\mathbf{m}^n\|), \tag{69}$$

where τ is a user-specified parameter. The algorithm is considered to have converged when both of the above equations are satisfied. The default value of τ is 0.01.

In case the algorithm happens directly upon the minimum, an additional condition is tested:

$$\|\mathbf{g}^n\| \;\leq\; \epsilon, \tag{70}$$

where ϵ is a small number close to zero, and where the gradient, $\mathbf{g}^n$, at the nth iteration is given by

$$\mathbf{g}^n \;=\; -2\,\underline{\mathbf{J}}^{n\,T}\underline{\mathbf{W}}_d^T\underline{\mathbf{W}}_d\big(\mathbf{d}^{\rm obs} - \mathbf{d}^n\big) \;-\; 2\,\beta^n\sum_{i=1}^{2}\underline{\mathbf{W}}_i^T\underline{\mathbf{W}}_i\big(\mathbf{m}_i^{\rm ref} - \mathbf{m}^n\big) \;-\; \gamma^n\,\hat{\underline{\mathbf{X}}}^{n2T}\mathbf{m}^n. \tag{71}$$

3. References

Anderson, E., Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenny & D. Sorensen, 1999. LAPACK Users' Guide (3rd ed.), SIAM, Philadelphia.

Anderson, W.L., 1982. Fast Hankel transforms using related and lagged convolutions, ACM Transactions on Mathematical Software, 8, 344–368.

Farquharson, C.G., D.W. Oldenburg & P.S. Routh, 2000. Simultaneous one-dimensional inversion of electromagnetic loop-loop data for both magnetic susceptibility and electrical conductivity, to be submitted to Geophysics.

Farquharson, C.G., & D.W. Oldenburg, 2000. Automatic estimation of the trade-off parameter in nonlinear inverse problems, to be submitted to Geophysics.

Gill, P.E., W. Murray & M.H. Wright, 1981. Practical Optimization, Academic Press, London.

Haber, E., 1997. Numerical Strategies for the Solution of Inverse Problems, Ph.D. thesis, University of British Columbia.

Haber, E., & D.W. Oldenburg, 2000. A GCV based method for nonlinear ill-posed problems, Computational Geosciences, 4, 41–63.

Hansen, P.C., 1998. Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion, SIAM, Philadelphia.

Li, Y., & D.W. Oldenburg, 1999. 3-D inversion of DC resistivity data using an L-curve criterion, Extended Abstracts, SEG, Houston, 31 October – 5 November 1999.

Li, Y., & D.W. Oldenburg, 2000. Fast inversion of large-scale magnetic data using wavelet transforms, Geophysics (accepted subject to revision).

Paige, C.C., & M.A. Saunders, 1982. Algorithm 583, LSQR: Sparse linear equations and least-squares problems, ACM Transactions on Mathematical Software, 8, 195–209.

Press, W.H., B.P. Flannery, S.A. Teukolsky & W.T. Vetterling, 1986. Numerical Recipes: The Art of Scientific Computing, Cambridge University Press, Cambridge.

Spies, B.R., & F.C. Frischknecht, 1987. Electromagnetic Sounding, in Electromagnetic Methods in Applied Geophysics, vol. 2, M.N. Nabighian (ed.), 285–425, Society of Exploration Geophysicists, Tulsa, Oklahoma.

Wahba, G., 1990. Spline methods for observational data, SIAM, Philadelphia.

Ward, S.H., & G.W. Hohmann, 1987. Electromagnetic Theory for Geophysical Applications, in Electromagnetic Methods in Applied Geophysics, vol. 1, M.N. Nabighian (ed.), 131–311, Society of Exploration Geophysicists, Tulsa, Oklahoma.

Zhang, Z., & D.W. Oldenburg, 1997. Recovering magnetic susceptibility from electromagnetic data over a one-dimensional earth, Geophysical Journal International, 130, 422–434.