

Background for Program “EM1DTM”

Version 1.0

Developed under the consortium research project
TIME DOMAIN INVERSION AND MODELLING OF
ELECTROMAGNETIC DATA

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

1 June 2006



Contents

Acknowledgments	ii
Contact Information	ii
Abstract	1
1. Program EM1DTM: An Overview	2
2. Theoretical Background	6
2.1 The observations	6
2.2 The Earth model	6
2.3 Forward modelling	6
2.4 Frequency- to time-domain transformation	9
2.5 Computing the sensitivities	11
2.6 The inversion algorithms	12
2.6.1 General formulation	12
2.6.2 General measures	13
2.6.3 General algorithm	14
2.6.4 Algorithm 1: fixed trade-off parameter	15
2.6.5 Algorithm 2: discrepancy principle	15
2.6.6 Algorithm 3: GCV criterion	16
2.6.7 Algorithm 4: L-curve criterion	17
2.6.8 Convergence criteria	17
3. References	18

Acknowledgments

This project was made possible by funding from the “TIME” consortium, of which the following companies are participants: Anglo American Corp., ENI Exploration & Production Division, Falconbridge Ltd. (Exploration), INCO Exploration & Technical Services Inc., M.I.M. Exploration Pty. Ltd., Newmont Exploration Ltd., Placer Dome Inc., Teck-Cominco Ltd., and from the Natural Sciences and Engineering Research Council of Canada.

Contact Information

UBC-GIF Director: Doug Oldenburg,
Author: Colin Farquharson,
Outreach Coordinator: Francis Jones.

UBC – Geophysical Inversion Facility,
Department of Earth & Ocean Sciences,
University of British Columbia,
6339 Stores Road,
Vancouver, B.C., V6T 1Z4,
CANADA.

<http://www.eos.ubc.ca/research/ubcgif>

ubcgif@eos.ubc.ca

Abstract

The name: EM1DTM – electromagnetics (“EM”), one-dimensional models (“1D”), time-domain observations (“T”), and magnetic sources and receivers (“M”).

The observations are values of voltage (i.e., dB/dt) or magnetic field. Receiver coils can be oriented in the x -, y - or z -directions, and they can be at any position relative to the transmitter loop. The transmitter loop can have any number of sides (greater than 2), and can be at any height above the ground surface. It is assumed to be horizontal. The transmitter current waveform can be a step off, a linear ramp turn off, or a general waveform that is provided in discretized form. Observations can be for any time after the step or ramp turn off, or any time after or during a discretized waveform. All the observations (in any combination) that are provided for a particular transmitter loop constitute a “sounding”, and are used to construct the one-dimensional model for that sounding. Measurement uncertainties can be in the same units as the observations or as relative uncertainties in percent.

The product: an electrical conductivity model. The Earth models are composed of layers of uniform conductivity with fixed interface depths. The value of the conductivity 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, with the sequence of one-dimensional models written out. These can be viewed directly as a composite two-dimensional image using the graphical user interface, or converted to a format which is suitable for viewing as a three-dimensional image using MeshTools3D.

General measures for both the measure of data misfit and the measure of the amount of model structure:

- Huber M -measure for data misfit, and
- Ekbloom p -measure for model structure, allow for a whole suite of variations, from the traditional sum-of-squares measures, to more robust measures which can ignore outliers in the observations and which can generate piecewise-constant models.

Four possible methods for determining the degree of regularization:

- the trade-off parameter is specified by the user, either as a single constant value, or with a cooling schedule to some final value, 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.

Documentation is provided in two parts:

- the description herein of the forward modelling and inversion algorithms used by program EM1DTM, and
- the document “Manual for Program EM1DTM” containing detailed descriptions of the inputs to and outputs from program EM1DTM (and description of the related independent forward-modelling program).

1. Program EM1DTM: 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 EM1DTM. Detailed descriptions of the mathematics of these algorithms can be found in the second section.

Time-domain electromagnetic surveys

Program EM1DTM is designed to interpret typical time-domain electromagnetic survey data, whether the measurements are of the magnetic field or its time derivative or both, whether the transmitter current waveform is a step or ramp turn-off or some general waveform, and whether the transmitter loop and receiver or receivers are on the ground or in the air. Such data are mostly affected by the spatial variation of the electrical conductivity of the subsurface, as this controls the strength, distribution and decay rate of electric currents which are induced in the ground by the time-varying current in the transmitter, and it is the magnetic field (or its time derivative) associated with these induced currents which is measured. The goal of EM1DTM is to construct a model of the conductivity variation in the subsurface from the measured magnetic field values. Examples of what knowledge about the subsurface conductivity variation can reveal are the location and size of an ore deposit in its surrounding host rocks, the thickness of overburden that has to be removed during a mining operation, or the depth to an aquifer and its extent. For a review of time-domain electromagnetic methods in geophysics see Nabighian & Macnae (1987).

One-dimensional models of the Earth

Program EM1DTM 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 within each layer that is determined. In EM1DTM, unlike in program EM1DFM, only the conductivity in each layer is sought in an inversion. (The magnetic susceptibilities of the layers are assumed to be all equal to that of free space.) When measurements from different locations are being interpreted, the one-dimensional models from all the locations are written out, and can be juxtaposed to create a two-dimensional image of the subsurface if the data were collected along a line, or a three-dimensional image if the data were collected over a grid.

The forward-modelling algorithm

For a given layered model, the values of the magnetic field for particular arrangements of source and receivers are computed using a matrix propagation, spectral expansion approach in the frequency domain, with the required time-domain values of the magnetic field or its time derivative calculated via inverse Fourier transform and convolution. This is described in more detail in Sections 2.3 & 2.4 of this document. The frequency-domain computations follow those described in Farquharson & Oldenburg (1996) and Farquharson, Oldenburg & Routh (2003), and the frequency-to-time conversion follows that in Farquharson & Oldenburg (1993). For a comprehensive presentation of the mathematics relevant to geophysical electromagnetic methods see Ward & Hohmann (1987).

The source is an extended, multi-sided, closed, horizontal loop. The fundamental building-block for this kind of source is a horizontal electric dipole, with the response of the whole loop obtained by integrating the responses of the electric dipoles around the loop. Such an electric dipole source has two distinct contributions: one from the flow of alternating current along the length of the dipole, and one from the flow of alternating current into and out of the ends of the dipole. Because the actual source under consideration is a closed loop, the contributions from the ends of the electric dipoles would cancel out, leaving only the contributions from the current flowing along the dipoles. To avoid any possible round-off errors, the contribution from the ends of the electric dipoles is never included in the numerical integration around the whole transmitter loop.

The contribution to the response coming from the flow of alternating current along the axis of an electric dipole is purely inductive, and so the magnetic field can be derived from a scalar potential – the vertical component of the Schelkunoff F-potential (just as for the magnetic dipole sources in EM1DFM). Also, 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 spatial derivatives of the potential 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.

Frequency- to time-domain transformation

The basic responses in the time-domain, that is, the H-field and its time derivative for a step-off in the transmitter current waveform, are obtained from the inverse Fourier transform of the frequency-domain expressions for the H-field due to the many-sided transmitter loop. Because of causality, these responses can be written in terms of cosine and sine transforms of the imaginary part of the H-field (Newman, Hohmann & Anderson, 1986), and are evaluated using Anderson's digital filtering routines (Anderson, 1982). The H-field and its time derivative for a linear ramp turn-off in the transmitter current, or a general current waveform, are then obtained by direct convolution of these basic responses with the (time derivative of the) transmitter current waveform. Program EM1DTM expects a general transmitter current waveform to be provided as a sequence of time-current pairs. It is then assumed the current varies linearly between the provided values. Also, for the general transmitter current waveform option, observations can be during the on-time as well as during the off-time.

The sensitivity-computing algorithm

The inverse problem of determining conductivity from electromagnetic measurements is nonlinear. Program EM1DTM therefore uses 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 EM1DTM using essentially a sensitivity-equation approach. (See Section 2.5 of this document, and Farquharson et al., 2003, 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 does the forward modelling to compute the magnetic fields. These matrices are saved within program EM1DTM 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 strategy

For reasons of stability, flexibility and proper treatment of the nonuniqueness inherent in this inverse problem, it is formulated as an under-determined optimization problem in which there are significantly more layers in the one-dimensional model than there are observations for any one sounding, and the model that minimizes the objective function:

$$\Phi = \phi_d + \beta \phi_m,$$

is sought. The first term above, ϕ_d , is a measure of the misfit between the observations and the data computed for a particular conductivity model. The second term involves ϕ_m , which is a measure of the complexity of the model. The model which minimizes this objective function will therefore have both a small data misfit, and thus match the observations to a certain extent, and a small measure of complexity, and thus be relatively simple and without any unnecessary features. The trade-off between how closely the model reproduces the observations and how simple it is depends on the coefficient β .

General measures

The traditional form of the measures used for the data misfit, ϕ_d , and the measure of model complexity, ϕ_m , is the sum-of-squares, or l_2 -norm, measure. This is because it is the most appropriate measure, statistically speaking, if the noise in a set of observations is Gaussian, and because it results in the objective function Φ being a quadratic functional of the model parameters if the relationship between the observations and model parameters is linear. However, if the noise in the observations is not Gaussian, for example, if there are outliers, any fit that is achieved by minimising a sum-of-squares misfit may be biased. Also, models generated by minimizing a sum-of-squares measure of complexity tend to have a smeared-out, fuzzy appearance because minimizing such a measure biases against large variations in the model parameter between neighbouring cells. Diffuse features in a model might be appropriate in certain scenarios, but might not be so in others. Because of the above issues, program EM1DTM implements general measures of data misfit and model complexity, specifically, the Huber M -measure for the data misfit, and Eklblom's l_p -like measure for the measure of model complexity. These two measures each have one variable parameter which can change their behaviour from essentially sum-of-squares measures to a robust measure of misfit and a measure that can produce piecewise-constant, blocky models. The implementation follows that of Farquharson & Oldenburg (1998), which uses the iteratively reweighted least squares (IRLS) algorithm. Robust fits to data and/or blocky models can be generated with only slightly more computational effort than that required using traditional sum-of-squares measures. This feature was not available in program EM1DFM.

The inversion algorithms

As mentioned above in conjunction with the need for the Jacobian matrix of sensitivities, the inverse problem of determining the layer conductivities from the observed magnetic fields (or their time derivatives) is nonlinear. Program EM1DTM uses an iterative procedure to solve this nonlinear inverse 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.

There are four different variations of the inversion algorithm within program EM1DTM. They differ in how the trade-off parameter β is specified for an inversion, or automatically determined during the course of an inversion. 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. The four are summarized below; detailed descriptions of each are given in Section 2.6, and in Farquharson & Oldenburg (2004).

1. *User-supplied cooling schedule*

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 final 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. (This algorithm has been expanded from its equivalent in program EM1DFM: now an initial, large value of the trade-off parameter, and a factor by which the trade-off parameter is decreased at each iteration, can be supplied in addition to the final value of the trade-off parameter.)

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.

The above four variations of the inversion algorithm have demonstrated their usefulness when used with sum-of-squares measures, in program EM1DFM and elsewhere. In testing to date, they have also been successfully used with general measures in program EM1DTM.

2. Theoretical Background

2.1 The observations

Program EM1DTM is designed to interpret time-domain electromagnetic data for arbitrarily sized loop sources. These data are measurements of the magnetic field, or its time derivative, due to currents induced in the Earth by a time-varying current in the transmitter loop. The receiver loops are assumed to be sufficiently small that the measurements can be considered as occurring at points in space. The transmitter loop is a closed horizontal loop which can be made up of any number of straight segments. Program EM1DTM can handle any number of receivers and measurement times for a single transmitter loop, for any separations and heights of the transmitter and receiver loops on or above the Earth's surface. The receiver loops can be oriented in the x -, y - or z -directions. The transmitter current waveform can be of three types: a step off, a linear ramp turn off, or a general discretized waveform which is specified in terms of time-current pairs, with the current assumed to vary linearly between the specified current values. All three forms of the waveform can also have multiple alternating previous cycles. Program EM1DTM accepts observations in six different forms: values of the time derivative of the magnetic field in micro-Volts, milli-Volts and/or Volts, and values of the magnetic field in nano-Tesla, micro-Tesla and/or milli-Tesla. For the step-off transmitter current waveform, the measurement times must be after the step off; for the linear ramp turn-off, the measurement times must be after the end of the ramp; but for the general, discretized transmitter current waveform the measurement times can be during the waveform's on-time as well as its off-time.

2.2 The Earth model

Program EM1DTM models the Earth beneath a measurement location as a stack of uniform, horizontal layers. It is the conductivities 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 can be juxtaposed to create a two- or three-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". In program EM1DTM, all data provided for a single transmitter loop constitute a single sounding, and a sounding can comprise only one transmitter loop. 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.

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 & Oldenburg (1996) and Farquharson, Oldenburg & Routh (2003). Computations are done in the frequency domain, then the fields transformed to the time domain. The fields for the horizontal, many-sided transmitter loop are computed by summing together the fields for horizontal electric dipoles joined end-to-end around the transmitter loop. The contributions to the magnetic fields from the grounding (or galvanic) terms associated with the exposed ends of each electric dipole are omitted as they would all cancel during the integration around the closed loop. The only contribution that need be considered is the inductive effect from the current flowing along the length

of the dipole. This means that, just as for program EM1DFM, only the z -component of the Schelkunoff \mathbf{F} -potential is required. The heart of the forward-modelling (and sensitivity-computing) algorithm in program EM1DTM is therefore essentially the same as that for program EM1DFM.

The Schelkunoff \mathbf{F} -potential is defined as follows (Ward & Hohmann, 1987):

$$\mathbf{E} = -\nabla \times \mathbf{F}, \quad (1)$$

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

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 j th 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, and the permeability of the layer is equal to that of free space μ_0):

$$\nabla^2 F_j - i\omega\mu_0\sigma_j F_j = 0. \quad (3)$$

Applying the two-dimensional Fourier transform to eq. (3) gives:

$$\frac{d^2 \tilde{F}_j}{dz^2} - u_j^2 \tilde{F}_j = 0, \quad (4)$$

where $u_j^2 = k_x^2 + k_y^2 + i\omega\mu_0\sigma_j$, and k_x and k_y are the horizontal wavenumbers. The solution to this equation is:

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

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:

$$\tilde{F}_{j-1}|_{z=z_j} = \tilde{F}_j|_{z=z_j}, \quad (6)$$

$$\left. \frac{d\tilde{F}_{j-1}}{dz} \right|_{z=z_j} = \left. \frac{d\tilde{F}_j}{dz} \right|_{z=z_j}. \quad (7)$$

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

$$\begin{pmatrix} e^{-u_{j-1}t_{j-1}} & e^{u_{j-1}t_{j-1}} \\ -u_{j-1}e^{-u_{j-1}t_{j-1}} & u_{j-1}e^{u_{j-1}t_{j-1}} \end{pmatrix} \begin{pmatrix} D_{j-1} \\ U_{j-1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -u_j & u_j \end{pmatrix} \begin{pmatrix} D_j \\ U_j \end{pmatrix}, \quad (8)$$

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:

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

where

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

for $j \geq 2$.

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

$$\tilde{F}_0 = D_0 e^{-u_0 z} + U_0 e^{u_0 z}, \quad (11)$$

which leads to

$$\begin{pmatrix} D_0 \\ U_0 \end{pmatrix} = \underline{\mathbf{M}}_1 \begin{pmatrix} D_1 \\ U_1 \end{pmatrix} \quad (12)$$

where

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

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:

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

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

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

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

$$\underline{\mathbf{P}} = \underline{\mathbf{M}}_1 \prod_{j=2}^M \underline{\mathbf{M}}_j, \quad (16)$$

and the factor E by

$$E = \exp\left(\sum_{j=2}^M u_{j-1} t_{j-1}\right). \quad (17)$$

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

$$D_M = \frac{1}{E} \frac{1}{P_{11}} D_0^S, \quad (18)$$

and the second gives

$$U_0 = E P_{21} D_M. \quad (19)$$

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

$$U_0 = \frac{P_{21}}{P_{11}} D_0^S, \quad (20)$$

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

$$\tilde{F}_0 = D_0^S \left(e^{-u_0 z} + \frac{P_{21}}{P_{11}} e^{u_0 z} \right). \quad (21)$$

For a horizontal x -directed electric dipole at a height h (i.e., $z = -h$, $h > 0$) above the surface of the layered Earth,

$$D_0^S = -\frac{i\omega\mu_0}{2u_0} \frac{ik_y}{k_x^2 + k_y^2} e^{-u_0 h} \quad (22)$$

(Ward & Hohmann, 1987, eq. 4.137). Substituting this into eq. (21), converting the required inverse two-dimensional Fourier transform to a Hankel transform, and applying eq. (2), gives the three components of

the complete H-field above the Earth model ($z < 0$) for the x -directed electric dipole (which is assumed to be at $x = 0, y = 0$):

$$H_x(x, y, z, \omega) = \frac{1}{4\pi} \frac{\partial}{\partial x} \frac{y}{r} \int_0^\infty \left(\pm e^{-\lambda|z+h|} + \frac{P_{21}}{P_{11}} e^{\lambda(z-h)} \right) J_1(\lambda r) d\lambda, \quad (23)$$

$$H_y(x, y, z, \omega) = \frac{1}{4\pi} \int_0^\infty \left(\pm e^{-\lambda|z+h|} + \frac{P_{21}}{P_{11}} e^{\lambda(z-h)} \right) \lambda J_0(\lambda r) d\lambda \\ - \frac{1}{4\pi} \frac{\partial}{\partial x} \frac{x}{r} \int_0^\infty \left(\pm e^{-\lambda|z+h|} + \frac{P_{21}}{P_{11}} e^{\lambda(z-h)} \right) J_1(\lambda r) d\lambda, \quad (24)$$

$$H_z(x, y, z, \omega) = \frac{1}{4\pi} \frac{y}{r} \int_0^\infty \left(e^{-\lambda|z+h|} + \frac{P_{21}}{P_{11}} e^{\lambda(z-h)} \right) \lambda J_1(\lambda r) d\lambda. \quad (25)$$

(The plus/minus is to do with whether or not the observation location is above or below the dipole: more below.) To obtain the fields for the many-sided, closed transmitter loop, the above expressions are integrated (with respect to the location of the dipole) around the closed transmitter loop. The terms above involving differentiation with respect to x are the galvanic contributions coming from the flow of current into and out of the ends of the dipole. These will cancel during the integration around the closed transmitter loop, and so can be dropped before that integration occurs. Hence, the relevant inductive contributions to the H-field for an x -directed electric dipole are:

$$H_x(x, y, z, \omega) = 0, \quad (26)$$

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

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

The Hankel transforms in eqs.(27) & (28) 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 and the minimum and maximum values, can be specified in the input file "em1dtm.in"; see "line 13" in the input file description (Section 3.1.1 of the Manual).

The integration of the fields for the dipole around the closed transmitter loop is done using Romberg integration (see, for example, Press et al. 1986). This is done in the frequency domain.

The first terms in the integrands of eqs. (27) & (28) are the contributions from the dipole in a wholespace, that is, the primary fields. The second terms are the contributions from currents induced in the layered Earth, that is, the secondary fields. In program EM1DTM, the primary field contributions, when required, are calculated directly, and so the Hankel transforms are only ever evaluated for the secondary fields.

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 Frequency- to time-domain transformation

To understand how the magnetic field or its time derivative is calculated for the various possible transmitter current waveforms in program EM1DTM, first consider the H-field for a step-off in the transmitter current waveform. Schematically:

$$I(t) = s(t) \quad \begin{array}{c} | \\ \text{---} \\ \text{---} \end{array} \quad \rightarrow \quad h_s(t) \quad \begin{array}{c} | \\ \text{---} \\ \text{---} \end{array}$$

where $s(t)$ is the unit step-off function, and $h_s(t)$ is the H-field for this step-off source. At the instant the transmitter current is switched off, there is a drop in the H-field. This is followed by the characteristic decay as the currents induced in the ground diffuse and decay. Let this decay of the H-field be denoted by $f(t)$. It's sense may or may not be the same as the primary H-field, and it may change sign, depending on the location of the receiver relative to the transmitter and its orientation. As the conductivity of the subsurface goes to zero, and so any currents that would be induced in the subsurface become vanishingly small, $f(t)$ also vanishes leaving an abrupt step-off in the H-field at the receiver.

Now suppose there is a step on in the transmitter current. Then:

$$I(t) = u(t) \quad \begin{array}{c} | \\ \text{---} \end{array} \rightarrow h_u(t) \quad \begin{array}{c} | \\ \text{---} \end{array}$$

where $u(t)$ is the Heaviside function (i.e., the unit step-on function), and $h_u(t)$ is the H-field for this source waveform. This H-field can be expressed using a Heaviside function and the decay of the H-field after the step-off transmitter waveform as:

$$h_u(t) = h_p u(t) (1 - f(t)), \quad (29)$$

where h_p is the wholespace H-field for a steady unit current in the transmitter loop. This expression is valid for all time.

The fundamental response required in order to calculate the responses for the possible waveforms in program EM1DTM is that for an impulse in the transmitter current waveform. This impulse is equal to the derivative with respect to time of the Heaviside function. Differentiating eq. (29) gives

$$G_h(t) = h_p \delta(t) (1 - f(t)) - h_p u(t) f'(t), \quad (30)$$

where $G_h(t)$ is the H-field Green's function, i.e., the H-field due to a delta-function source. Schematically:

$$I(t) = \delta(t) \quad \begin{array}{c} \uparrow \\ \text{---} \end{array} \rightarrow G_h(t) \quad \begin{array}{c} \uparrow \\ \text{---} \end{array}$$

This shows that the H-field Green's function comprises both an impulsive term, and a decaying term.

The time-domain H-field response for a general transmitter current waveform, $I(t)$, is obtained by the convolution:

$$h(t) = \int_{t'=-\infty}^{\infty} G_h(t-t') I(t') dt'. \quad (31)$$

Using the expression in eq. (30) for the H-field Green's function, the H-field for a general transmitter current waveform is given by:

$$h(t) = h_p (1 - f(0)) I(t) - h_p \int_{t'=-\infty}^t f'(t-t') I(t') dt'. \quad (32)$$

From the numerical perspective it is more convenient to work with the derivative of the piecewise-linear transmitter current waveform. Integrating the second term in eq. (32) by parts gives:

$$h(t) = h_p (1 - f(0)) I(t) + h_p f(0) I(t) - h_p \int_{t'=-\infty}^t f(t-t') I'(t') dt', \quad (33)$$

$$= h_p I(t) - h_p \int_{t'=-\infty}^t f(t-t') I'(t') dt'. \quad (34)$$

The time derivative of the H-field for a general transmitter current waveform $I(t)$, and hence the voltage in a receiver coil, is obtained by differentiating eq. (34):

$$h'(t) = h_p I'(t) - h_p f(0) I'(t) - h_p \int_{t'=-\infty}^t f'(t-t') I'(t') dt'. \quad (35)$$

From the argument above, it is the function $f(t)$ and its time derivative, which are the decaying parts of the H-field for a step-off and impulse respectively, and which do not include any delta-function term, that are effectively the Green's functions required to compute the responses for a general transmitter current waveform. The function $f(t)$ and its time derivative can be obtained from the frequency-domain expression for the H-field via cosine and sine transforms (see Newman, Hohmann & Anderson, 1986):

$$f(t) = -\frac{2}{\pi} \int_0^\infty \text{Im } H(\omega) \frac{1}{\omega} \cos \omega t \, d\omega. \quad (36)$$

$$f'(t) = \frac{2}{\pi} \int_0^\infty \text{Im } H(\omega) \sin \omega t \, d\omega, \quad (37)$$

The above transforms are evaluated using the digital filtering algorithms of Anderson which are analogous to those for evaluating the Hankel transforms (Anderson, 1982). The convolutions required to compute the response for a linear ramp turn-off and for a general, discretized transmitter current waveform are done directly.

2.5 Computing the sensitivities

The inverse problem of determining the conductivity of the Earth from electromagnetic measurements is nonlinear. Program EM1DTM 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, \mathbf{J} , where:

$$J_{ij} = \frac{\partial d_i}{\partial \log \sigma_j}, \quad (38)$$

in which d_i is the i th observation, and σ_j is the conductivity of the j th layer.

The algorithm for computing the sensitivities is obtained by differentiating the expressions for the relevant parts of the H-field in the frequency domain (eqs. 26–28, Section 2.3) with respect to the model parameters. This is analogous to the case for program EM1DFM: see Farquharson, Oldenburg & Routh (2003). For example, the sensitivity with respect to m_j (i.e., the logarithm of the conductivity of the j th layer) of the z -component of the inductive part of the H-field for an x -directed electric dipole source is given by differentiating eq. (28):

$$\frac{\partial H_z(x, y, z, \omega)}{\partial m_j} = \frac{1}{4\pi} \frac{y}{r} \int_0^\infty \left(e^{-\lambda|z+h|} + \frac{\partial}{\partial m_j} \left[\frac{P_{21}}{P_{11}} \right] e^{\lambda(z-h)} \right) \lambda J_1(\lambda r) \, d\lambda. \quad (39)$$

The derivative of the coefficient is simply

$$\frac{\partial}{\partial m_j} \left[\frac{P_{21}}{P_{11}} \right] = \frac{\partial P_{21}}{\partial m_j} \frac{1}{P_{11}} - \frac{P_{21}}{P_{11}^2} \frac{\partial P_{11}}{\partial m_j}, \quad (40)$$

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

$$\frac{\partial \mathbf{P}}{\partial m_j} = \mathbf{M}_1 \mathbf{M}_2 \cdots \mathbf{M}_{j-1} \left(\frac{\partial \mathbf{M}_j}{\partial m_j} \mathbf{M}_{j+1} + \mathbf{M}_j \frac{\partial \mathbf{M}_{j+1}}{\partial m_j} \right) \mathbf{M}_{j+2} \cdots \mathbf{M}_M. \quad (41)$$

The sensitivity with respect to the conductivity of the basement halfspace is given by

$$\frac{\partial \mathbf{P}}{\partial m_M} = \mathbf{M}_1 \mathbf{M}_2 \cdots \mathbf{M}_{M-1} \frac{\partial \mathbf{M}_M}{\partial m_M}. \quad (42)$$

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

The same integrations and transformations that are used in the forward modelling process to get from eqs. (26)–(28) to the time-domain responses for the many-sided transmitter loop and for the particular transmitter current waveform are then used to get from eq. (39), and the corresponding one for the sensitivity in the y -component, to the required time-domain sensitivities for the large-loop data.

The partial propagation matrices

$$\underline{\mathbf{P}}_k = \underline{\mathbf{M}}_1 \prod_{j=2}^k \underline{\mathbf{M}}_j, \quad k = 2, \dots, M \quad (43)$$

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.

2.6 The inversion algorithms

Just as for program EM1DFM, there are four different variations of the one general inversion algorithm in program EM1DTM. The variations differ in their treatment of the trade-off parameter. The general formulation is described in Section 2.6.1 below, and the different ways of treating the trade-off parameter are described in Sections 2.6.4 to 2.6.7. Unlike program EM1DFM, general measures of misfit and model structure are used in program EM1DTM. These are described in Section 2.6.2.

2.6.1 General formulation

The aim of all the inversion algorithms 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:

$$\Phi = \phi_d + \beta \phi_m. \quad (44)$$

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

$$\phi_d = M_d(\underline{\mathbf{W}}_d(\mathbf{d} - \mathbf{d}^{\text{obs}})), \quad (45)$$

where \mathbf{d}^{obs} is the vector containing the N observations, \mathbf{d} is the forward-modelled data, and $M_d(\mathbf{x})$ is some measure of the “length” of a vector \mathbf{x} (more on this in Section 2.6.2). It is assumed that the noise in the observations is uncorrelated, and that the estimated standard deviation of the noise in the i th observation is of the form $s_0 \hat{s}_i$, where \hat{s}_i indicates the amount of noise in the i th 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:

$$\underline{\mathbf{W}}_d = \text{diag}\{1/(s_0 \hat{s}_1), \dots, 1/(s_0 \hat{s}_N)\}. \quad (46)$$

The model-structure component of the objective function is ϕ_m :

$$\phi_m = \alpha_s M_m^s(\underline{\mathbf{W}}_s(\mathbf{m} - \mathbf{m}_s^{\text{ref}})) + \alpha_z M_m^z(\underline{\mathbf{W}}_z(\mathbf{m} - \mathbf{m}_z^{\text{ref}})), \quad (47)$$

where \mathbf{m} is the vector containing the logarithms of the layer conductivities, and $M_m^s(\mathbf{x})$ and $M_m^z(\mathbf{x})$ are measures of a length of the vector \mathbf{x} (more in Section 2.6.2). The matrix $\underline{\mathbf{W}}_s$ is:

$$\underline{\mathbf{W}}_s = \text{diag}\{\sqrt{t_1}, \dots, \sqrt{t_{M-1}}, \sqrt{t_{M-1}}\}, \quad (48)$$

the true l_p -norm suffers. In program EM1DTM, the values of p can be different for the smallest and flattest components of ϕ_m .

2.6.3 General algorithm

As mentioned in Section 2.5, the inverse problem considered here is nonlinear. It is solved using an iterative procedure. At the n th iteration, the actual objective function being minimized is:

$$\Phi^n = \phi_d^n + \beta^n \phi_m^n. \quad (54)$$

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:

$$\mathbf{d}^n = \mathbf{d}^{n-1} + \underline{\mathbf{J}}^{n-1} \delta \mathbf{m}, \quad (55)$$

where $\delta \mathbf{m} = \mathbf{m}^n - \mathbf{m}^{n-1}$, and $\underline{\mathbf{J}}^{n-1}$ is the Jacobian matrix given by eq. (38) and evaluated for the model from the previous iteration. At the n th iteration, the problem to be solved is that of finding the change, $\delta \mathbf{m}$, to the model which minimizes the objective function Φ^n . Differentiating eq. (54) with respect to the components of $\delta \mathbf{m}$, and equating the resulting expressions to zero, gives the system of equations to be solved. This is slightly more involved now that ϕ_d and ϕ_m comprise Huber's M -measure (eqs. 50–51) and Ekblom's l_p -like measure (eqs. 52–53), rather than the usual sum-of-squares measures. Specifically, the derivative of eq. (50) gives:

$$\frac{\partial M_d}{\partial \delta m_k}(\mathbf{x}) = \sum_{i=1}^N \rho'_H(x_i) \frac{\partial x_i}{\delta m_k}, \quad (56)$$

where $\mathbf{x} = \underline{\mathbf{W}}_d(\mathbf{d}^{n-1} + \underline{\mathbf{J}}^{n-1} \delta \mathbf{m} - \mathbf{d}^{\text{obs}})$. The vector of the derivatives with respect to the perturbations of the model parameters in all the layers can be written in matrix notation as:

$$\frac{\partial M_d}{\partial \delta \mathbf{m}_k} = -\underline{\mathbf{J}}^T \underline{\mathbf{W}}_d^T \underline{\mathbf{R}}_d \mathbf{x}, \quad (57)$$

where $\underline{\mathbf{R}}_d = \text{diag}\{\rho'_H(x_1)/x_1, \dots, \rho'_H(x_N)/x_N\}$, and the superscript $n-1$ has been dropped for the Jacobian matrix. Similarly for the Ekblom l_p -like measure:

$$\frac{\partial M_m}{\partial \delta \mathbf{m}_k} = -\underline{\mathbf{W}}_m^T \underline{\mathbf{R}}_m \mathbf{x}, \quad (58)$$

where $\mathbf{x} = \underline{\mathbf{W}}_m(\mathbf{m}^{n-1} + \delta \mathbf{m} - \mathbf{m}^{\text{ref}})$, and $\underline{\mathbf{R}}_m = \text{diag}\{\rho'_E(x_1)/x_1, \dots, \rho'_E(x_N)/x_N\}$. The linear system of equations to be solved at each iteration in the inversion algorithm therefore becomes:

$$\left[\underline{\mathbf{J}}^T \underline{\mathbf{W}}_d^T \underline{\mathbf{R}}_d \underline{\mathbf{W}}_d \underline{\mathbf{J}} + \beta^n \sum_{i=1}^2 \alpha_i \underline{\mathbf{W}}_i^T \underline{\mathbf{R}}_i \underline{\mathbf{W}}_i \right] \delta \mathbf{m} = \underline{\mathbf{J}}^T \underline{\mathbf{W}}_d^T \underline{\mathbf{R}}_d \underline{\mathbf{W}}_d (\mathbf{d}^{\text{obs}} - \mathbf{d}^{n-1}) + \beta^n \sum_{i=1}^2 \alpha_i \underline{\mathbf{W}}_i^T \underline{\mathbf{R}}_i \underline{\mathbf{W}}_i (\mathbf{m}_i^{\text{ref}} - \mathbf{m}^{n-1}), \quad (59)$$

where the subscript i for the model terms represents s and z . The presence of the diagonal weighting matrices $\underline{\mathbf{R}}$ in eq. (59), which results from the way the derivatives of the Huber and Ekblom measures are expressed in eqs. (57) & (58), makes this what is known as the iteratively reweighted least squares (IRLS) algorithm in the context of linear inverse problems (see, for example, Farquharson & Oldenburg, 1998). These weighting matrices are updated at each iteration, just as the Jacobian matrix is updated.

Once the step $\delta \mathbf{m}$ has been determined by the solution of eq. (59), the new model is given by:

$$\mathbf{m}^n = \mathbf{m}^{n-1} + \nu \delta \mathbf{m}. \quad (60)$$

The condition on the step length ν is that the objective function must be decreased by the addition of the step to the model:

$$\phi_d^n + \beta^n \phi_m^n < \phi_d^{n-1} + \beta^n \phi_m^{n-1}, \quad (61)$$

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 1 is tried. If eq. (61) is true for this step length, it is accepted. If eq. (61) is not true, ν is decreased by factors of 2 until it is true.

2.6.4 Algorithm 1: user-supplied cooling schedule

There are now two slightly different versions of this algorithm. Just as for program EM1DFM, it is possible to supply a single value for the trade-off parameter, β , which is used for all iterations in the inversion procedure. However, it is now possible to also supply an initial value for the trade-off parameter and a factor by which the trade-off parameter is to be reduced from one iteration to the next. Once the supplied final value is reached, the algorithm will continue using this value until the convergence criteria are met (see Section 2.6.8). This second, cooling-schedule version is more robust than simply using the final value of the trade-off parameter throughout the iterative procedure: a value that is appropriate at the conclusion of an inversion can often be too small for the initial iterations, causing too much and possibly incorrect structure to appear too early in the model, which can then hard to undo. The solution to eq. (59) is computed via LU decomposition using the subroutine “SGESV” from LAPACK (Anderson et al., 1999). If the desired final 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 an 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.

2.6.5 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 ($i = 1, \dots, N$) are known, and the measure of misfit is the traditional sum-of-squares measure, 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` $\times N$, where `chifac` is supplied by the user. If the noise in the observations is well known, and the c parameter in the Huber M -measure is large (so that the misfit is essentially a sum-of-squares measure), `chifac` should equal 1. However, `chifac` can be adjusted by the user to give a target misfit appropriate for a particular data-set, and for a particular choice of the Huber M -measure parameter c . (For example, the expectation of an l_1 measure of misfit, and so the Huber M -measure with c small, is $(2/\pi)^{1/2}N$; Parker & McNutt, 1980.) 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 EM1DTM, the target misfit at the n th iteration is given by

$$\phi_d^{n,\text{tar}} = \max(\text{mfac} \times \phi_d^{n-1}, \text{chifac} \times N), \quad (62)$$

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

The step $\delta\mathbf{m}$ is found from the solution of eq. (59) using subroutine `SGESV` from LAPACK (Anderson et al., 1999). The line search at each iteration moves along the ϕ_d versus $\log \beta$ curve until either the target misfit, $\phi_d^{n,\text{tar}}$, is bracketed, in which case a bisection search is used to converge to the target, or the minimum misfit ($> \phi_d^{n,\text{tar}}$) 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. (Specifically, \mathbf{m}^\dagger is a model whose top $M/5$ layers have a conductivity of 0.02 S/m, and

whose remaining layers have a conductivity of 0.01 S/m. Also, the reference models used in the computation of $\phi_m(\mathbf{m}^\dagger)$ are homogeneous halfspaces of 0.01 S/m.) The line search is efficient, but does involve the full forward modelling to compute the misfit for each trial value of β .

2.6.6 Algorithm 3: GCV criterion

If only the relative amount of noise in the observations is known, that is, \hat{s}_i ($i = 1, \dots, N$) 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, 2003; Farquharson & Oldenburg, 2004). From eq. (59), the GCV function for the n th iteration is given by:

$$GCV(\beta) = \frac{\|\mathbf{R}_d^{1/2} \mathbf{W}_d \hat{\mathbf{d}} - \mathbf{R}_d^{1/2} \mathbf{W}_d \mathbf{J} \mathbf{M}^{-1} (\mathbf{J}^T \mathbf{W}_d^T \mathbf{R}_d \mathbf{W}_d \hat{\mathbf{d}} + \mathbf{r})\|^2}{[\text{trace}(\mathbf{I} - \mathbf{R}_d^{1/2} \mathbf{W}_d \mathbf{J}^{n-1} \mathbf{M}^{-1} \mathbf{J}^{n-1T} \mathbf{W}_d^T \mathbf{R}_d^{1/2})]^2}, \quad (63)$$

where:

$$\mathbf{M}(\beta) = \left[\mathbf{J}^T \mathbf{W}_d^T \mathbf{R}_d \mathbf{W}_d \mathbf{J} + \beta \sum_{i=1}^2 \alpha_i \mathbf{W}_i^T \mathbf{R}_i \mathbf{W}_i \right], \quad (64)$$

$$\mathbf{r} = \beta \sum_{i=1}^2 \alpha_i \mathbf{W}_i^T \mathbf{R}_i \mathbf{W}_i (\mathbf{m}_i^{\text{ref}} - \mathbf{m}^{n-1}), \quad (65)$$

and $\hat{\mathbf{d}} = \mathbf{d}^{\text{obs}} - \mathbf{d}^{n-1}$. If β^* is the value of the trade-off parameter that minimizes eq. (63) at the n th iteration, the actual value of β used to compute the new model is given by:

$$\beta^n = \max(\beta^*, \text{bfac} \times \beta^{n-1}), \quad (66)$$

where the user-supplied factor **bfac** is such that $0.01 < \text{bfac} < 0.5$. 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 \mathbf{M} required in eq. (63), and the solution to eq. (59) 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 $\text{bfac} \times \beta^{n-1}$ 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 \mathbf{M} .

The statistical arguments on which the GCV criterion is based assume that the measures used are sum-of-squares measures. However, during the course of a steadily converging inversion, it is anticipated that the linearized problem at each iteration is sufficiently quadratic, even when the Huber M -measure and Ekblom's l_p -like measure differ considerably from the sum-of-squares measure, that the GCV criterion will give a useful value for the trade-off parameter. This has proved to be the case in the testing done to date.

2.6.7 Algorithm 4: L-curve criterion

As for the GCV-based method described in Section 2.6.6, the L-curve method provides a means of estimating an appropriate value of the trade-off parameter if only \hat{s}_i , $i = 1, \dots, N$, 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, 2004). In this situation, the L-curve is defined using the linearized misfit, which uses the approximation given in eq. (55) for the forward-modelled data. The curvature of the L-curve is computed using the formula (Hansen, 1998):

$$C(\beta) = \frac{\zeta'\eta'' - \zeta''\eta'}{((\zeta')^2 + (\eta')^2)^{3/2}}, \quad (67)$$

where $\zeta = \log \phi_d^{\text{lin}}$ 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 n th iteration is given by eq. (66), where β^* 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. (67) prone to numerical noise. The line search along the L-curve used in program EM1DTM 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. (67) 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 β^* . 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 solution to eq. (59) is computed using subroutine SGESV from LAPACK (Anderson et al., 1999).

2.6.8 Convergence criteria

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

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

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

$$\|\mathbf{g}^n\| < \sqrt[3]{\tau}(1 + \Phi^n), \quad (70)$$

where τ is a user-specified parameter, and the gradient for the n th iteration is given by:

$$\mathbf{g}^n = -2\mathbf{J}^n \mathbf{W}_d^T \mathbf{R}_d \mathbf{W}_d (\mathbf{d}^{\text{obs}} - \mathbf{d}^n) - 2\beta^n \sum_{i=1}^2 \alpha_i \mathbf{W}_i^T \mathbf{R}_i \mathbf{W}_i (\mathbf{m}_i^{\text{ref}} - \mathbf{m}^n). \quad (71)$$

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:

$$\|\mathbf{g}^n\| \leq \epsilon, \quad (72)$$

where ϵ is a small number close to zero.

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, 1993. Inversion of time-domain electromagnetic data for a horizontally layered Earth, *Geophysical Journal International*, 114, 433–442.
- Farquharson, C.G., & D.W. Oldenburg, 1996. Approximate sensitivities for the electromagnetic inverse problem, *Geophysical Journal International*, 126, 235–252.
- Farquharson, C.G., & D.W. Oldenburg, 1998. Nonlinear inversion using general measures of data misfit and model structure, *Geophysical Journal International*, 134, 213–227.
- Farquharson, C.G., & D.W. Oldenburg, 2004. A comparison of automatic techniques for estimating the regularization parameter in nonlinear inverse problems, *Geophysical Journal International*, 156, 411–425.
- Farquharson, C.G., D.W. Oldenburg & P.S. Routh, 2003. Simultaneous one-dimensional inversion of loop-loop electromagnetic data for magnetic susceptibility and electrical conductivity, *Geophysics*, 68, 1857–1869.
- 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, 2003. Fast inversion of large-scale magnetic data using wavelet transforms and a logarithmic barrier method, *Geophysical Journal International*, 152, 251–265.
- Nabighian, M.N., & J.C. Macnae, 1987. Time Domain Electromagnetic Prospecting Systems, in *Electromagnetic Methods in Applied Geophysics*, vol. 2, M.N. Nabighian (ed.), 427–520, Society of Exploration Geophysicists, Tulsa, Oklahoma.
- Newman, G.A., G.W. Hohmann, & W.L. Anderson, 1986. Transient electromagnetic response of a three-dimensional body in a layered Earth, *Geophysics*, 51, 1608–1627.
- Parker, R.L., & M.K. McNutt, 1980. Statistics for the one-norm misfit measure, *Journal of Geophysical Research*, 85, 4429–4430.
- Press, W.H., B.P. Flannery, S.A. Teukolsky & W.T. Vetterling, 1986. *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Cambridge.
- 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.