Full-Jacobian Gauss-Newton 3D marine magnetotelluric inversion

Sébastien de la Kethulle de Ryhove and Rune Mittet*, EMGS

Summary

In the past, when solving 3D magnetotelluric (MT) inverse problems, computing the Jacobian (sensitivity matrix) at each inversion iteration has been avoided due to the relatively large memory and complexity requirements. We have developed a model-space Gauss-Newton 3D marine MT inversion scheme where the full Jacobian is computed at each inversion iteration. The memory and complexity requirements are sufficiently modest to allow the inversion of 3D data sets on a small computer cluster. For a real data example from the Barents Sea, our scheme successfully converged starting from a homogeneous half-space model. This is in contrast to a previous quasi-Newton 3D MT inversion scheme we developed, which for the same data set failed to converge starting from homogeneous half-space models. In addition, testing revealed noticeable improvements in terms of the required number of inversion iterations for convergence and receiver-imprint related artefacts and when going from a quasi-Newton based scheme to our new Gauss-Newton inversion scheme.

Introduction

Magnetotelluric (MT) data can be a valuable complement to seismic data in geologic environments where seismic imaging is challenging; e.g. below highly heterogeneous basalt layers or in the presence of complex salt structures. In such situations, resistivity models obtained via MT inversion can for example be used to improve velocity models for seismic processing or in workflows for joint interpretation (Hovertsen et al., 1998; Key et al., 2006; Avdeeva et al., 2012).

Consider the problem of inverting an MT data set consisting of \( N \) data points assuming the earth has been divided into \( M \) rectangular cells with conductivities \( \sigma_1, \ldots, \sigma_M \). Assembling the problem’s \( N \times M \) Jacobian (or sensitivity) matrix \( J \) comes at a relatively large computational cost. As a consequence, in the literature on MT inversion written during the last few decades, significant attention has been devoted to gradient-type inversion schemes which altogether avoid assembling the Jacobian such as the conjugate gradient method (Mackie et al., 1993a), the non-linear conjugate gradient (NLCG) method (Newman and Alumbaugh, 2000) and quasi-Newton methods (Avdeev and Avdeeva, 2009; de la Kethulle de Ryhove and Mittet, 2014). For inversion schemes based on the Gauss-Newton method, which do require knowledge of the Jacobian, focus has been on computational complexity reductions. Sasaki (2004) achieves this via approximations, whereas Siripunvaraporn (2005) replaces the model-space \( M \times M \) system of linear normal equations by a data-space \( N \times N \) system of linear normal equations. The rationale behind this is that in 3D we typically have that \( N \ll M \).

We would like to put forward the idea of performing 3D marine MT inversion using a model-space Gauss-Newton scheme where the full Jacobian is computed at each inversion iteration. The motivation is that the Gauss-Newton approximation to the data Hessian \( J^tJ + \text{c.c.} \) is expected to be of higher quality than that which can be obtained by gradient-type schemes. This can be intuitively understood by noting that the former is based on partial derivatives of each data point with respect to each model parameter \( (N \times M \text{ quantities in total}) \), whereas in the latter, all that is available to obtain an approximation to the data Hessian is the history of \( M \)-dimensional gradient vectors from each inversion iteration. As a result, gradient-type inversion schemes often require a large number of iterations to converge – which has led researchers to consider preconditioning schemes to remedy the situation (Newman and Boggs, 2004; Plessix and Mulder, 2008).

If the Gauss-Newton scheme is implemented with care, the memory and complexity requirements are sufficiently modest to allow the inversion of 3D marine MT data sets on a small computer cluster. We find computing the Jacobian at each inversion iteration well worth the computational expense. Indeed, in comparison to our former quasi-Newton scheme based on the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) update method (Byrd et al., 1995; de la Kethulle de Ryhove and Mittet, 2014), testing indicates that the proposed Gauss-Newton scheme has at least three advantages: reduced start model dependence, reduced number of iterations for convergence without any need for preconditioning, and a noticeable reduction of shallow receiver-imprint related artefacts.

The remainder of this paper is organized as follows: we start by presenting our Gauss-Newton 3D marine MT inversion algorithm. Some inversion results are discussed thereafter before drawing some concluding remarks.

Efficient Gauss-Newton 3D marine MT inversion

Let the earth be divided into \( M \) isotropic cells, let each cell be assigned a conductivity value \( \sigma_i \), and let \( \mathbf{m} \) be a vector of length \( M \) containing these values. Our inversion scheme employs a Gauss-Newton algorithm (Nocedal and Wright, 1999) to minimize the cost function

\[
\varepsilon(\mathbf{m}) = \varepsilon_{\text{data}}(\mathbf{m}) + \lambda \varepsilon_{\text{reg}}(\mathbf{m}),
\]

where \( \lambda \in \mathbb{R}^+ \) is the Tikhonov regularization parameter, \( \varepsilon_{\text{reg}} \) is a regularization term, and the data cost term reads

\[
\varepsilon_{\text{data}}(\mathbf{m}) = \sum_{i,j \in \Omega} \left| W_{i,j,\omega,n}(Z_{i,j,\omega,n}^{\text{obs}} - Z_{i,j,\omega,n}^{\text{syn}}(\mathbf{m})) \right|^2.
\]

Here, the indices \( i, j \in \{x,y\} \) denote components of the \( 2 \times 2 \) impedance tensor \( Z \), \( \omega \) denotes angular frequency, \( n \) is a site index, \( W_{i,j,\omega,n} \) is a data weight, and the superscripts \( \text{obs} \) and \( \text{syn} \) respectively denote observed and forward-modelled synthetic data. The regularization term used in this work is a quasi-\( L_p \) norm \( (p \in \{1,2\}) \) of the horizontal and vertical derivatives of the model parameters \( \mathbf{m} \), and is based on an extension to three dimensions and general \( p \) of the regularization scheme presented by Hansen and Mittet (2009). Different weighting
coefficients $\alpha_h, \alpha_v$ can be used for the horizontal and vertical derivatives. The value of the Tikhonov parameter is also controlled as described by Hansen and Mittet (2009).

We now describe our implementation of the Gauss-Newton scheme. Most of the points discussed below need to be carefully dealt with in order to obtain an efficient implementation.

Forward modelling is performed using the finite-difference time-domain (FDTD) approach presented by de la Kethulle de Ryhove and Mittet (2014). For a specific source configuration, it has the advantage of allowing the computation of the electromagnetic field unknowns at all frequencies of interest in a single simulation. In comparison to modelling based on a frequency-domain approach (Mackie et al., 1993b; Smith, 1996), where for a specific source configuration a new forward modelling job needs to be run for each frequency of interest, this in principle allows to reduce the number of modelling jobs by a factor $n_f$ (number of frequencies). However, as for each FDTD job the modelling time is roughly proportional to $N_s \sqrt{T_{max}}$, with $T_{max}$ the longest period one wishes to compute electromagnetic field responses for and $N_s$ the number of electromagnetic field unknowns, we find it useful to divide the periods of interest into 2-3 groups and run separate modelling jobs for each group. If, for each group, the finite-difference (FD) stencils are gradually coarsened as the maximum period $T_{max}$ in the group increases, we find that it usually is possible to achieve computational savings in spite of the increase in the total number of jobs. For example, if one wishes to produce modelling results for periods in the range $1.5s < T < 375s$, splitting the period range into two groups ($G_1$ with periods such that $1.5s < T \leq 25s$, and $G_2$ with periods such that $25s < T \leq 375s$) and using an FD stencil that is four times coarser for the long-period group gives a total modelling time that is roughly proportional to $N_s \sqrt{25} + \frac{1}{2} N_s \sqrt{375}$, whereas the total modelling time if one produces results for all periods at once using the finer stencil would be roughly proportional to $N_s \sqrt{375}$. Here the computational complexity can thus be reduced by a factor $\sim (\sqrt{375}/4 + 25)/\sqrt{375} \approx 1.97$.

It is essential to compute the sensitivities $\partial Z_{ij}^{syn}/\partial \alpha_m$ using an adjoint approach (McGillivray and Oldenburg, 1990; Avdeev, 2005) to restrict the number of forward modelling jobs. The MT impedance is defined in terms of four field components ($E_x, E_y, H_z, H_y$). For each site and frequency group, the Jacobian can hence be constructed based on the forward modelling results for four different sources. In addition, for each frequency group, forward modelling for two plane-wave MT sources is required to compute the direct fields. This gives a total of $4 \times$ (number of receivers) + 2 forward modelling jobs for each frequency group to assemble the Jacobian.

In order to avoid running four different forward modelling jobs per frequency group at each site $x$, it is tempting to use the superposition principle to construct linear combinations $\alpha_x J_x + \alpha_y J_y + \beta_x K_x + \beta_y K_y$ of $x$ and $y$-oriented unit electric and magnetic dipole sources, where the coefficients $\alpha_x, \alpha_y, \beta_x, \beta_y$ are a function of frequency and depend on the site’s impedance misfits $\Delta Z_{ij}(x, \omega)$. This idea was used for computing gradients in our former quasi-Newton inversion scheme (de la Kethulle de Ryhove and Mittet, 2014). However, it comes with two disadvantages. First, the forward modelling jobs for the direct fields need to be finished in order to determine $\alpha_x, \alpha_y, \beta_x, \beta_y$ for each site and frequency before the remaining adjoint modelling jobs can start. This prevents parallelism and introduces delays in wall clock time. Second, generating fictitious time-domain sources for our FDTD algorithm with spectra determined by the values of $\alpha_x, \alpha_y, \beta_x, \beta_y$ at each frequency is a numerically difficult problem because the transform from frequency to fictitious time is non-unique (de Hoop, 1996). This limits the number of periods in each group to a maximum of five (de la Kethulle de Ryhove and Mittet, 2014) and may lead to accuracy issues if the fictitious time-domain sources are not generated with care. In our implementation of the Gauss-Newton scheme, we therefore choose not to construct such linear combinations.

We now discuss the assembly of the Jacobian once all modelling jobs have been successfully executed. The Jacobian is a matrix with $N$ (number of data points) rows and $M$ (number of unknowns) columns. For a reasonably-sized 3D marine MT survey with say 100 receivers and 20 periods, inverting all four impedance tensor components gives $N = 8000$ data points. For purposes of 3D discretizations of the earth used in our experiments contained in the region of $M \approx 500000$ unknowns corresponding to the isotropic conductivities of $M$ rectangular cells. Following standard practice, these discretizations were finer close to the region where the receivers were deployed, and became coarser as one moved away from this region horizontally or in depth. In principle, these earth discretizations need not be related to the FD stencils used in the forward modelling though for accuracy reasons it is advisable to try to have at least one Yee-grid node per cell edge.

With $M = 500000$ and $N = 8000$, the memory requirement for storing the Jacobian in complex double precision is $\sim 60$ GB. In order to assemble the Jacobian efficiently, and thereafter solve the system of linear normal equations to find the search direction, we find it important to parallelize the task with the help of a Message Passing Interface (MPI) implementation.

This is done as follows. A number of compute nodes is set aside to assemble the Jacobian. Together, these nodes must have sufficient memory for distributed storage of the Jacobian. A number of MPI process $n_p$ is chosen. The set of receivers is then partitioned into $n_p$ disjoint groups. For efficient load balancing, the number of receivers per group should vary as little as possible. Each receiver group is assigned to a specific MPI process. The Jacobian assembly can then start. Each MPI process assembles the Jacobian rows corresponding to the data points belonging to the receivers it has been assigned. For this purpose, in addition to the direct fields, each MPI process only needs to know the adjoint fields corresponding to the receivers it has been assigned.

The system of linear equations $(J^T J + \epsilon c) \Delta m = -g$ now needs to be solved for $\Delta m$ in order to obtain the search direction, with $H_{reg}$ the Hessian of $\epsilon_{reg}(m)$ and the gradient $g = (\partial \epsilon/\partial \sigma_1, \cdots, \partial \epsilon/\partial \sigma_M)$. Since it is prohibitively expensive to form the $M \times M$ matrix $J^T J$, we use a conjugate gradient (CG) solver (Nocedal and Wright, 1999; Avdeev, 2005) with Jacobi preconditioning for this purpose. At each CG iteration,
two matrix-vector products $y = Jx$ and $z = J^\dagger y$ need to be evaluated. In our implementation, each MPI process takes care of the rows of $J$ it has assembled. The partial products computed by each MPI process are simply summed at the end to calculate $y = Jx$ or $z = J^\dagger y$. The computation of $H_{\text{reg}}\Delta m$ is simpler and is not discussed further here due to space limitations.

Finally, once $\Delta m$ has been found, a line search based on a backtracking approach (Nocedal and Wright, 1999) is carried out. Since only the value of the cost function $\epsilon(m)$ needs to be evaluated at each line search iteration, this requires only two modelling jobs per frequency group to compute the direct fields. The associated computational expense is hence small.

**Inversion examples**

The Gauss-Newton algorithm presented above was applied to a 2010 data set from the western Barents Sea. See (de la Kethulle de Ryhove and Mittet, 2014) for a description of this data set together with inversion results obtained using our former quasi-Newton L-BFGS scheme. The MT data used in the inversion consist of impedance tensors for 17 periods per receiver ranging from 1.45s to 375s. We had data up to $T \approx 10000\,$s, though since only depths 0km $\leq \Delta \leq 10\,$km were of interest to us the longest periods were excluded. Initial inversion tests were carried out with 75 of the 80 deployed receivers as in (de la Kethulle de Ryhove and Mittet, 2014). However, the inversion struggled to achieve a good data fit for three of these 75 receivers. Examining the observed data revealed that in comparison to their neighbours these receivers displayed anomalously high apparent resistivities for one of the off-diagonal modes, which did not seem physical to us. These receivers were consequently excluded from later inversion runs. The results presented in this section are for 72 of the 80 deployed receivers. This gives $N = 4896$ data points.

The north-east-depth dimensions of the 3D inversion resistivity model were $82 \times 62 \times 20\,$km. This volume was divided into $M = 536210$ rectangular cells. The start model consisted of a $3.3\,\Omega\,\text{m}$ water layer and a homogeneous formation layer with conductivity $0.26\,\Omega\,\text{m}^{-1}$. The 17 periods were divided into one short-period group (9 periods in the range $1.45\,$s to $23.4\,$s) and one long-period group (8 periods in the range $32.8\,$s to $375\,$s). This leads to a total of 580 forward modelling jobs per Gauss-Newton iteration. The FD stencil for the long-period group was four times coarser than for the short-period group.

The inversion converged in eleven Gauss-Newton iterations to a root mean square (RMS) data misfit of 1.46, with the weights in equation (2) set to $W_{ij,\omega,n} = 3\% |\rho_{obs}^{\omega,n}|$ for off-diagonal impedance tensor components, and respectively $W_{xx,\omega,n} = 10\% W_{xx,\omega,n}$ and $W_{yy,\omega,n} = 10\% W_{yy,\omega,n}$ for the diagonal components. These eleven Gauss-Newton iterations can be compared to 196 iterations for our former quasi-Newton scheme which, in addition, failed to converge starting from homogeneous half-space models. The data, regularization, and total RMS misfits as a function of the iteration number are displayed in Figure 1.

The inversion was run on a computer cluster consisting of a combination of nodes with dual Intel Xeon E5-2650 2.6GHz CPUs, nodes with dual Intel Xeon E5-2660 2.2GHz CPUs, and nodes with dual Intel Xeon E5-2680 2.7GHz CPUs. All nodes had 16 processors and either 64GB or 128GB memory. For the first and tenth Gauss-Newton iterations, the average forward modelling runtimes were of respectively $5.5$ and $30.1\,$min. The increase in average runtime from the first to the tenth iteration is due to the increase in maximum resistivity from $4\omega\,$m for the start model to $\sim 100\omega\,$m at the tenth iteration (the runtime of our FDTD forward modelling algorithm is proportional to the square root of the maximum resistivity in the input earth model (de la Kethulle de Ryhove and Mittet, 2014)). For both the first and tenth Gauss-Newton iterations, it took approximately $9\,$min. to assemble the Jacobian using two nodes and four MPI processes per node. Solving the system of linear normal equations with these same compute resources thereafter took respectively $401\,$CG iterations and $\sim 19\,$min. for the first Gauss-Newton iteration, and $997\,$CG iterations and $\sim 48\,$min. for the tenth Gauss-Newton iteration. The CG solver termination criterion was a reduction in the 2-norm of the preconditioned residual of $10^{-5}$. We believe this may be stricter than necessary, though additional testing is needed to confirm this.

The central part of the resistivity model the inversion converged to is shown in Figure 2. The resistivity remained below $\sim 1.3\,\Omega\,\text{m}$ below the 10-km investigation depth. As with our former quasi-Newton inversion, three resistive bodies, marked A, B and C in Figure 2, were identified. Resistive body A, interpreted to be a salt layer, extended to the east and south boundaries of the earth model in the current inversion, but in our former inversion had a maximum extension of 9km in the south-north direction and 13km in the east-west direction. The larger horizontal extent of this body in the current inversion may be due to the horizontal regularization. The maximum depth of this body was $\sim 7.5\,$km in the current inversion and $\sim 6.5\,$km in our former inversion. Resistive body B, interpreted to be salt, had a maximum depth of $\sim 6.5\,$km in the current inversion and $\sim 5.5\,$km in our former quasi-Newton inversion. Differences in the geometry of this body were observed between the two inversions. This was also the case for the shallow object marked C.

In addition to the inversion discussed above, a number of inversions on the same data set with different regularization settings were run. Among others we observed that object A, which...
Full-Jacobian Gauss-Newton 3D marine MT inversion

Figure 2: Final resistivity model. As with our former quasi-Newton inversion, three resistive bodies (A, B and C) were identified.

we would not expect to affect short-period data, did not appear in over-regularized inversions; and that detailed object geometries, in particular that of object B, were dependent on the relative weighting of the horizontal and vertical derivatives in the regularization scheme. This underlines the importance of proper regularization tuning. A test plan involving synthetic inversions in different scenarios is one way of achieving this.

In order to examine receiver-imprint related inversion artefacts, synthetic inversion tests where deep resistive targets had to be recovered starting from a half-space with the correct $2\Omega m$ background resistivity were carried out. Vertical cross-sections extending to depths $z = 1.5km$ for the final models obtained with a quasi-Newton L-BFGS inversion similar to that presented in (de la Kethulle de Ryhove and Mittet, 2014) and the Gauss-Newton inversion scheme discussed in this paper are shown in Figure 3. Receiver-imprint related artefacts are clearly less visible in the model obtained with the Gauss-Newton scheme. This can be at least partly understood by noting the following: the first step taken by the L-BFGS scheme will be in the steepest descent direction, and if the start model minimizes the regularization term of the cost function, the regularization gradient vanishes as long as $\epsilon_{\text{reg}}(m)$ is a twice differentiable function of $m$, which is usually the case. Hence only the data misfit gradient steers the L-BFGS scheme at the first iteration. This is different for the Gauss-Newton scheme because $H_{\text{reg}}$ is non-zero even in the above situation. As a result, the regularization term will play a role in the determination of the Gauss-Newton direction at the first iteration and penalize updates with sharp contrasts.

Conclusions

We have developed a Gauss-Newton 3D marine MT inversion scheme where the full Jacobian is computed at each inversion iteration. Its computational requirements allow the inversion of 3D data sets on a small computer cluster. Although assembling the Jacobian at each iteration comes at a relatively large cost, we find this is clearly worth the effort. Indeed, synthetic and real inversion tests indicate that, in comparison to gradient-based quasi-Newton inversion algorithms, the proposed scheme has at least three advantages: reduced start model dependence, reduced number of iterations for convergence without any need for preconditioning, and a noticeable reduction of shallow receiver-imprint related artefacts.

Acknowledgments

The authors wish to thank EMGS for permission to publish the results, as well as Dr Kristian Rymann Hansen for numerous interesting discussions, and in particular his insightful comments related to receiver imprints in quasi-Newton and Gauss-Newton schemes which have been included in the paper.
EDITED REFERENCES
Note: This reference list is a copyedited version of the reference list submitted by the author. Reference lists for the 2016 SEG Technical Program Expanded Abstracts have been copyedited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

REFERENCES
