Finite element based inversion of AEM data using stochastic optimization

Christoph Schwarzbach
University of British Columbia
Department of Earth and Ocean Sciences
6339 Stores Road
Vancouver, BC, V6T 1Z4
Canada
cschwarz@eos.ubc.ca

Eldad Haber
University of British Columbia
Department of Earth and Ocean Sciences
6339 Stores Road
Vancouver, BC, V6T 1Z4
Canada
haber@math.ubc.ca

SUMMARY

Modern data acquisition technology in airborne electromagnetics (AEM) produces huge data sets which cover areas of considerable extend. Both the large size of the domain of interest, in general a three-dimensional volume, and the large number of transmitters and receivers pose challenges to any type of modelling or inversion software. Solution of the inverse problem requires repeated solves of the forward problem. The time to solve one forward problem in turn scales linearly with the number of transmitters. In this paper we examine stochastic optimization techniques for the solution of the inverse problem which essentially allow us to work with small subsets of transmitters/receivers and, thus, reduce the computational load significantly.

Key words: electromagnetics, modelling, inversion, finite elements

INTRODUCTION

In recent years significant advances have been made in our ability to model and invert electromagnetic data (Newman and Alumbaugh, 1997a, b; Haber *et al.*, 2004; Newman and Alumbaugh, 2000; Zhdanov and Fang, 1996; Mackie *et al.*, 1993; Schwarzbach *et al.*, 2011). Most forward modeling and inversion codes rely on a combination of staggered grids finite difference/volume methods in order to model the electric/magnetic field and the regularized output least squares approach in order to invert the data. While this approach has worked for synthetic and field data sets it is computationally expensive. In the bottleneck of the computations stands the solution of the forward problem, that is, the solution of the discretized system of Maxwell's equations in space and frequency/time. Upon discretization of the forward problem a large sparse and ill-conditioned matrix is generated. For realistic problems the size of the matrix can easily be in the order of millions of degrees of freedom. Recent data collection has put forward another great challenge. Since the number of sources and receivers is very large one has to solve the forward problem many times. Furthermore, for the computation of the Gauss-Newton step, the fields for each source have to be saved. This constitutes a major difficulty when the number of sources and receivers is large as computer memory needs to be allocated. As a result, solving problems where the number of sources/receivers is in the thousands is

done only when the significant computational resources are available. In this paper we experiment with a number of computational techniques to overcome these difficulties.

First, in order to reduce the number of degrees of freedom of the forward problem we use a finite element formulation that allows the use of semi-structured (OcTree) and unstructured grids. Second, in order to solve the problem with many sources/receivers we use stochastic optimization techniques that allow us to work only with a subset of source/receiver combinations at each iteration. A brief outline of these ideas is given in the following section. We will conclude this extended abstract with a small numerical example and a summary.

METHOD

Forward problem

For the forward problem we consider Maxwell's equations in frequency domain

$$\nabla \times \mu^{-1} \nabla \times \boldsymbol{E}_j - i\omega \sigma \boldsymbol{E}_j = i\omega \boldsymbol{q}_j, \qquad (1)$$

$j = 1, ..., n_s$, accompanied with Neumann boundary conditions. Here, $\boldsymbol{E}_j$ denotes the electric field, σ is the (possibly complex) electrical conductivity, μ is the magnetic permeability, ω is the angular frequency and $\boldsymbol{q}_j$ is the source. We assume that there are n_s sources which give rise to n_s electric fields.

We solve problem (1) employing an edge finite element discretization on a tetrahedral mesh (Jin, 1993; Monk, 2003; Schwarzbach *et al.*, 2011). The use of semi-structured (OcTree) and unstructured meshes enables us to easily deal with bathymetry as well as with the rapid decay of the fields away from the sources without generating a fine mesh everywhere. Upon discretization we obtain a system of linear equations

$$A(m)U = Q \qquad (2)$$

where the jth column of Q and U respectively contain the source contribution and the electric field degrees of freedom of the jth transmitter. In view of the inverse problem, the system matrix $A(m)$ is assumed to be parameterized by the model vector m whose k^{th} entry corresponds to the log conductivity of the kth element, this is, $\sigma = e^m$. Equation (2) is solved with the parallel, sparse direct solver PARDISO (Schenk and Gärtner, 2004, 2006). A single factorization of

the system matrix allows for the solution of many right hand sides, the computation of the gradient and the computation of the Gauss-Newton step (Pratt, 1999; Haber *et al.*, 2000).

We assume that the model–data relation is given by a matrix function

$$D(m) = P^T A(m)^{-1} Q \qquad (3)$$

where P is a matrix which interpolates the electric field degrees of freedom $U(m) = A(m)^{-1} Q$ into the receiver location. The $(i, j)^{th}$ entry of data matrix $D(m)$ is the measured datum that corresponds to the ith receiver and the jth source.

Inverse problem

Assume that we observe the data matrix D^{obs}. Our goal is to find a "reasonable" model m that fits this data up to some tolerance. Using the regularized output least squares approach we seek a model that minimizes a functional of the form

$$F(m) = 0.5 \parallel W \bullet (D(m) - D^{obs}) \parallel^2_F \\ + 0.5 \, \alpha \parallel G \, (m - m_{ref}) \parallel^2 \qquad (4)$$

where $\bullet$ denotes the Hadamard product, $\parallel \parallel_F$ denotes the Frobenius norm and α is a regularization parameter. The dense matrix W contains the inverse of the standard deviation of each datum. The matrix G is chosen here as a (generalized) derivation operator which is applied to promote smoothness of deviations between the model m and a reference model m_{ref}.

To minimize the functional $F(m)$ we consider two techniques. First, if possible, we apply the inexact Gauss-Newton method (Nocedal and Wright, 1999; Haber *et al.*, 2004) to the first derivative of $F(m)$ with respect to m

$$\nabla_m F(m) = g(m) = 0. \qquad (5)$$

That is, at each iteration we (approximately) solve the linear system

$$(J(m)^T J(m) + \alpha \, G^T G) \, \delta m = -g(m) \qquad (6)$$

and then update m by $m + \nu \, \delta m$ where $\nu \leq 1$ is a parameter chosen to ensure that the value of the objective function decreases (line search). $J(m) = \nabla_m D(m)$ denotes the sensitivity matrix. There is no need to compute the sensitivity matrix $J(m)$ directly. To solve equation (6), we use the preconditioned conjugate gradient (PCG) solver (Barret *et al.*) with the Hessian of the regularization operator as a preconditioner. We only require matrix vector products with $J(m)$ and $J(m)^T$. $J(m)$ can be expressed as a product of sparse matrices and their inverses; consequently, the matrix vector products essentially reduce to solving the forward and adjoint problems.

While the Gauss-Newton method has been one of the most popular methods of choice for most least squares problems it is difficult to apply when the number of sources is very large. The difficulty stems from the need to store the fields that are associated with all sources. For a large number of sources such storage is impossible. In this case we use a more "modest" optimization algorithm, the limited memory BFGS (L-BFGS) algorithm (Nocedal and Wright, 1999). L-BFGS is a quasi-Newton method that requires only the gradient and the step update of the previous L steps. Since computing the gradient can be done sequentially, it does not require the

storage of all fields. However, the efficiency in storage comes with a price as convergence of L-BFGS is significantly slower compared with the convergence of the Gauss-Newton method.

Stochastic optimization

The solution of (6) with a large number of sources is very time consuming since the evaluation of the data misfit scales linearly with the number of sources (Table 1).

	LDL^T-factors	solves
data misfit	1	n_s
gradient		n_s
n_{cg} PCG steps		$2 \, n_{cg} n_s$
n_{ls} linesearch steps	n_{ls}	$n_{ls} \, n_s$
total	$1 + n_{ls}$	$(2 + 2 \, n_{cg} + n_{ls}) \, n_s$

Table 1. Cost of forward solution of one Gauss-Newton step. Note that we store the matrix factors L and D and reuse them for the computation of the gradient and of the search direction (PCG).

For the computation of the gradient one requires $2n_s$ solutions of Maxwell's equations as well as the storage of these fields. For large scale problems the computation and the storage requires excessive computational resources. In recent years, problems with multiple sources have been studied (Krebs *et al.*, 2009; Haber *et al.*, 2010) and efficient techniques that are based on stochastic optimization have been proposed for the solution. Here we explore and experiment with a variant of stochastic optimization. To use stochastic optimization we introduce the random variable ξ that takes the values 1, ..., n_s and rewrite the problem in its stochastic form

$$F(m) = 0.5 \, \mathbf{E} \parallel W_{:,\xi} \bullet (D_{:,\xi}(m) - D_{:,\xi}{}^{obs}) \parallel^2 \\ + 0.5 \, \alpha \parallel G \, (m - m_{ref}) \parallel^2 \qquad (7)$$

where $\mathbf{E}$ is the expected value and $W_{:,\xi}$ denotes the ξth column of W. This is a classical formulation of stochastic optimization (Shapiro *et al.*, 2009; Juditsky *et al.*, 2009). Such problems have been extensively studied in the optimization literature. In particular, we make use of a stochastic optimization algorithm that has a very simple and remarkable property. At each step it requires only a single (or a small) number of samples from ξ. It has been shown in Shapiro *et al.* (2009) and Juditsky *et al.* (2009) that if the samples are replaced at each iteration and if the stepsize is damped appropriately then such algorithms are guaranteed to converge. The obvious advantage of such an algorithm is that at each iteration only a (random) subset of the sources is treated and therefore it requires a much smaller storage compared with standard optimization techniques. The use of a small number of sources (at each iteration) also allows for the incorporation of the Gauss-Newton method thus gaining on the L-BFGS method that does not use explicit information about the sensitivities.

RESULTS

We illustrate the three approaches discussed above by an example problem. The experimental setup comprises 128 transmitters and 1,920 receivers, this is, a data set of 245,760 data points. We seek to reconstruct a model vector m with 109,065 entries. The full data set with all 128 sources is

inverted using (a) the classical Gauss-Newton method, (b) L-BFGS with storage of 20 vectors for the approximation of the inverse Hessian, and (c) stochastic optimization. For the Gauss-Newton method, we stop the PCG solver [equation (6)] when the tolerance is 10^{-1} or after 20 iterations. For the stochastic inversion we utilize only 6.25% of the full data set, chosen randomly at each iteration. 8/128 sources are chosen at random and 3 Gauss-Newton steps are performed for these 8 sources. The next iteration is started on the final model of the previous step but with another instance of randomly chosen sources. This procedure is repeated here to make a total of 16 random realizations. Using the same regularization parameter, the deterministic variants (a) and (b) reduce the initial data misfit to about 3.3% and the stochastic variant (c) to about 7.7%. The predicted models obtained by the three optimization variants differ only slightly.

	GN	L-BFGS	S-GN
iterations	20	100	16 x 3
LDL^T-factors	21	132	66
solves	118,016	33,792	14,544

Table 2. Iteration numbers, number of matrix factorisations and backward/forward solves for three optimization strategies: Gauss-Newton (GN) and L-BFGS inversion of the full data set with 128 sources. Stochastic inversion (S-GN) with data subsets of 8 randomly chosen sources, 16 realisations, 3 Gauss-Newton steps each.

Table 2 summarizes the major computational load of the inversion, that is, computing the LDL^T-factors of the system matrix $A(m)$ and carrying out forward/backward solves for a large number of right hand sides. Due to its faster convergence, the Gauss-Newton method requires less iterations and, consequently, less LDL^T-factors than L-BFGS. However, the computation of the Gauss-Newton model update requires the solution of a linear system with the Hessian matrix. Even a moderate number of PCG steps leads to a substantial amount of backward/forward solves. Despite the slightly higher data misfit the stochastic inversion strategy proves to be the most efficient of the three methods.

CONCLUSIONS

In this work we have explored the solution of large scale electromagnetic inverse problems. We have used a combination of techniques to obtain an efficient and robust code. The number of unknowns in the forward problem is reduced by employing a finite element discretization of Maxwell's equations on semi-structured or unstructured tetrahedral meshes. Stochastic optimization techniques reduce the number of forward problems to be solved. Combining these techniques we are able to solve electromagnetic inverse problems that are traditionally solved using intensive computational platforms on modest ones. We anticipate even bigger improvements for problems where iterative methods are needed for the solution of the forward problem.

REFERENCES

Barrett, R., Berry, M., Chan, T., Demmel, J., , Donato, J., Dongarra, J., Eijkhout, V., Pozo, R., Romine, C., and Van der Vorst, H., 1994, Templates for the solution of linear systems: Building blocks for iterative methods: SIAM.

Haber, E., Ascher, U., and Oldenburg, D., 2000, On optimization techniques for solving nonlinear inverse problems: Inverse Problems, 16, 1263–1280.

Haber, E., Ascher, U., and Oldenburg, D., 2004, Inversion of 3D electromagnetic data in frequency and time domain using an inexact all-at-once approach: Geophysics, 69, 1216–1228.

Haber, E., Chung, M., and Herrmann, F., 2010, Solving inverse problems with many right hand sides: Submitted, 1263–1280.

Jin, J., 1993, The finite element method in electromagnetics: John Wiley and Sons.

Juditsky, A., Lan, G., Nemirovski, A., and Shapiro, A., 2009, Stochastic approximation approach to stochastic programming: SIAM Journal on Optimization, 19, 1574–1609.

Krebs, J. R., Anderson, J. E., Hinkley, D., Neelamani, R., Lee, S., Baumstein, A., and Lacasse, M.-D., 2009, Fast full-wavefield seismic inversion using encoded sources: Geophysics, 74, WCC177–WCC188.

Mackie, R., Madden, T., and Wannamaker, P., 1993, Three dimensional magnetotelluric modeling using difference equations – Theory and comparison to integral equation solutions: Geophysics, 58, 215–226.

Monk, P., 2003, Finite element methods for Maxwell's equations: Oxford University Press.

Newman, G., and Alumbaugh, D., 1997a, Three-dimensional massively parallel electromagnetic inversion—I. Theory: Geophysical Journal International, 128, 345–354.

Newman, G., and Alumbaugh, D., 1997b, Three-dimensional massively parallel electromagnetic inversion—II. Analysis of a crosswell electromagnetic experiment: Geophysical Journal International, 128, 355–367.

Newman, G. A., and Alumbaugh, D. L., 2000, Threedimensional magnetotelluric inversion using nonlinear conjugate gradients: Geophysical Journal International, 140, 410–418.

Nocedal, J., and Wright, S., 1999, Numerical optimization: Springer.

Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, Part 1: Theory, and verification in a physical scale model: Geophysics, 64, 888–901.

Schenk, O., and Gärtner, K., 2004, Solving unsymmetric sparse systems of linear equations with PARDISO: Journal of Future Generation Computer Systems, 20, 475–487.

Schenk, O., and Gärtner, K., 2006, On fast factorization pivoting methods for symmetric indefinite systems: Electronic Transactions on Numerical Analysis, 23, 158–179.

Schwarzbach, C., Börner, R.-U., and Spitzer, K., 2011, Three-dimensional adaptive higher order finite element simulation for geo-electromagnetics—a marine CSEM example: Geophysical Journal International, to appear.

Shapiro, A., Dentcheva, D., and Ruszczynski, D., 2009, Lectures on stochastic programming: Modeling and theory: SIAM.

Zhdanov, M., and Fang, S., 1996, Three-dimensional quasilinear electromagnetic inversion: Radio Science, 31, 741–748.