Large-scale Gauss-Newton inversion of transient CSEM data using the model reduction framework

M. Zaslavsky, V. Druskin, A. Abubakar, T. Habashy and V. Simoncini*

Schlumberger-Doll Research, Cambridge, MA,

*Dipartimento di Matematica, Università di Bologna, Italy

(December 13, 2012)

Running head: Model reduction for tCSEM inversion

ABSTRACT

Usually transient data controlled-source electromagnetic measurements is interpreted via extracting few frequencies and solving the corresponding inverse frequency-domain problem. However, typically ad hoc, frequency sampling may result in loss of information and affect the quality of interpretation. On the other hand, fitting data directly in the time domain automatically exploits all available information. The main drawback of such an approach is its large computational cost, in particular, when the Gauss-Newton (GN) algorithm is employed for the misfit minimization. That cost is mainly comprised of the multiple solutions of the forward problem and linear algebraic operations using the Jacobian matrix for calculating the GN step. For large-scale 2.5D and 3D problems with multiple sources and receivers, the corresponding cost grows enormously for inversion algorithms using conventional finite-difference time-domain (FDTD) algorithms. To partially mitigate this problem, a fast three-dimensional (3D) forward solver based on the rational Krylov subspace reduction algorithm using an optimal subspace selection was proposed earlier. In this work, we apply the same approach to reduce the size of the time-domain
Jacobian matrix. The reduced-order model (ROM) is obtained by projecting a discretized large-scale Maxwell system onto a rational Krylov subspace (RKS) with optimized poles. The RKS expansion replaces the time discretization for both forward and inverse problems; however, for the same or better accuracy, its subspace dimension is much smaller than the number of time steps of the conventional FDTD. The crucial new development of this work is the space-time data compression of the ROM forward operator and decomposition of the ROM’s time-domain Jacobian matrix via chain rule, as a product of time-dependent and space-dependent terms, thus effectively decoupling the discretizations in the time and parameter spaces. All the above, together with the weighted $L_2$-norm regularization technique for the GN inversion, allowed us to arrive at an efficient algorithm for inversion of tCSEM data. We illustrate our approach using synthetic 2.5D examples of hydrocarbon reservoirs in the marine environment.
INTRODUCTION

The controlled-source electromagnetic (CSEM) method is becoming an increasingly popular tool for oil and gas exploration. In practice, CSEM acquisition systems can be subdivided into frequency-domain (fCSEM) and time-domain (tCSEM) methods.

In the former technique, the electromagnetic field broadcast by a transmitter is periodic. Then, for interpretation, one or a few frequencies are extracted from the measured signal. Flosadottir and Constable (1996) developed the theoretical aspects and performed a synthetic analysis; an experimental study and results of real survey are considered, e.g., in Constable and Cox (1996). 2.5D and 3D inverse fCSEM problems were successfully solved by Abubakar et al (2008), Mitsuhata et al (2002) and Abubakar et al (2009), Commer and Newman (2008), Plessix and van der Sman (2008), respectively.

The excitation current in tCSEM system is given by sequences of impulse or switch-off/switch-on modes (Barsukov et al., 2007). One way to interpret tCSEM data is to extract the data for few frequencies and perform frequency-domain inversion. However, typically ad hoc choices of frequency sampling may cause loss of information, worsening of inversion results and/or increase of computational costs. Indeed, when a rather small number of frequencies is used, the approximation of time-domain signal becomes very poor, and lost information sometimes may significantly affect the interpretation. With an increase in the number of frequencies and by refining of frequency sampling, the loss of information can be avoided, but the complexity of the problem may become very high.

Another way to interpret the tCSEM data is by inverting the transient responses directly. It allows one to avoid any potential loss of information, but has quite a few challenges. The tCSEM forward problem is governed by parabolic Maxwell equations assuming that the
displacement current is neglected. The inverse problem is nonlinear and is usually solved by using iterative algorithms such as the Gauss-Newton method or the nonlinear conjugate gradient (NLCG) method (Haber et al., 2004; Zhdanov, 2002), calling computationally intensive forward modeling solvers in the loop. The Gauss-Newton (GN) method is especially expensive because of the calculation of its update direction and adjoint computations. Therefore, a fast and robust forward solver is crucial for the inversion efficiency.

CSEM forward modeling problems have several challenges. They require very large computational domains to approximate the boundary conditions at infinity and very fine meshes around sources, receivers, and targets. That results in a very large (up to $10^7$, or even more, unknowns) and an extremely ill-conditioned discretized linear system. Furthermore, very large survey time intervals, ranging from milliseconds to tens and hundreds of seconds, cause conventional finite-difference time-domain (FDTD) algorithms to be very expensive.

Another important question is the choice of an iterative algorithm for the nonlinear inversion. Here, we consider the unpreconditioned nonlinear inverse problem only. Indeed, some preconditioners may significantly reduce the number of nonlinear iterations, however, there is typically a trade-off between their efficiency and computational complexity, and this discussion is out of the scope of this paper. The NLCG method (Wang et al., 1994; Newman and Alumbaugh, 2000; Zhdanov, 2002) does not require the computation of the full Jacobian matrix. Instead, it employs a back-propagation (electromagnetic migration) algorithm to compute the gradient, and it is typically performed by using the FDTD. NLCG-type methods, unless they are preconditioned, have a slow convergence rate (Wang et al., 1994) and often reach just local minima with unsatisfactory quality of inverted models. The GN-type methods usually yield more reliable results with a better convergence rate than the NLCG methods. However, computation of the Jacobian matrix in each GN iteration
requires a significant number of additional forward solves, even when using an adjoint formulation (Newman and Alumbaugh, 2000). The number of columns in the Jacobian matrix is the product of the number of sources and receivers as well as sampling times. The latter can be up to $O(10^4)$ or higher in a typical large-scale survey and results in an extremely large Jacobian matrix. Indeed, such a fine sampling of transient data is redundant. However, ad hoc removing part of sampling times from signals is not optimal and may result in loss of information. Moreover, having such a redundancy in the data can be beneficial for interpretation since it may help to suppress the noise.

One of the most popular model reduction techniques, and solution of parabolic equations in particular, is based on the projection onto Krylov subspaces (for a detailed review, see, e.g., Frommer and Simoncini, 2008). Such an approach was efficiently applied to the electromagnetic problems by Druskin and Knizhnerman (1988). A more recent modification of the Krylov projection approach, called the rational Krylov subspace reduction (RKSR), was applied to the transient problems by Boerner et al. (2008), Druskin et al. (2009), Druskin et al. (2010), and Zaslavsky et al. (2011b). Main tool of the RKSR is the direct projection of the semi-discretized time-domain Maxwell system onto the rational Krylov subspace (spanned on time slices of the implicit FDTD solutions, or, equivalently, solutions of Laplace domain problems). In particular, the most efficient modifications of the RKSR were developed by Knizhnerman et al. (2009), Druskin et al. (2009), and Druskin et al. (2010). It used the rational Krylov subspaces (RKS) with shifts (poles) chosen from the point of view of the optimal rational approximation of the matrix exponent on the entire time interval. The algorithm was also generalized to solve electromagnetic forward problems with induced polarization (Druskin and Zaslavsky, 2012; Zaslavsky and Druskin, 2011). For extensive benchmarking of this approach against the FDTD and other algorithms, see
In this work, we extend our RKSR approach for the Gauss-Newton method for the inverse time-domain problem using the model reduction framework. We consider the transient measurements as a multi-input multi-output (MIMO) transfer function (see, e.g., Antoulas, 2005), with inputs and outputs being sources and receivers, respectively. The reduced order model (ROM) is obtained by projecting a discretized large-scale Maxwell system onto a RKS. The RKS expansion plays a role of the time discretization for both forward and inverse problems; however, thanks to the pole optimization, the subspace dimension is much smaller than the number of time steps of the conventional FDTD for the same or better accuracy. The RKSR typically reduces the solution of a large-scale discretized Maxwell’s systems with $O(10^7)$ unknowns to ROMs with $O(10^2)$ unknowns. The procedure allows the compression of the transfer function for multiple source and receiver surveys both in time and space by an accurate approximation via a superposition of a small number of exponentially decaying modes. By representing the nonlinear operator of the forward problem as the composition of the spectral problem and the above mentioned mode superposition, Druskin et al. (2012) showed that the ROM’s Jacobian in the single-input/single output (SISO) formulation allowed a chain rule decomposition via the product of two rectangular matrices. One of them was a low-rank matrix of the discretized Laplace transform and its first derivative (at the spectrum of the projected system), and the other was the Jacobian of the projected spectral problem conveniently computed using standard perturbation theory. This decomposition allowed to decouple time and parameter space discretizations in the Jacobian representation. We extend this approach to the MIMO case. As the result, we dramatically reduce complexity and storage requirements for the Gauss-Newton method. To solve an ill-posed inverse problem, we employed a Tikhonov approach with weighted $L_2$-
The efficiency of the developed approach is confirmed by results on the inversion associated with synthetic 2.5D models of hydrocarbon reservoirs in the marine environment. This paper is organized as follows. In Section 2 we introduce the state (parabolic Maxwell’s) equations and present the forward problem in the control theory MIMO framework. In Section 3 we formulate the inverse problem as a regularized misfit minimization and we numerically solve it by means of the Gauss-Newton minimization algorithm. The subspace projection ROM framework for both the forward and inverse problems is given in Section 4. We outline there the main results of the paper, that are space-time compression of the MIMO transfer function (forward problem operator) and compact representation of the Jacobian via the chain rule. Numerical experiments with full scale 2.5D inversion of synthetic data are reported in Section 5. We outsourced some technical details of the subspace projection approach to the Appendices: Appendix A describes optimal rational Krylov subspaces suitable for our problem; efficient decomposition for ROM Jacobian and Hessian via tensor-products are given in Appendix B. The computation of Ritz pairs derivatives is shown in Appendix C. Also, some supplemental details of regularization and misfit weighting are described in Appendix D.

**FORWARD PROBLEM**

Consider the three-dimensional quasi-stationary time-domain Maxwell system

\[
\nabla \times H^l(r, t) = \sigma(r)E^l(r, t) + J^l(r, t), \\
\n\nabla \times E^l(r, t) = -\mu \frac{\partial H^l(r, t)}{\partial t},
\]

for multiple exciting (external) currents \( J^l(r, t), l = 1, \ldots, m_s \). Here \( H^l(r, t) \) and \( E^l(r, t) \) are the magnetic and electric fields, respectively, due to source \( J^l(r, t) \); \( \sigma(r) \) is the electrical conductivity and \( m_s \) is the number of sources. It is assumed that the magnetic perme-
ability is a positive constant. A quasi-stationary approximation – when the displacement current is neglected – is commonly assumed in low-frequency electromagnetic applications (Davydycheva et al., 2006).

We impose the initial condition

\[ E^i(r, t) = 0, \quad H^i(r, t) = 0, \quad \text{for} \ t < 0. \]

(2)

Expressing \( H^i \) in terms of \( E^i \), we obtain the parabolic quasi-stationary Maxwell’s system

\[ \nabla \times \nabla \times E^i(r, t) + \mu \sigma(r) \frac{\partial}{\partial t} E^i = -\mu \frac{\partial}{\partial t} J^i(r, t). \]

(3)

We consider the switch-off source case, for which

\[ J^i(r, t) = \varphi^i(r)(1 - h(t)), \]

(4)

where \( h(t) \) is the Heaviside function and the vector function \( \varphi^i(r) \) is the initial spatial current distribution.

Reciprocally (in the sense of symmetry of Green’s function, see, e.g., McGillivray et al., 1994), the receiver can be simulated by the current distribution \( \psi^j(r) \), so the time-domain measurements for the \( i \)-th source and the \( j \)-th receiver can be represented as

\[ \int_{\mathbb{R}^3} \psi^j(r) \cdot E^i(r, t) \, dr. \]

For example, if sources and receivers are ideal (point) electric dipoles then the current distributions \( \varphi^i \) and \( \psi^j \) are vector delta functions. For finite electrode lengths, they can be represented as compactly supported distributions (superpositions) of vector delta functions.

Without loss of generality, we can combine all \( \psi^j(r) \) and \( \varphi^i(r) \) in one set of linearly independent vectorial weighting functions \( \{\tilde{\varphi}^j\}_{i=1}^n \). We define the square MIMO transfer or
impedance function as the time-dependent symmetric matrix \( \tilde{F}(t) \in \mathbb{R}^{m \times m} \) with elements

\[
\tilde{F}_{jl}(t) = \int_{\mathbb{R}^3} \tilde{\varphi}_j(r) \cdot E_l(r, t) \, dr, \quad j = 1, \ldots, m, \quad l = 1, \ldots, m,
\]

assuming that \( E_l(r, t) \) is due to the initial current distribution \( \tilde{\varphi}_l(r) \). Then, the measurement set is a subset of the elements of this matrix.

For the spatial discretization, we employ the Cartesian Lebedev finite-difference scheme following Davydycheva et al. (2003). The main advantage of this grid is that it is fully conservative for arbitrary anisotropic media (similar to the Yee grid for isotropic formations). This approach also allows an efficient finite-difference homogenization for accurate representation of sharp high-contrast inhomogeneities and the spectrally matched (a.k.a. optimal) grid in the exterior for an efficient truncation of the computational domain. The discretization on the Lebedev grid with \( N_1 \) spatial nodes yields the semi-discrete Maxwell’s system

\[
A E^l(t) + M \frac{\partial}{\partial t} E^l(t) = \tilde{\varphi}^l \delta(t), \quad E^l(t)|_{t<0} = 0,
\]

where \( \delta(t) \) is the Dirac’s delta function, \( \tilde{\varphi}^l \) and \( E^l(t) \) are \( N \)-dimensional vectors and vector functions of \( t \), respectively, \( N = 3N_1, \mathbb{R}^{N \times N} \ni A = A^T \approx \nabla \times \nabla \times \) is a real symmetric nonnegative-definite matrix, and \( M \) is a positive-definite mass matrix function with entries obtained by the discretization of \( \mu \sigma \) on the computational grid. For brevity, here and below we use the same notations for discrete variables and their continuum counterparts.

For \( l = 1, \ldots, m \), equations 5 can be transformed to the Cauchy problem for the symmetrized first-order ODE system

\[
A u^l(t) + \frac{\partial}{\partial t} u^l(t) = 0, \quad u^l(t)|_{t=0} = b^l,
\]

where \( u^l = M^{\frac{1}{2}} E^l, \quad b^l = M^{-\frac{1}{2}} \tilde{\varphi}^l \) and \( A = M^{\frac{1}{2}} A M^{-\frac{1}{2}} \) is symmetric nonnegative (positive definite on subspace of \( u \) with divergence-free \( M^{-\frac{1}{2}} u \)) matrix (Druskin and Knizhnerman,
For brevity, here and below we will omit the dependence of operators and vectors on \( \sigma \) unless it is required to be shown explicitly.

For each \( l \), the solution to equation 6 can be formally written as the matrix exponential

\[
u_l(t) = \exp(-tA)b^l.
\]

The transfer matrix valued function \( \tilde{F}(t) \) for the semidiscrete problem is given by (7) as

\[
\tilde{F}(t)(\sigma) = B^T e^{-tA(\sigma)}B,
\]

where \( B \in \mathbb{R}^{N \times m} \) is the matrix with columns \( b^l, l = 1, \ldots, m \) and \( T \) denotes transposition.

Each element \( \tilde{F}_{jl}(t) \) of the transfer function corresponds to measurements with the \( j \)-th receiver and the \( l \)-th transmitter. In addition, we assume that the measurements are restricted to discrete times \( t_1, t_2, \ldots, t_K \), so we denote the forward operator as that mapping the conductivity vector \( \sigma \) to the three-dimensional array \( F(\sigma) \in \mathbb{R}^{m \times m \times K} \) of the elements of \( \tilde{F}(t_1)(\sigma) \), i.e., of the scalars \( F_{ijk} = \tilde{F}_{jl}(t_k) \). Obviously, for the SISO case \( (m = 1) \), the transfer function \( F(\sigma) \) is just a vector from \( \mathbb{R}^K \). Unless specified otherwise, we will operate with \( F(\sigma) \) and other three-dimensional arrays from \( \mathbb{R}^{m \times m \times K} \) as vectors of one-dimensional arrays, e.g., \( F(\sigma) = (F_1(\sigma), \ldots, F_L(\sigma))^T \) with \( F_i(\sigma) = F_{ijk}(\sigma) \) assuming lexicographic ordering \( i = m^2(j-1) + m(l-1) + k \).

Due to symmetry (reciprocity) \( F_{ijk}(\sigma) = F_{ljk}(\sigma) \), the space dimension of the admissible \( F(\sigma) \) does not exceed \( 0.5m \cdot (m+1) \cdot K \).

**INVERSE PROBLEM**

To formulate the nonlinear discrete inverse problem, we employ the framework described by Abubakar et al. (2008). The unknown discrete electrical conductivity \( \sigma \) can be obtained
by minimization of the data misfit functional

\[ \sigma = \arg \min_{\sigma \in \mathcal{S}} \Phi^d(\sigma), \tag{9} \]

where \( \mathcal{S} \) is a compact set in the domain of positive conductivities and

\[ \Phi^d(\sigma) = \frac{1}{2} \sum_{k=1}^{K} \sum_{j,l=1}^{m} W_{jlk}((F_{jlk}(\sigma) - d_{jlk}))^2 = (F(\sigma) - d)^T W (F(\sigma) - d). \tag{10} \]

Here, \( W_{jlk} \) are non-negative weights and \( d_{jlk} \) are the measured data, both of the same dimensionality as \( F_{jlk} \). Here we will treat array \( d = (d_{jlk})_{j,l,k=1}^{m,m,K} \in \mathbb{R}^{m \times m \times K} \) as a vector from \( \mathbb{R}^{m^2K} \) (similarly to \( F(\sigma) \)) and denote \( W = \text{diag}(W_{jlk}) \in \mathbb{R}^{m^2K \times m^2K} \). If, for some \( j, l, k \) the measured data are not collected, then we set \( W_{jlk} = 0 \).

Let \( \sigma \in \mathbb{R}^D \) be defined at cell centers of the subgrid \( M \) with \( D = P \times Q \times R \) cells of discretization grid:

\[ \sigma = (\sigma^{pqr})_{p,q,r=1}^{P,Q,R}. \]

Here, we assume that the conductivity distribution outside \( M \) is known and, for brevity, use the same notation \( \sigma \) as for the discrete conductivity in the entire domain.

The inverse problem 9 is known to be unstable and even to have a non-unique solution. However, the set of admissible solutions \( \sigma \) can be significantly restricted by employing a priori information regarding the smoothness of the solution. In our work, we employ the Tikhonov approach and formulate the inverse problem as a minimization of the following cost function

\[ \Phi(\sigma) = \Phi^d(\sigma) + \lambda \Phi^r(\sigma), \tag{11} \]

where \( \Phi^r(\sigma) \) and \( \lambda \) are the regularization functional and regularization parameter, respectively. We employ the finite-difference analogue of the weighted \( H^1 \) norm as a regularization
functional:

\[ \Phi^r(\sigma) = \int \frac{1}{6} \left( b_x^2 \left( (\nabla_x \sigma)^2 + \delta_x^2 \right) + b_y^2 \left( (\nabla_y \sigma)^2 + \delta_y^2 \right) + b_z^2 \left( (\nabla_z \sigma)^2 + \delta_z^2 \right) \right) dV, \]

(12)

where \( \nabla_x, \nabla_y, \) and \( \nabla_z \) are the \( x-, y-, \) and \( z- \) components of the finite-difference gradient, respectively, \( b_x = \left( b^x_{pqr} \right)_{p,q,r=1}^{P-Q,R}, b_y = \left( b^y_{pqr} \right)_{p,q,r=1}^{P-Q,R}, \) and \( b_z = \left( b^z_{pqr} \right)_{p,q,r=1}^{P-Q,R} \) are weights associated with norm, and \( \delta_x, \delta_y, \) and \( \delta_z \) are constants that ensure the regularization term to be non-zero. We discuss the choice of weighting parameters \( W_{jlk} \) as well as a regularization technique in detail in Appendix D.

All the minima \( \sigma \) of \( \Phi(\sigma) \) satisfy the equation

\[ \frac{\partial \Phi^d(\sigma)}{\partial \sigma} + \lambda \frac{\partial \Phi^r(\sigma)}{\partial \sigma} = 0. \]

Taking equations 10 and 12 into account, we obtain the nonlinear extremal equation with respect to \( \sigma \):

\[ \frac{\partial F}{\partial \sigma} W (F(\sigma) - d) + \left( F(\sigma) - d \right) W \frac{\partial F}{\partial \sigma} + \lambda \mathcal{L} \sigma = 0, \]

(13)

with linear operator \( \mathcal{L} = -\nabla_x b_x^2 \nabla_x - \nabla_y b_y^2 \nabla_y - \nabla_z b_z^2 \nabla_z \). To solve equation 13, we employ a Gauss-Newton approach. Let \( \sigma^n, n = 0, 1, \ldots \) be iterates of the algorithm and \( J(\sigma^n) \) be the Jacobian of the forward operator \( F(\sigma) \) with respect to \( \sigma \) at \( \sigma = \sigma^n \). Depending on the context, \( J(\sigma^n) \) can be treated as rank-four tensor with elements \( J_{jlk,\alpha}(\sigma^n) = \left. \frac{\partial F_{jlk}(\sigma)}{\partial \sigma^\alpha} \right|_{\sigma = \sigma^n}, \alpha = 1, \ldots, D \) or \( m^2 K \times D \)-matrix with \( (i, \alpha) \) element \( J_{i,\alpha} \) given by \( \frac{\partial F_i(\sigma)}{\partial \sigma^\alpha} \).

From definition of \( F_{jlk} \) and \( F_i \) it obviously follows \( J_{i,\alpha} = J_{jlk,\alpha} \) with \( i = m^2(j-1) + m(l-1) + k \).

At the \( q \)-th iteration of the Gauss-Newton algorithm, we obtain the linear equation for the search vector \( s^q = \sigma^q - \sigma^{q-1} \)

\[ H(\sigma^{q-1}) s^q = -r(\sigma^q), \]

(14)
where $H$ is the Hessian of the functional $\Phi$:

$$H(\sigma^q) = (J(\sigma^q))^T W J(\sigma^q) + \lambda \mathcal{L},$$

and

$$r(\sigma^q) = J(\sigma^q)^T W (F(\sigma^q) - d) + \lambda \mathcal{L} \sigma^q.$$

Although the structure of the cost function in equation 11 is very similar to the one used by Abubakar et al. (2008) for the frequency-domain inversion, there is a principal difference between them, caused by the complexity of the data. The number of operating frequencies in the frequency-domain problem is typically of order one, while the number of time nodes $K$ in the time-domain problem can be of order $10^3$ and even $10^4$. That results in the proportional increase of the problem complexity. In the next section, we will consider the algorithm of compression of the time-domain transfer function and its Jacobian by means of low-rank approximations, reducing the computational complexity by several orders.

**Remark 1** Typically, when solving the inverse problem, constraints on the upper and lower bounds of the conductivity are imposed. To incorporate that into the inversion algorithm, we make a nonlinear change of variables

$$\sigma = \frac{\sigma^\text{max} e^c + \sigma^\text{min} e^{-c}}{e^c + e^{-c}}.$$

Here $\sigma^\text{max}$ and $\sigma^\text{min}$ are the upper and lower bounds for conductivity, respectively. Another feature of our inversion algorithm is a so-called line search that follows the Gauss-Newton step. It allows us to ensure that the cost function is monotonically decreasing at each iteration. For details of both line search and change of variables we refer the reader to Abubakar et al. (2008).
In this section, we describe the concept of the subspace projection method to obtain a low rank approximation of the MIMO transfer function and correspondingly, of the forward problem operator and its Jacobian. The essence of this approach is to obtain accurate approximations of the semi-discretized problem via projection onto a subspace $U_n$ of rather small dimension $n$. Different approaches to choices of such subspaces are discussed in Appendix A. Here we briefly outline the Galerkin-projection model reduction approach, see, e.g., Frommer and Simoncini (2008) for more references. Let $G^n = [g_1, \ldots, g_n] \in \mathbb{R}^{N \times n}$, $N \ll n$, be the matrix of an orthogonal basis on $U_n$. Subsequently, the approximate solution of equation 6 is obtained by projection onto the $U_n$ as

$$u_n(t) = G^n e^{-tH_n(G^n)^TB},$$

where $H_n = (G^n)^T A G^n \in \mathbb{R}^{n \times n}$ is the Galerkin-Ritz projection of $A$ on the subspace. Consequently, we obtain the following expression for $\tilde{F}_n(t) \approx \tilde{F}(t)$:

$$\tilde{F}_n(t) = B^T G^n e^{-tH_n(G^n)^TB}.$$

Assuming that $n$ is small, $\tilde{F}_n$ can be computed analytically using the spectral decomposition of $H_n$. Indeed, let $(\lambda^i, \phi^i) \in \mathbb{R} \times \mathbb{R}^n$ be the eigenpairs (a.k.a. Ritz values) of $H_n$: $H_n \phi^i = \lambda^i \phi^i, \|\phi^i\| = 1, \ i = 1, \ldots, n$. We can present the reduced order transfer function via spectral decomposition as

$$\tilde{F}_n(t) = B^T G^n e^{-tH_n(G^n)^TB} = \sum_{i=1}^{n} e^{-\lambda^i t} Y_i Y_i^T,$$

where $Y_i = B^T G^n \phi^i \in \mathbb{R}^m, \ i = 1, \ldots, n$. The reduced order forward operator $F^n \approx F$ is given by the elements of $\tilde{F}_n(t_k)$. The reduced order spectral decomposition formula 17
allows to represent it as

\[ F^n = \Psi(\Xi^n), \]

where \( \Xi^n = (\lambda_1, Y_1^T, \ldots, \lambda_1, Y_1^T)^T \in \mathbb{R}^{n(m+1)} \). Therefore, we arrived at the key property of our reduced order formula 17: it represents the 3D data set as a superposition of products of one-dimensional arrays, given by the components of \( Y \) and \( \exp(-t_k \lambda) \), i.e., it compresses the forward operator both in space and time. The approximate forward operator \( F^n(\sigma) \) can be parametrized in terms of \( n(m+1) \) components of \( \Xi^n \), compared with \( 0.5(m+1)mK \) parameters (accounting for the reciprocity) of the exact operator \( F \).

The subspace projection approach allows us to compress not only the transfer function, but also its Jacobian. The reduced order forward operator is a composition of two nonlinear mappings (from right to left): the Ritz spectral problem \( \Xi^n(\sigma) : \mathbb{R}^D \mapsto \mathbb{R}^{n(m+1)} \), and the reduced order spectral decomposition \( \Psi(\Xi^n) : \mathbb{R}^D \mapsto \mathbb{R}^{n(m+1)} \), i.e., it can be written as

\[ F^n(\sigma) = (\Psi \circ \Xi^n)(\sigma). \]

The chain rule allows us to decompose the Jacobian of the ROM \( J^n \) as a product of \( m^2K \times n(m+1) \) and \( n(m+1) \times D \) Jacobian matrices of the composition mappings, i.e.,

\[ J^n(\sigma) = E(\Xi^n(\sigma))B(\sigma), \tag{18} \]

where

\[ E(\Xi^n(\sigma)) = \frac{d}{d\Xi^n} \Psi|_{\Xi^n(\sigma)} \in \mathbb{R}^{m^2K \times n(m+1)}, \]

and

\[ B(\sigma) = \frac{d}{d\sigma} \Xi^n(\sigma) \in \mathbb{R}^{n(m+1) \times D}. \]

Now, with the help of decomposition 18 we can rewrite the equation for Gauss-Newton
search vector 14 as

$$H^n(\sigma^q)s^q = -\tilde{r}^n(\sigma^q),$$

(19)

where $H^n$ is the reduced order Hessian

$$H^n(\sigma^q) = (B(\sigma^q))^T W^n B(\sigma^q) + \lambda \mathcal{L},$$

$$W^n = E(\Xi^n(\sigma^q))^T W E(\Xi^n(\sigma^q)),$$ and

$$r^n(\sigma^q) = B(\sigma^q)^T E(\Xi^n(\sigma^q))^T W (F(\sigma^q) - d) + \lambda \mathcal{L} \sigma^q.$$

The exact Jacobian $J(\sigma)$ is a $m^2K \times D$ matrix. Obviously, the number of rows can be halved (more precisely, reduced to $\frac{m(m+1)}{2}$) due to reciprocity, but for real scale tCSEM measurements it can still be a significant number. Coupling together with conductivity dimensions $D$ used for large scale imaging, it results in prohibitively large $J(\sigma)$. Using ROM 17 we decoupled two the largest dimensions $K$ and $D$ in the Jacobian representation. For example, instead of $O(m^2KD^2)$ operations for the direct computation of the exact Jacobian $H$, to compute $H^n$ we need $O(n^2K)$ operations for $W^n$ and $O(D^2mn)$ operations for multiplications by $B$ and $B^T$. The derivation of these estimates and efficient formulas of the Hessian computation is given in Appendix B. Indeed, as was already mentioned, our approach can be only efficient if $n \ll Km$ and $n \ll mD$, that can be easily achieved by proper subspace selection for the MIMO problems, see Appendix A.

We should note, that the compression of the model parameters $\sigma$ (under some a priori given assumptions regarding their distributions) can further reduce the Jacobian’s complexity (see Lin et al., 2012), however this procedure is out of the scope of this paper.
NUMERICAL RESULTS

In our numerical examples, we employed a Lebedev staggered finite-difference scheme to discretize the problem in space (Davydycheva et al., 2003). Similarly to Zaslavsky et al. (2011a), we combined the fine uniform grid in the domain of interest and the coarse optimal grid outside. We used block rational Krylov subspace with a priori optimally chosen shifts (Appendix A) as a projection subspace $U_n$. For solving real symmetric linear systems to construct the RKS, we used the preconditioned conjugate gradient method. Following the ideas of the finite-difference integral equation approach developed by Zaslavsky et al. (2011a), we employed an efficient preconditioner obtained by using the finite-difference Green’s function in layered medium. All the tests, which are related to marine CSEM environment, were performed on a PC with a 2-GHz CPU and 4Gb of memory.

For all numerical tests, we considered inverse marine CSEM problem with water depth 1 km, constant water, reservoir and background conductivities 3 S/m, 0.01 S/m and 0.5 S/m, respectively. the domain $[-10 \text{ km}; 10 \text{ km}] \times [-2 \text{ km}; 2 \text{ km}] \times [-3 \text{ km}; 0 \text{ km}]$ is discretized using 40 nodes along the $x$-direction with a cell size of 0.5 km, 8 nodes along the $y$-direction with a cell size of 0.5 km, and 60 nodes along the $z$-direction with a cell size of 0.05 km. The computational domain was extended to infinity using 20 cells along the $x$- and $y$- directions and 36 cells along the $z$-direction. Therefore, the total number of grid nodes is $60 \times 28 \times 96$. Although the problem is fully 3D, in our numerical examples, we reconstructed 2D models only, which are invariant in the $y$-direction (2.5D problem). The inversion domain in $(x; z)$ coordinates is $[-10 \text{ km}; 10 \text{ km}] \times [-3 \text{ km}; -1 \text{ km}]$ and, correspondingly, the inversion grid consists of $40 \times 40$ cells. In all the tests, we employ the same locations for $x$-oriented source and $x$-oriented receiver arrays (indicated by black dots on the top of the inversion domain).
with $m = 15$ and the data were measured at $K = 10^4$ temporal nodes over a time interval from 0.001 s to 100 s, so accounting for reciprocity, we had $m \cdot (m + 1) \cdot K/2 = 1.2 \cdot 10^6$ measurements. For the inversion, we used the RKS with $n = 225$, i.e. subspace of size $n/m = 15$ per source. The approximation error of the corresponding ROM is much smaller compared to the typical accuracy of the measured data (convergence curves can be found in Druskin et al. (2009) and Zaslavsky et al. (2011b)). Hence, we have $n \cdot m = 3.375 \cdot 10^3$ and, roughly speaking, we reduced the Jacobian matrix’s dimension approximately by a factor of $1.2 \cdot 10^6/3.375 \cdot 10^3 \approx 0.35 \cdot 10^3$. Thanks to real arithmetics in our approach, our storage requirements and computational cost are equivalent to the ones for 7-8 and 3-4 frequencies, respectively, in the frequency domain inversion. For the measured data, we used the simulated responses (computed on a finer grid), as well as the ones perturbed by 1% and 3% relative random uncorrelated noise. No additive environmental noise was introduced. A homogeneous background was used as an initial guess for inversion in all the tests.

In the first test we considered a single homogeneous rectangular 200m thick resistive block embedded in conductive formation (Figure 1, top). The initial relative data misfit was $2.4 \cdot 10^{-2}$. In Figure 1, we plotted the results of the reconstructions with 0% (second from the top), 1% (third from the top), and 3% noise (bottom). The inversions with no noise and 1% noise data converged after 9 iterations and took 21 hours; 3% noise case took 8 iterations and 18 hours. The relative data misfits were decreased to $1 \cdot 10^{-3}$, $5 \cdot 10^{-3}$, and $1.7 \cdot 10^{-2}$ for the 0%, 1%, and 3% noise cases, respectively. Here and in the example below, the error for given transmitter is distributed more or less uniformly with respect to different receivers. The inversions with 0% and 1% noise found the lateral boundaries rather accurately and the rectangular shape of the object was preserved; however, the inverted object is slightly
thicker than the original. For the 3% noise case, the lateral boundaries became blurry and the shape of the object was no longer preserved.

In the second test we slightly modified the shape of the object from the previous test (Figure 2, top). Its thickness varies from 200m to 400m. The initial data relative misfit was $2.5 \cdot 10^{-2}$. The inversion results for 0%, 1%, and 3% noise are plotted in Figure 2 second from the top, third from the top and at the bottom, respectively. For the 0% noise case, the shape of the object was preserved, although, similar to the previous test, the reconstructed object is slightly thicker than the original one. Adding 1% noise made the boundaries of the object blurry, while the results for the 3% noise case give just an idea about the shape of the object and its lateral extension. The inversion process for the 0%, 1%, and 3% noise cases converged after 12 (25 hours of computation time), 11 (23 hours) and 7 (15 hours) iterations and reached $9 \cdot 10^{-4}$, $5 \cdot 10^{-3}$, and $1.8 \cdot 10^{-2}$ relative data misfit, respectively.

In the third test, we considered two homogeneous 200m thick resistive objects embedded in conductive formation (see Figure 3, top). The initial relative data misfit was $1.7 \cdot 10^{-2}$. The inversion results are shown in the three bottom plots in Figure 3 (0%, 1% and 3% noise cases are the second, the third from the top and the bottom one, respectively). As one can observe, for the 0% and 1% noise cases the lateral boundaries and the shape of the objects were found accurately; however, the objects are still slightly thicker than the original ones, and adding 1% noise resulted in blurry boundaries. For the 3% noise case, the reconstructed values of conductivity are significantly higher than the original ones, and the inversion results become unreliable. For this example, the inversion algorithm converged after 13 (28 hours of computation time), 12 (25 hours), and 7 (15 hours) iterations and reached $5.7 \cdot 10^{-4}$, $5.8 \cdot 10^{-3}$, and $1.25 \cdot 10^{-2}$ relative data misfit for the cases with 0%, 1%, and 3% noise, respectively.
CONCLUSIONS

We developed the regularized Gauss-Newton inversion algorithm for large-scale, three-dimensional, time-domain CSEM problems using the model reduction framework. This approach allowed us to significantly reduce, particularly for dense time-domain data samplings, both the complexity of the forward problem and linear algebraic operations with the Jacobian matrix without losing significant information. Large-scale numerical examples performed on a PC show the efficiency of the developed approach.

ACKNOWLEDGMENTS

We thank Dr. Garret Flagg for invaluable contribution to the formulation of the original version of our approach as well as for a very useful discussions.

APPENDIX A

OPTIMAL SUBSPACES

Here we will give a brief (not claimed to be complete) overview of efficient choices of projection subspaces. Krylov subspace projection methods have been a popular tool for computing the solution of parabolic equations since the 1980s. The simplest and most popular Krylov subspace (originally developed for solution of eigenproblems and inverting matrices) is a polynomial one: for source term $b^l$

$$U_n^l = \text{span}\{b^l, Ab^l, \ldots, A^{n-1}b^l\}.$$
For the problems with multiple source terms $b^1, \ldots, b^m$, a so-called block polynomial Krylov subspace

$$U_n = \text{span}\{b^1, \ldots, b^m, Ab^1, \ldots, Ab^m, \ldots, A^{p-1}b^1, \ldots, A^{p-1}b^m\}, \ n = mp$$

can be employed. The construction of the orthogonal basis $G^n$ can be done by means of a three-term block recursion (Golub and Underwood, 1977). However, for marine CSEM problems, the condition number of the operator $A$ can be up to $10^{13}$ (Zaslavsky et al., 2011b). For such ill-conditioned operators $A$, the size of the polynomial subspace required for an accurate approximation can be as high as $m \cdot 10^4$ (or even higher). That results in a very poor (if any) compression of the data and Jacobian.

A projection-based approach using the so-called rational Krylov subspaces was developed by Ruhe (1994). Its block analogue for multiple sources has a form

$$U_n = \text{span}\{(A+s_1)^{-1}b^1, \ldots, (A+s_1)^{-1}b^m, \ldots, (A+s_p)^{-1}b^1, \ldots, (A+s_p)^{-1}b^m\}, \ n = mp.$$  

The orthogonal basis $G^n$ can be constructed via the rational Arnoldi algorithm (see Ruhe, 1994). The proper choice of parameters $s_1, \ldots, s_p$ called shifts is crucial for the efficiency of the projection method. An a-priori asymptotically optimal choice of $s_i, i = 1, \ldots, p$ was proposed by Druskin et al. (2010) and was further investigated with application to CSEM problems by Zaslavsky et al. (2011b). It was shown that an optimal choice of shifts is given by the solution of the third Zolotarev problem in the complex plane. In particular, let $r_n(z) = \prod_{j=1}^{n} \frac{z-p_j}{z-s_j}$. Then, we choose parameters $s_i$ that deliver the minimum

$$\min_{(s_1, \ldots, s_n) \in \mathbb{C}^n} \sigma_n(s_1, \ldots, s_n),$$

where

$$\sigma_n(s_1, \ldots, s_n) = \min_{(p_1, \ldots, p_n) \in \mathbb{C}^n} \frac{\sup_{z \in \partial \Lambda | A} |r_n(z)|}{\inf_{\lambda \in \mathbb{R}} |r_n(\lambda)|}.$$
Here $Λ[A]$ is a spectral interval of $A$ (or its estimate). This problem appears to have real solutions $s_i ∈ ℝ$, $i = 1, \ldots, p$. For details, we refer the reader to Druskin et al. (2010) and Zaslavsky et al. (2011b). As was shown in these papers, the size $n = m \cdot 15$, or at most $n = m \cdot 30$, is sufficient to approximate the transfer function rather accurately. This choice of subspace was used for performing the numerical tests.

**Remark 2** An adaptive approach to choose parameters $s_i$ (Druskin et al., 2010) allows us to take advantage of the nonuniformity of the spectrum using information gathered during the iterative process to construct $U_n$. Generally, it gives the same convergence speed as the a-priori approach for weakly uniform media and produces superior convergence for strongly nonuniform formations. However, we did not use it in this work.

In principle, we may reduce the size of the subspace even further, especially for a problem with multiple sources. For example, intuitively it is clear that for small $s_i$, the vectors $(A + s_i)^{-1}b^1, \ldots, (A + s_i)^{-1}b^m$ are rather close to each other. Therefore, we can exclude some of these vectors from $U_n$. Generally speaking, we may seek the subspace $U_n$ in the form

$$U_n = \text{span}\{(A + s_1)^{-1}f_1, (A + s_2)^{-1}f_2, \ldots, (A + s_n)^{-1}f_n\},$$

where $s_i, i = 1, \ldots, n$ are shifts and $f_i ∈ \text{span}\{b^1, \ldots, b^m\}, i = 1, \ldots, n$ are vectors. That is called the problem of tangential interpolation (Ball et al., 1990). To find $s_i, f_i, i = 1, \ldots, n$, there exist several approaches like $H_2$-optimal approximation, balanced truncation, optimal Hankel norm approximation, and others. Very promising results were obtained by Druskin et al. (2012) where the authors investigated $H_2$-optimal approximation with application to inverse 1D SISO problems. It was shown that, if the regularization parameter $λ$ is equal to 0, the error of the $H_2$-optimal approximant does not propagate into the
inverse problem solution. That means that, even for rather inaccurate approximations, the exact solution of the initial full inverse problem is still the solution of the projected inverse problem. However, work on extending the $H_2$-optimal approximation for the solution of inverse multidimensional MIMO problems is still in progress.

**APPENDIX B**

**JACOBIAN IN REDUCED ORDER FORMULATION**

To obtain computationally efficient formulas for the reduced order Jacobian, we will take advantage of tensor structure of $F^n$ and $J^n$ and rewrite some above mentioned derivations in the tensorial form. To avoid using an additional index $n$ and possible confusion with the tensor indices, in the following we abuse notation and use $\hat{\cdot}$ instead of $.n$ for reduced order models of size $n$.

Elements $\hat{F}_{jk}(\sigma)$ of $\hat{F}(\sigma)$ can be written as

$$\hat{F}_{jk}(\sigma) = (b_j)^T G^n e^{-t_k H_n(\sigma)} (G^n)^T b_l. \quad (B-1)$$

Then,

$$\hat{F}_{jk}(\sigma) = (b_j)^T G^n e^{-t_k H_n(\sigma)} (G^n)^T b_l = \sum_{i=1}^{n} (b_j)^T G^n \phi^i(\sigma) e^{-\lambda^i(\sigma) t_k} (\phi^i(\sigma))^T (G^n)^T b_l = \sum_{i=1}^{n} (b_j)^T Z^i(\sigma)(Z^i(\sigma))^T b_l e^{-\lambda^i(\sigma) t_k}, \quad (B-2)$$

where $Z^i(\sigma) = G^n \phi^i(\sigma), \ i = 1, \ldots, n$ are the Ritz vectors of $A(\sigma)$ on $U_n$. Or, equivalently, the approximate transfer function can be rewritten in the form

$$\hat{F}_{jk}(\sigma) = e(t_k)^T C_{jl}, \quad (B-3)$$
where a tensor of rank two with elements $E$ elements of vectors We may rewrite the Jacobian in B-5 as an interior product of two tensors. Denote the function has a form $\lambda^i, (Z^i)^T b^l, i = 1, \ldots, n; l = 1, \ldots, m$, compared to $m \times m \times K$ parameters of $F_{nil}$. 

In fact, from equation B-2, it follows that the elements of Jacobian $\hat{J}$ written as the 3D array have the form

$$\hat{J}_{nil,\alpha} = (\hat{F}_{nil})'$$

$$= \sum_{i=1}^{n} \left( (b^i)^T (Z^i(\sigma))' (Z^i(\sigma))^T b^l + (b^i)^T (Z^i(\sigma))(Z^i(\sigma))' (Z^i(\sigma))^T b^l \right) e^{-\lambda^i(\tau) t_k}$$

$$- \sum_{i=1}^{n} (b^i)^T (Z^i(\sigma))(Z^i(\sigma))' T b^l \lambda^i(1) \lambda^i(\tau), \quad \text{(B-4)}$$

where $'$ denotes $\frac{\partial}{\partial \sigma^\alpha}$. Thus, we obtained that the Jacobian of the approximate transfer function has a form

$$\hat{J}_{nil,\alpha} = \mathcal{E}(t_k)^T B_{nil,\alpha}, \quad \text{(B-5)}$$

where

$$\mathcal{E}(t_k)^T = [e^{-t_k \lambda^i(\tau)}, \ldots, e^{-t_k \lambda^i(\tau)}], \quad t_k e^{-t_k \lambda^i(\tau)}, \ldots, t_k e^{-t_k \lambda^i(\tau)}$$

and

$$B_{nil,\alpha} = \left[ (b^i)^T (Z^i(\sigma))' (Z^i(\sigma))^T b^l + (b^i)^T (Z^i(\sigma))(Z^i(\sigma))' (Z^i(\sigma))^T b^l, \ldots, \right.$$ 

$$\left. (b^i)^T (Z^i(\sigma))' (Z^i(\sigma))^T b^l \lambda^i(1), \ldots, (b^i)^T (Z^i(\sigma))' (Z^i(\sigma))^T b^l \lambda^i(\tau) \right]. \quad \text{(B-6)}$$

We may rewrite the Jacobian in B-5 as an interior product of two tensors. Denote the elements of vectors $\mathcal{E}(t_k)$ and $B_{nil,\alpha}$ as $\mathcal{E}_i(t_k)$ and $B_{nil,\alpha}^p$, respectively. Let $\mathcal{E} \in \mathbb{R}^{2n} \otimes \mathbb{R}^K$ be a tensor of rank two with elements $\tilde{E}_{ik} = \mathcal{E}_i(t_k)$ and $\tilde{B} \in \mathbb{R}^{2n} \otimes \mathbb{R}^m \otimes \mathbb{R}^m \otimes \mathbb{R}^D$ be a tensor.
of rank four with elements $B_{pjl,\alpha}$. Formula B-5 is equivalent to the representation of the Jacobian tensor $\tilde{J} \in \mathbb{R}^m \otimes \mathbb{R}^m \otimes \mathbb{R}^K \otimes \mathbb{R}^D$ with elements $J_{jlk,\alpha}$ as

$$\tilde{J} = \tilde{E} \times_1 \tilde{B}, \quad (B-7)$$

where $\tilde{E} \times_1 \tilde{B}$ denotes the interior product of tensors $\tilde{E}$ and $\tilde{B}$ with respect to the first lower index of $\tilde{E}$ and the first upper index of $\tilde{B}$ (Ruiz-Tolosa and Castillo, 2005). The first term in equation B-7 represents the discretized Laplace transform and its derivative evaluated at the Ritz values, and the second term represents the Jacobian of the spectral (Ritz) problem on $U_n$.

**Remark 3** Formula B-7 generalizes the results for the SISO case obtained by Druskin et al. (2012). Indeed, for $m = 1$ and $b^1 = b$ the tensor $\tilde{B}$ is reduced to the matrix with columns

$$B_{\alpha} = [(b)^T (Z^1(\sigma))^T b + (b)^T (Z^1(\sigma))'(Z^1(\sigma))^T b, \ldots, (b)^T (Z^n(\sigma))^T b + (b)^T (Z^n(\sigma))'(Z^n(\sigma))^T b,\ldots - (b)^T Z^1(\sigma)(Z^1(\sigma))^T b(\lambda^1(\sigma))', \ldots, - (b)^T Z^n(\sigma)(Z^n(\sigma))^T b(\lambda^n(\sigma))')]. \quad (B-8)$$

The Jacobian $J = \{J_{k,\alpha}\}_{k=1,\alpha=1}^{K,D}$ for the SISO case can be obtained as a product of two matrices

$$J = (E)^T B, \quad (B-9)$$

where, for brevity, we identified the matrix $E$ with the tensor $\tilde{E}$.

Similar to expression B-3 for the transfer function, the dependence on $t$ in equation B-5 appears in $E(t)$ only. Therefore, for given $\sigma$ and set of times $t_1, \ldots, t_K$, the Jacobian $\hat{J}$ depends on $Z^t$, $\lambda^t$ as well as on $(Z^t)'$ and $(\lambda^t)'$, totaling $mn+n+mnD+nD = n(m+1)(D+1)$
parameters. That represents a major memory saving compared to the FDTD approach. The computation of \((\lambda^i)'\) and \((Z^i)'\) can be performed using finite differences; however, such approach is computationally expensive. In our paper, we employed a different approach to compute the derivative of Ritz pairs \((\lambda^i; Z^i)_{i=1}^n\), i.e., using the perturbation theory (see Appendix C). It allows us to cheaply compute the approximation to \((\lambda^i)'\) and \((Z^i)'\) that converges while \(n \to N\).

Substituting \(\hat{F}\) and \(\hat{J}\) instead of \(F\) and \(J\) in equation 14, we obtain that on the \(q\)–th Gauss-Newton iteration, we compute \(s^q = \sigma^q - \sigma^{q-1}\) by solving

\[
\hat{H}(\sigma^q)s^q = -\hat{r}(\sigma^q),
\]

where

\[
\hat{H}(\sigma^q) = \left(\hat{J}(\sigma^q)\right)^T W\hat{J}(\sigma^q) + \lambda\mathcal{L},
\]

and

\[
\hat{r}(\sigma^q) = (\hat{J}(\sigma^q))^T W \left(\hat{F} - d\right).
\]

Thanks to the decomposition B-7 of the Jacobian for the subspace projection method, formula B-11 can be rewritten using the interior tensor products as

\[
\hat{H}(\sigma^q) = \sim B_{1|2} \bullet \sim V_{1|2} \bullet \sim B + \lambda \sim \mathcal{L}.
\]

Here, matrix \(\mathcal{L}\) is identified with tensor \(\sim \mathcal{L}\) and

\[
\mathbb{R}^{2n} \otimes \mathbb{R}^{2n} \otimes \mathbb{R}^m \otimes \mathbb{R}^m \otimes \mathbb{R}^m \otimes \mathbb{R}^m \ni \mathcal{V} = \sim E_{1|2} \bullet \sim C_{1|2} \bullet \sim E_{1|2},
\]

where the components of tensor \(\sim \mathcal{C} \in \mathbb{R}^{K} \otimes \mathbb{R}^{K} \otimes \mathbb{R}^m \otimes \mathbb{R}^m \otimes \mathbb{R}^m \otimes \mathbb{R}^m\) are given by \(C_{kk'jj'hh'} = W_{ijkl} W_{j'k'li'} \delta_{jj'} \delta_{hh'} \delta_{kk'}\) with \(\delta\) being Kronecker delta. Hence, we can compute the tensor \(\mathcal{V}\) with \(O(n^2 K)\) operations first, and after that, we obtain the Hessian using
formulas B-12 with $O(D^2 mn)$ operations. Therefore, for large-scale inverse problems, the leading terms of the total computational cost do not contain product of largest dimensions $K$ and $D$ as for the exact case. Thanks to the small dimension $n$ of projection subspace $U_n$, the computation of the Hessian for the subspace projection approach is by several orders of magnitude cheaper compared to the standard approach, whose cost is $O(D^2 m^2 K)$. The computational cost of the solution of linear system B-10 and the cost of a forward solver is comparably insignificant and is absorbed by the above cost. It is also well-known that matrix-tensor operations can be easily parallelized (Johnson et al., 1991). Therefore, the algorithm can be further accelerated.

**APPENDIX C**

**DERIVATIVES OF RITZ PAIRS**

Consider the case $n = N$ first. In this case, Ritz pairs of $A$ on $U_n$ coincide with eigenpairs of $A$. The lemma below is a well-known result from the perturbation theory (Landau and Livshitz, 1977; Golub and Van Loan, 1996)

**Lemma 1** For $i = 1, \ldots, N$, let $(\lambda^i, Z^i)$ be the Ritz pairs of $A$ in $U_N$. Then

$$
(b^i)^T (Z^i)' = \sum_{j=1, j \neq i}^{N} \frac{(Z^i)^T A' Z^j}{\lambda^i - \lambda^j} (b^j)^T (Z^j) \quad (\lambda^i)' = (Z^i)^T A' Z^i, \quad i = 1, \ldots, N, \quad (C-1)
$$

where $A' = \frac{\partial A}{\sigma}$

Because $N$ may be large, the computation of the true Jacobian becomes expensive. Assume next that the approximation space $U_n$ has size $n < N$, and let $(\lambda_i, \phi^i), i = 1, \ldots, n$ be the eigenpairs of $H_n = (G^n)^T A G^n$, so that $Z^i = G^n \phi^i$ are the corresponding Ritz vectors. We
then approximate

\[(Z') = (G^n \phi')' \approx G^n (\phi')', \quad (C-2)\]

that is, we consider the derivative of the basis \(G^n\) negligible. This is clearly a strong assumption; however, it allows us to keep the matrix simple and inexpensive to compute.

Intuitively, it is clear that the better the approximation properties of \(U_n\) the more accurate equation C-2 becomes. In our numerical experiments, we used a subspace size of 15, which was shown to have sufficient accuracy (Druskin et al., 2010; Zaslavsky et al., 2011b).

**APPENDIX D**

WEIGHTING PARAMETERS FOR DATA MISFIT AND REGULARIZATION

To normalize the data misfit, we employ the following data weighting: if for \((j; l)\)-th pair of transmitters and receivers the data is not collected, then \(W_{jlk} = 0, \ k = 1, \ldots, K\) otherwise

\[W_{jlk} = \frac{\Delta t_k f_{jlk}}{\sum_{k=1}^{K} \sum_{m=1}^{M} (d_{jlk})^2 \Delta t_k f_{jlk}}, \quad (D-1)\]

where \(\Delta t_k = t_{\min(k+1, K)} - t_{\max(k-1,1)}\) are time steps. The choice of parameters \(f_{jlk}\) is problem-dependent and can be made to emphasize specific features of the given survey. In particular, for time-domain CSEM problems, similar to frequency-domain CSEM problems, we exploit the wave-guide effect of resistive anomalies to be determined. This effect is mostly detectable by measurements at long-offset transmitter-receiver pairs at late times only. On the other hand, the short-offset transmitter-receiver pairs are used in the time-domain problems in a similar way as during the electromagnetic profiling. That differs time-domain problem from frequency-domain problems where short-offset measurements are not used as they are inaccurate due to near-source singularity. Therefore, the measurements at these
pairs at small times are important. However, these data are singular in time. For example, the response at the transmitter behaves like $t^{-\frac{5}{2}}$ around $t = 0$. Therefore, we choose the parameters $f_{jlk}$ to be equal to $t_{k}^{\frac{1}{2}}$ for short offsets to weaken that singularity, and to be equal to 1 otherwise.

The regularization functional is chosen to make the inverse problem more stable by incorporating a priori information regarding the smoothness of $\sigma$. To reconstruct smooth models, finite-difference (on the grid $M$) $H^1$-norm $\|\sigma\|_{H^1} = \int (\nabla \sigma)^2 dV$ is typically employed (Tikhonov and Arsenin, 1977). Here and below, $\int (\cdot) dV$ denotes a quadrature on grid $M$. For discontinuous models, one may use $TV$-norm $\|\sigma\|_{TV} = \int \|\nabla \sigma\| dV$ (or even $H^\epsilon$-norm for $0 < \epsilon < 1$); however, $\Phi^\epsilon(\sigma)$ is non-smooth in this case. We follow the approach introduced by van den Berg and Abubakar (2001), i.e., employ the finite-difference analogue of weighted $H^1$-norm from formula 12 as the regularization functional.

Following Abubakar et al. (2008), we choose the regularization parameter $\lambda$ and weighted $L_2$-norm parameters $b^x$, $b^y$, $b^z$, $\delta_x$, $\delta_y$, $\delta_z$ adaptively on each Gauss-Newton iteration:

$$\lambda = \lambda^n = \Phi^d(\sigma^n),$$

and for $\alpha = x, y, z$

$$\delta_{\alpha} = \sqrt{\frac{\Phi^d(\sigma^n)}{h_{\alpha}}},$$

$$b^\alpha = b^\alpha,n = \left((\nabla_{\alpha} \sigma^n)^2 + \delta_{\alpha}^2\right)^{-1},$$

where $h_{\alpha}$ is the average grid step along $\alpha$ direction within subgrid $M$. 
Figure 1: Test 1. Model (top) and inversion results for 0% (second from the top), 1% (third from the top) and 3% (bottom) noise.
Figure 2: Test 2. Model (top) and inversion results for 0% (second from the top), 1% (third from the top) and 3% (bottom) noise.
Figure 3: Test 3. Model (top) and inversion results for 0% (second from the top), 1% (third from the top) and 3% (bottom) noise.
REFERENCES

Abubakar, A., T. Habashy, V. Druskin, L. Knizhnerman, and D. Alumbaugh, 2008, 2.5D forward and inverse modeling for the interpretation of low-frequency electromagnetic measurements: Geophysics, 73, F165–F177.

Abubakar, A., T. Habashy, M. Li, and J. Liu, 2009, Inversion algorithms for large-scale geophysical electromagnetic measurement: Inverse Problems, 25, 30pp

Antoulas, A., 2005, Approximation of Large-Scale Dynamical Systems: Cambridge University Press, 479pp


Davydycheva, S., V. Druskin, and T. Habashy, 2003, An efficient finite-difference scheme
for electromagnetic logging in 3D anisotropic inhomogeneous media: Geophysics, 68, 1525–1536.


McGillivray, P., D. Oldenburg, R. Ellis, T. Habashy, 1994, Calculation of sensitivities for
the frequency-domain electromagnetic problem: Geophysical Journal International, **116**, 1-4


Plessix, RE, and P. van der Sman, 2008, Regularized and blocky 3D controlled source electromagnetic inversion: Proceedings of PIERS, 149-154


Zaslavsky, M., S. Davydycheva, V. Druskin, A. Abubakar, T. Habashy, and L. Knizhnerman 1, 2011a, Hybrid finite-difference integral equation solver for 3D frequency domain anisotropic electromagnetic problems: Geophysics, 76, F123-F137.