A low-rank approximation for large-scale 3D controlled-source electromagnetic Gauss-Newton inversion

Manuel Amaya[1], Jan Petter Morten[2], and Linus Boman[2]

ABSTRACT

We have developed an approximation to the Hessian for inversion of 3D controlled source electromagnetic data. Our approach can considerably reduce the numerical complexity in terms of the number of forward solutions as well as the size and complexity of the calculations required to compute the update direction from the Gauss-Newton equation. The approach makes use of "supershots," in which several source positions are combined for simultaneous-source simulations. The resulting Hessian can be described as a low-rank approximation to the Gauss-Newton Hessian. The structure of the approximate Hessian facilitates a matrix-free direct solver for the Gauss-Newton equation, and the reduced memory complexity allows the use of a large number of unknowns. We studied the crosstalk introduced in the approximation, and we determined how the dissipative nature of marine electromagnetic field propagation reduces the impact of this noise. Inversion results from recent field data demonstrated the numerical and practical feasibility of the approach.

INTRODUCTION

The marine controlled-source electromagnetic (CSEM) method is an efficient tool for offshore hydrocarbon exploration with the potential to significantly increase the drilling success rate (Hesthammer et al., 2010). Moreover, the technology has been demonstrated to be effective also in field appraisal (Morten et al., 2011) as well as structural imaging applications (Hoversten et al., 2013; Morten et al., 2013). CSEM data imaging today is based on full-waveform inversion approaches for all of these applications. Using inversion, interpretation challenges related to background complexity and hydrocarbon reservoir variations can be addressed by depth imaging and from quantitative resistivity information in the resulting subsurface models.

State-of-the-art 3D CSEM acquisition offers significant advantages by allowing imaging of lateral variations and also targets, which could be situated between source towlines. If subsurface resistors not related to hydrocarbons are present, then the resolution of interpretation ambiguity is often dependent on understanding the geometry in 3D. Our experience with CSEM data inversion suggests that Hessian-based optimization schemes can often be successful for imaging the 3D geometry of a hydrocarbon reservoir and other resistive structure that may be present, starting from a simple initial guess model. However, the Hessian-based inversion strategies that were originally devised and successfully applied for lower dimensional analysis using a 2D assumption on model geometry (Abubakar et al., 2006, 2009; Li and Key, 2007; Mittet et al., 2007), will lead to very large computational complexity when scaled up for 3D applications.

One of the numerical complexities of inversion approaches that rely on second-derivative information, such as Gauss-Newton, Levenberg-Marquard, or Occam (de Groot-Hedlin and Constable, 1990), arises due to the size of memory needed to store and carry out computations with the Hessian matrix. The number of parameters required to describe the subsurface region covered by a 3D CSEM survey can be of the order of $N \sim 10^6$, and the size of the Hessian will scale as N^2 making the matrix impractical to handle even on large high-performance computer systems. Li et al. (2011, 2013) introduce a model compression method that can significantly reduce the number of parameters for a 3D Gauss-Newton inversion approach. In this paper, we will present a method combining a low-rank approximation of the Hessian combined with a direct solver, so that the number of inversion parameters is no longer a bottleneck.

A second complication for inversion algorithms based on a Gauss-Newton optimization is the large number of forward simulations needed. To construct the Hessian matrix, it is necessary to

compute the Green's functions for the individual source and receiver sensors of the survey. For a large 3D CSEM survey, the number of independent source positions can be very large, often in the order of $N_{tx} \sim 10^5$. The resulting numerical computational load is formidable considering that 3D modeling is required. Grayver et al. (2013) study the use of a direct solver in forward modeling, which has a very gentle scaling with respect to the number of independent source positions that need to be evaluated once the system matrix factorization has been achieved. However, the memory requirement for the matrix factorization can be very large for an industry-scale 3D CSEM survey.

To cope with the above-mentioned numerical complexity of 3D CSEM inversion, descent-based inversion approaches that do not require the full-Hessian computation have been used (see, e.g., Mackie et al., 2007; Zhdanov et al., 2007; Støren et al., 2008). However, these approaches typically require significant manual work to ensure that the initial model reflects the large-scale features of the background geology to achieve acceptable convergence rates and feasible models (Loke and Dahlin, 2002). Higher order methods, e.g., Gauss-Newton, typically require less information in the initial models to achieve these two goals. Nocedal and Wright (2006) show convergence ratios of different optimization methods, and Pratt et al. (1998) show seismic waveform inversions comparing Gauss-Newton and full-Newton methods.

In this paper, we introduce an inversion scheme based on the Gauss-Newton algorithm, but with a significantly reduced numerical complexity. We make use of a "supershot" technique, where the superposition of several Green's functions is computed in single modeling jobs with simultaneously active sources. These constructs can be used to reduce the numerical load from the simulation of Green's functions and storage of the Jacobian for marine CSEM inversion. The scheme is effective when the number of source positions is much larger than the number of receiver positions, or vice versa. The use of such supershots has previously been introduced in seismic data imaging. In seismic prestack wave-equation migration, the use of source encoding techniques has been demonstrated to reduce the crosstalk noise following from the processing of simultaneously active sources (Morton and Ober, 1998; Jing et al., 2000; Romero et al., 2000). Such encoding techniques have also been used in seismic full-waveform inversion (see, e.g., Krebs et al., 2009; Boonyasiriwat and Schuster, 2010; Ben-Hadj-Ali et al., 2011; Bansal et al., 2013; Schiemenz and Igel, 2013). In this work, we will consider a related phase encoding technique for electromagnetic data inversion. The approach based on supershots can be particularly well-suited for CSEM data due to the strong attenuation of signal amplitude from dissipative propagation, which limits crosstalk.

The proposed inversion scheme is based on a low-rank approximation to the Gauss-Newton Hessian. However, the rank of the approximated Hessian is a parameter of the approach and will typically be much larger than the rank of quasi-Newton approximations to this quantity. In addition, the proposed higher rank approximation is not conditioned by past iterations, resulting in more accurate and larger updates. This allows more efficient inversions with higher convergence rates. In the inversion results shown in this paper, we will compare quasi-Newton inversion based on past iteration data for the Hessian approximation, Gauss-Newton inversion, and finally our proposed approaches to inversion. In our approach, we also do not modify the cost function or the gra-

dient computation, preserving the main structure of the Gauss-Newton scheme.

The Green's functions from the simultaneous source modeling can lead to a considerable compression of the Jacobian matrix. The inversion approach presented in this paper achieves further computational savings by using this fact, and in addition, we use a direct solution of the approximated Gauss-Newton equation. This solver avoids the explicit construction of a large (N^2) Hessian matrix, thus allowing inversion with a much larger number of free parameters.

In this paper, we first describe the theory of the low-rank approximation and how the data part of the Hessian is represented. We then analyze the error of the approximation. Next, we describe the solution of the Gauss-Newton equation with the low-rank data Hessian using a data Hessian matrix-free formulation. Then, we show inversion results and compare with a quasi-Newton approach and to a Gauss-Newton scheme with model parameter compression. Finally, we discuss the results and conclude.

THEORY

The inversion of CSEM data is formulated as an optimization problem:

$$\sigma(\mathbf{r}) = \arg \min_{\sigma \in \mathcal{M}} \varepsilon(\sigma), \quad (1)$$

where $\mathbf{r}$ defines the conductivity at position $\mathbf{r}$ of a 3D conductivity model in the set $\mathcal{M}$ of models compatible with a priori information and

$$\varepsilon(\sigma) = \varepsilon_D(\sigma) + \varepsilon_R(\sigma) \quad (2)$$

is the cost function. This cost function includes the regularization misfit term ε_R, and the data misfit term ε_D, which depends on the observations

$$\varepsilon_D(\sigma) = \sum_{F,i,f,\mathbf{r}_{rx},\mathbf{r}_{tx}} |W_i^F(\mathbf{r}_{rx}|\mathbf{r}_{tx}, f)\Delta F_i(\mathbf{r}_{rx}|\mathbf{r}_{tx}, f; \sigma)|^2. \quad (3)$$

Here, $\Delta F(\sigma) = F^{obs} - F^{synth}(\sigma)$ represents the difference between the observed and synthetic fields ($F = E$ for electric and $F = H$ for magnetic), W is a datum weight (typically inverse measurement uncertainty), i is the spatial components (x, y) of the field recordings, f is the frequencies, $\mathbf{r}_{rx}$ is a receiver position, and $\mathbf{r}_{tx}$ is a source position. The shorthand notation $\kappa = (F, i, f, \mathbf{r}_{rx}, \mathbf{r}_{tx})$ will uniquely label a measurement.

In this work, the nonlinear optimization problem is solved by iteratively updating the 3D conductivity model, following the Gauss-Newton method but with an approximate Hessian. At each iteration, the model update $\Delta\sigma$ is obtained by solving the linear equation system

$$\mathbf{H}\Delta\sigma = -\mathbf{g}, \quad (4)$$

where the Hessian is

$$\mathbf{H} = \mathbf{H}_D + \mathbf{H}_R, \quad (5)$$

and the gradient is

$$\mathbf{g} = \mathbf{g}_D + \mathbf{g}_R. \tag{6}$$

We will refer to $\mathbf{H}_D$ and $\mathbf{g}_D$ as the data Hessian matrix and the data gradient, which are derived from the data misfit term ε_D. The quantities $\mathbf{H}_R$ and $\mathbf{g}_R$ are the regularization Hessian matrix and the regularization gradient obtained from the regularization misfit term ε_R.

The data Hessian matrix and the data gradient vector are constructed from the Jacobian matrix $\mathbf{J}$ as follows:

$$(\mathbf{H}_D)_{\mathbf{r},\mathbf{r}'} = \sum_{\kappa} (\mathbf{J})_{\mathbf{r},\kappa}(\mathbf{J})_{\mathbf{r}',\kappa}^* + \text{c.c.}, \tag{7}$$

$$(\mathbf{g}_D)_{\mathbf{r}} = \sum_{\kappa} W_{\kappa} \Delta F_{\kappa}^* (\mathbf{J})_{\mathbf{r},\kappa} + \text{c.c.} \tag{8}$$

The superscript asterisk denotes complex conjugation, and the notation c.c. denotes the complex conjugate of the preceding expression. The Jacobian is a complex $N \times M$ matrix, where N is the number of model parameters, and M is the number of data samples. Note that in a Gauss-Newton approach, the Hessian in equation 7 is approximated by neglecting second-order derivative terms. The Jacobian can be constructed from Green's functions (Støren et al., 2008)

$$(\mathbf{J})_{\mathbf{r},\kappa} = W_i^F(\mathbf{r}_{\text{rx}}|\mathbf{r}_{\text{tx}}, f) \sum_p G_{p,i}^{F,J}(\mathbf{r}_{\text{rx}}|\mathbf{r}, f)$$
$$\times \sum_q G_{p,q}^{E,J}(\mathbf{r}|\mathbf{r}_{\text{tx}}, f)\hat{j}_q(\mathbf{r}_{\text{tx}}, f), \tag{9}$$

where $\mathbf{r}$ is the position in the conductivity model, $G_{p,q}^{F,J}$ denotes the Green's function for field F, component p, from a unit electric current source in direction q, and $\hat{j}_q$ is a vector component of the source dipole moment approximated as a point dipole $(p, q = x, y, z)$. When the field data have been normalized by the dipole moment, $\hat{\mathbf{j}}$ is a unit vector. It is straightforward to generalize the expression in equation 9 to the anisotropic and discrete case. From this expression, we see that explicit construction of any element $(\mathbf{r}, \kappa)$ of the Jacobian requires the two Green's functions associated with the receiver position $\mathbf{r}_{\text{rx}}$ and the source position $\mathbf{r}_{\text{tx}}$ to be simulated. Therefore, the total number of forward solutions (receiver simulations plus source simulations) needed for building the data Hessian matrix and the data gradient in a standard Gauss-Newton implementation is

$$N_{\text{Sim}} = \underbrace{N_F N_i N_{\text{rx}}}_{N_{\text{Sim}_{\text{rx}}}} + \underbrace{N_{\text{tx}}}_{N_{\text{Sim}_{\text{tx}}}}. \tag{10}$$

Here, and throughout, the notation N_a denotes the number of unique elements of index a; i.e., N_i is the number of spatial components of the field recorded at the receivers. The time-domain forward modeling code used computes the Green's function at all survey frequencies from a single simulation. Therefore, N_{Sim} in equation 10 does not scale with the number of frequencies N_f for our case.

The data Hessian matrix is a real $N \times N$ dense symmetric positive semidefinite matrix. The number of independent data samples (number of Jacobian columns)

$$M = N_F N_i N_f N_{\text{rx}} N_{\text{tx}} \tag{11}$$

could limit the maximum rank of the data Hessian matrix (Grayver et al., 2013), i.e.,

$$\text{rank}(\mathbf{H}_D) \leq \min(N, 2M). \tag{12}$$

In a 3D CSEM survey, the number of data samples M can be larger than the number of model parameters N.

The regularization $\varepsilon_R(\sigma)$ introduces a priori information about the solution model $\sigma(\mathbf{r})$ (Zhdanov, 2009) and is usually designed to give a positive definite Hessian that makes it feasible to solve the linear system in equation 4. In our implementation, the regularization can incorporate information on model conductivity $\varepsilon_{\text{ap mod}}$, model smoothness $\varepsilon_{\text{grad}}$, and conductivity anisotropy $\varepsilon_{\text{ap aniso}}$:

$$\varepsilon_R(\sigma) = \varepsilon_{\text{ap mod}}(\sigma) + \varepsilon_{\text{grad}}(\sigma) + \varepsilon_{\text{ap aniso}}(\sigma). \tag{13}$$

In this scheme, the regularization $\varepsilon_R(\sigma)$ is normalized by the number of model parameters, and the contributions to the total misfit for the regularization terms in equation 13 are balanced by the use of weights.

The regularization Hessian matrix derived from this scheme is

$$\mathbf{H}_R = \mathbf{H}_{\text{ap mod}} + \mathbf{H}_{\text{grad}} + \mathbf{H}_{\text{ap aniso}}, \tag{14}$$

where $\mathbf{H}_{\text{ap mod}}$ is a diagonal positive definite matrix, and the other two terms have off-diagonal structure.

Low-rank data Hessian matrix: Supershots and adjoint modeling

For a state-of-the-art 3D CSEM survey and with a realistic model representation, the numerical complexity involved with the construction of the data Hessian matrix $\mathbf{H}_D$ and the solution for the model update can be very large. The number of simulations required can be of order 10^5, and the dense linear system (see equation 4) can be of size $10^6 \times 10^6$ depending on the parameterization.

In this paper, we propose a low-rank approximation to the data Hessian matrix $\tilde{\mathbf{H}}_D$ in a Gauss-Newton scheme; i.e.,

$$\tilde{\mathbf{H}} \tilde{\Delta} \sigma = -\mathbf{g}, \tag{15}$$

where the approximate Hessian is

$$\tilde{\mathbf{H}} = \tilde{\mathbf{H}}_D + \mathbf{H}_R. \tag{16}$$

It is only the data Hessian that is approximated; the gradient and cost functions are not affected.

To compute the matrix $\tilde{\mathbf{H}}_D$, we consider a low-rank approximation in which sources in equation 9 are combined in superpositions, after being encoded with a random phase factor. The result is a contraction of Jacobian columns that can be written as

$$(\tilde{\mathbf{J}})_{\mathbf{r},\tilde{\kappa}} = \sum_{\mathbf{r}_{\mathrm{tx}} \in s} e^{i\phi_\kappa} (\mathbf{J})_{\mathbf{r},\kappa} \qquad (17)$$

for a group of source positions s, and where ϕ_κ is uniformly distributed random numbers in the interval $[0, 2\pi)$. The shorthand notation $\tilde{\kappa} = (F, i, f, \mathbf{r}_{\mathrm{rx}}, s)$ will uniquely label a source-group simulation.

The number of source groups N_s and the grouping scheme are discussed below. Following this approach, the quantities in equation 17 associated with the source positions in a source group and a specific receiver field-component are calculated from a single simultaneous-source (supershot) forward solution of the Maxwell equations. We denote the output of this simulation as

$$\tilde{G}^F_{p,i,\mathbf{r}_{\mathrm{rx}},s}(\mathbf{r}, f) = \sum_{q, \mathbf{r}_{\mathrm{tx}} \in s} G^{E,J}_{p,q}(\mathbf{r}|\mathbf{r}_{\mathrm{tx}}, f) W^F_i(\mathbf{r}_{\mathrm{rx}}|\mathbf{r}_{\mathrm{tx}}, f) e^{i\phi_\kappa} \hat{j}_q(\mathbf{r}_{\mathrm{tx}}, f),$$
$$(18)$$

where $W^F_i(\mathbf{r}_{\mathrm{rx}}|\mathbf{r}_{\mathrm{tx}}, f) e^{i\phi_\kappa} \hat{j}_q(\mathbf{r}_{\mathrm{tx}}, f)$ is the distributed source strength that corresponds to the receiver field component. For each source group, we need to simulate all receivers, fields, and components, so that the number of simulations is

$$N_{\mathrm{Sim}_s} = N_F N_i N_{\mathrm{rx}} N_s. \qquad (19)$$

With respect to the total number of simulations, the quantity N_{Sim_s} above will replace $N_{\mathrm{Sim}_{\mathrm{tx}}}$ in equation 10 when supershots are applied. The supershot simulations in equation 18 define an approximate Jacobian $\tilde{\mathbf{J}}$ as

$$(\tilde{\mathbf{J}})_{\mathbf{r},\tilde{\kappa}} = \sum_p G^{F,J}_{p,i}(\mathbf{r}_{\mathrm{rx}}|\mathbf{r}, f) \tilde{G}^F_{p,i,\mathbf{r}_{\mathrm{rx}},s}(\mathbf{r}, f), \qquad (20)$$

which in turn defines an approximate data Hessian matrix $\tilde{\mathbf{H}}_{\mathrm{D}}$:

$$(\tilde{\mathbf{H}}_{\mathrm{D}})_{\mathbf{r},\mathbf{r}'} = \sum_{\tilde{\kappa}} (\mathbf{J})_{\mathbf{r},\tilde{\kappa}} (\mathbf{J})^*_{\mathbf{r}',\tilde{\kappa}} + \mathrm{c.c.}, \qquad (21)$$

analogous to a Gauss-Newton Hessian. The number of columns (data samples) in the approximate Jacobian is

$$\tilde{M} = N_F N_i N_f N_{\mathrm{rx}} N_s. \qquad (22)$$

Here, N_s is typically selected small (i.e., of the order of 1–10 as discussed below), so $\tilde{M} \ll N$, and therefore

$$\mathrm{rank}(\tilde{\mathbf{H}}_{\mathrm{D}}) \leq 2\tilde{M}. \qquad (23)$$

Comparing equation 23 with equation 12, we see that this approach results in a low-rank data Hessian matrix approximation $\tilde{\mathbf{H}}_{\mathrm{D}}$. Note that the number of forward solutions (compare equations 10 and 19) and the storage requirement to construct $\tilde{\mathbf{H}}_{\mathrm{D}}$ (compare equations 11 and 22) scale by N_s instead of N_{tx} as for $\mathbf{H}_{\mathrm{D}}$. The reduction in numerical complexity from the approximation is described below. For now, we just note that realistic values of the ratio N_{tx}/N_s can be of order 10–100, i.e., a dramatic decrease in complexity.

The number of source groups determines the accuracy of the approximation. In the limit, where there is only one source per group, no approximation is made and the Gauss-Newton Hessian is recovered. We therefore consider N_s to be a tuning parameter controlling the accuracy.

The data gradient $\mathbf{g}_{\mathrm{D}}$ is computed following the adjoint scheme described in Støren et al. (2008), which does not involve any approximation of the expression for the gradient in equation 8. Following this scheme, we compute

$$(\mathbf{g}_{\mathrm{D}})_{\mathbf{r}} = \sum_{v=F,i,f,\mathbf{r}_{\mathrm{rx}},p} \mathrm{sign}(F) G^{F,J}_{p,i}(\mathbf{r}|\mathbf{r}_{\mathrm{rx}}, f) L_v(\mathbf{r}) + \mathrm{c.c.}, \quad (24)$$

where $\mathrm{sign}(F)$ is $+1$, when F is a magnetic field, and -1, when F is an electric field, and

$$L_v(\mathbf{r}) = \sum_{q, \mathbf{r}_{\mathrm{tx}}} G^{E,J}_{p,q}(\mathbf{r}|\mathbf{r}_{\mathrm{tx}}, f) [W^F_i(\mathbf{r}_{\mathrm{rx}}|\mathbf{r}_{\mathrm{tx}}, f)]^2 \Delta F \hat{j}_q(\mathbf{r}_{\mathrm{tx}}, f).$$
$$(25)$$

The Green's function $G^{F,J}_{p,i}$ in equation 24 is the same as the one used for the approximated Jacobian $\tilde{\mathbf{J}}$ in equation 20. To compute $L_v(\mathbf{r})$, it is necessary to run additional simulations

$$N_{\mathrm{Sim}_{\mathrm{Adj}}} = N_F N_i N_{\mathrm{rx}}, \qquad (26)$$

because the source strength factors are different in the superpositions in equations 18 and 25.

The total number of forward solutions needed to compute the approximated data Hessian matrix in equation 21 and the data gradient in equation 24 in the low-rank approach is

$$\tilde{N}_{\mathrm{Sim}} = N_{\mathrm{Sim}_{\mathrm{rx}}} + N_{\mathrm{Sim}_s} + N_{\mathrm{Sim}_{\mathrm{Adj}}}$$
$$= N_F N_i N_{\mathrm{rx}} (N_s + 2). \qquad (27)$$

For typical modern 3D CSEM surveys, N_{Sim} is dominated by the number of sources N_{tx}. We thus see a decrease in the number of forward solutions compared with Gauss-Newton inversion, whenever $N_F N_i N_{\mathrm{rx}}(N_s + 2)/N_{\mathrm{tx}} < 1$, where factors N_F and N_i are of order unity. Note that the number N_{Sim_s} from equation 19 corresponds to the number of additional forward solutions required by the low-rank approach compared with gradient-based approaches, such as quasi-Newton and nonlinear conjugate gradients (CG).

Source grouping strategies

We will now discuss the selection of source points for the groups introduced in equation 17. We will consider the three different strategies shown schematically in Figure 1, where all source positions are arranged in four groups.

In Figure 1a, the sources in each group are selected, such that each group covers a contiguous area. In Figure 1b, the sources are grouped randomly, and in Figure 1c, the groups are created by maximizing the distance between the sources in each group.

Numerical studies described below show that using groups with more distant sources, as in Figure 1c, gives the best result. This is in agreement with the qualitative argument in the next section, which predicts that approximation errors decay with increasing separation between simultaneous sources in equation 18. Note that the specific grouping illustrated in the example Figure 1c is only approximately optimal, because further increase in separation could be possible.

To illustrate some properties of different grouping strategies, we will consider synthetic data from the model and source-receiver layout shown in Figure 2. The example survey includes 25 receivers, five towlines (sampling 330 source positions in total), and a thin resistor at 50 ohm-m. The background resistivity is 1 ohm-m, and the water resistivity is 0.25 ohm-m. The input synthetic data for the study were the E_x and E_y fields at frequencies 0.25 and 1.0 Hz, respectively, without noise.

In Figure 3, we show the eigenvalue distribution for the data Hessian from the three different grouping strategies described above, and the Gauss-Newton Hessian. The eigenvalue distribution for the distant-sources grouping is in closest agreement with the Gauss-Newton data Hessian eigenvalue distribution, and the grouping based on random selections is also similar to the grouping based on distant sources. The number of forward simulations is the same for the different grouping strategies, and identical to the rank of the approximated data Hessian, $\mathrm{rank}(\tilde{\mathbf{H}}_D) = 600$. The number of source groups was $N_s = 3$.

As mentioned, the number of source groups controls the accuracy and computational cost of the approximation and is thus a tuning parameter. In particular, the value of this parameter determines the level of crosstalk noise (discussed below). Moreover, the rank of the approximate data Hessian $\tilde{\mathbf{H}}_D$ is proportional to the number of groups. In Figure 4, we show an example, based on the survey shown in Figure 2, of how increasing the number of groups increases the rank of the low-rank data Hessian in equation 21. We use the grouping strategy based on distant sources described above. From Figure 4, we see how increasing N_s leads to an increase in the number of nonzero eigenvalues, and an eigenvalue distribution closer to that of the Gauss-Newton data Hessian $\mathbf{H}_D$.

In a typical application of the 3D CSEM Gauss-Newton inversion scheme, the regularization terms in the cost function at equation 13 will stabilize the solution of the linear system in equation 4, and limit the condition number of the Hessian matrix $\mathbf{H}$. Therefore, though the matrix $\tilde{\mathbf{H}}_D$ is typically rank deficient as shown in Figures 3 and 4, the eigenvalue distribution of the matrix in equation 4 that determines the update for the low-rank approximation can be close to that of the original Gauss-Newton scheme. Figure 5 demonstrates this for the survey shown in Figure 2. Note that the magnitude scale of the eigenvalues in Figure 5 is different than in Figures 3 and 4 due to a scaling factor applied to balance the contributions of data misfit and regularization in the cost function. The black curve (E6) shows how the a priori regularization acts to limit

the smallest matrix eigenvalue for the low-rank approximation with $N_s = 3$. For the small and largest eigenvalues, the distribution is similar to that following from the Gauss-Newton Hessian (green curve E1). The magenta curve (E4) shows the eigenvalues, when only the data Hessian is considered. When we increase the number of source groups, it is possible to obtain a regularized low-rank Hessian that has a similar eigenvalue distribution as a Gauss-Newton Hessian (compare the red curve E2 with the yellow curve E5).

For each iteration of the inversion, it is possible to vary the sources that are included in each group, while keeping the source-grouping strategy. Once the groups of sources are established, one forward solution per group of sources and receiver field component is performed (a total of $N_F N_i N_{rx} N_s$ forward solutions),
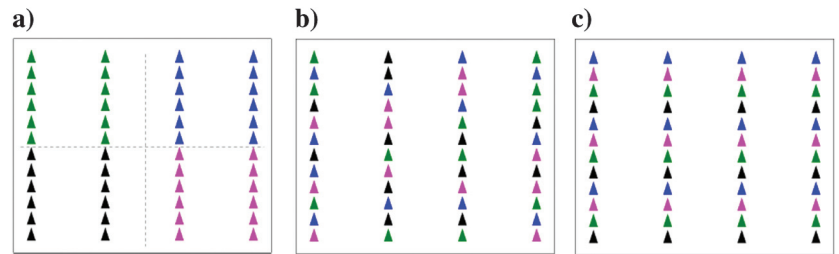

Figure 1. Examples of different strategies for source grouping. The triangles represent source positions, and the sources with the same color are grouped. (a) Four groups of sources gathering closest ones in a single group simulation. (b) Four groups of sources gathering the different sources randomly in a single group simulation. (c) Four groups of sources gathering distant sources in a single group simulation.
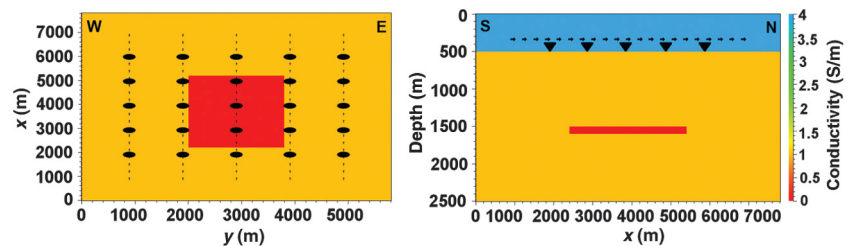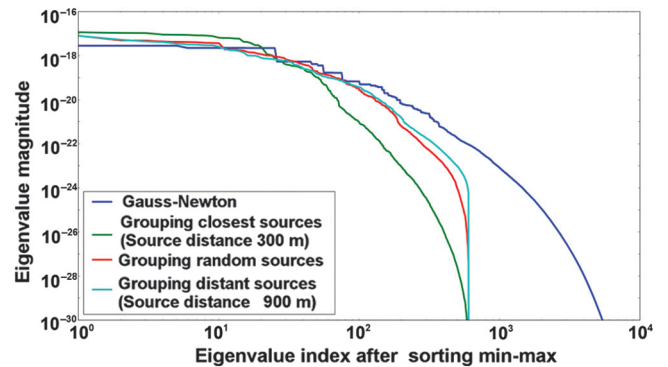

Figure 2. Survey layout for synthetic study.

Figure 3. Example of eigenvalue distribution for different source-grouping strategies. We use the same number of simulations in the three different grouping strategies. The eigenvalues are shown in sorted order from the largest to the smallest along the horizontal axis.
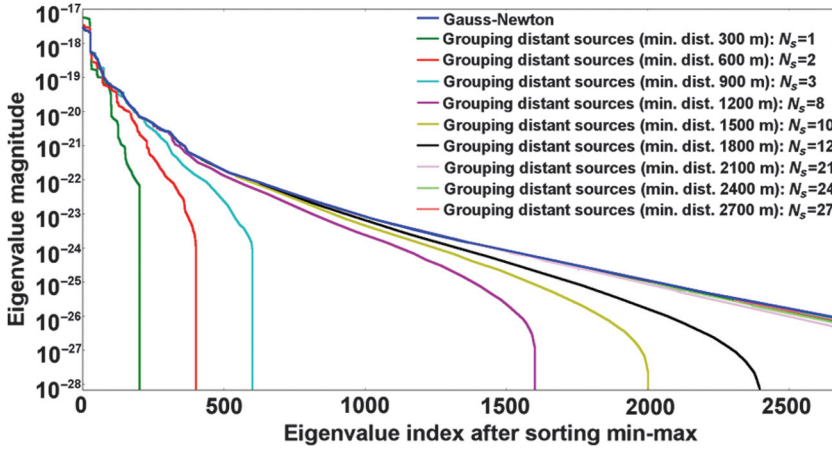
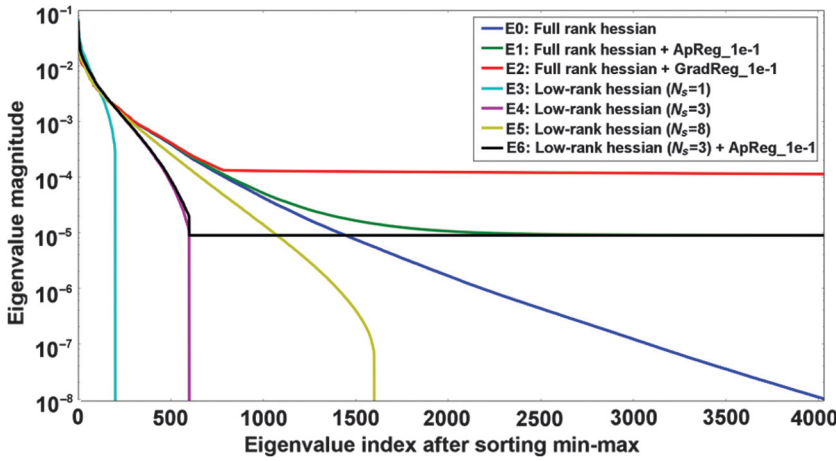Figure 4. Example of eigenvalue distribution for different number of distant-source groups N_s.

Figure 5. Hessian matrix **H** eigenvalue distributions. Blue curve E0: Gauss-Newton data Hessian $\mathbf{H}_D$ only. Green curve E1: Gauss-Newton data Hessian and regularization $\mathbf{H}_D + \mathbf{H}_{ap\,mod}$. Red curve E2: Gauss-Newton data Hessian and regularization $\mathbf{H}_D + \mathbf{H}_{grad}$. Cyan E3, purple E4, and yellow E5 curves: low-rank data Hessian $\mathbf{H}_D$ only for 1, 3, and 8 source groups. Black curve E6: low-rank data Hessian and regularization $\mathbf{H}_D + \mathbf{H}_{ap\,mod}$.
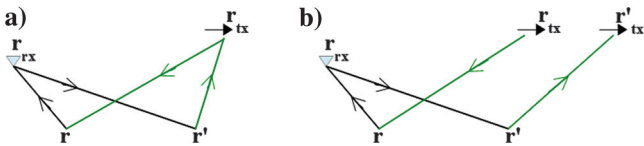

Figure 6. Diagrams representing the two types of terms included in the data Hessian matrix of a low-rank approximation (equation 21). Each source position (subscript "tx") is associated with its corresponding source strength factors (see discussion following equation 18). In these diagrams, an arrow from, e.g., position $\mathbf{r}_{tx}$ to $\mathbf{r}$ represents a Green's function $G_{p,q}^{F,J}(\mathbf{r}|\mathbf{r}_{tx}, f)$, which is a factor in the corresponding term of the Hessian. "Reverse time" arrows, where the origin for the propagation represented by arrows is located at a receiver position (subscript "rx") appear with complex conjugation in the Hessian expressions. (a) Diagram representing those terms of $\tilde{\mathbf{H}}_D$ that also make up the data Gauss-Newton Hessian (equation 7). (b) Diagram representing the additional crosstalk noise terms introduced into $\tilde{\mathbf{H}}_D$, where two different source positions contribute.

using $W_i^F(\mathbf{r}_{rx}|\mathbf{r}_{tx}, f)e^{i\phi_\kappa}$ as the source strength for each source in a group s. Linearity of the Maxwell equations implies that this is equivalent to the summation of individual terms in equation 18. The random phase factors $e^{i\phi_\kappa}$ are sampled independently for ϕ_κ in each group $\kappa_s = (F, i, f, \mathbf{r}_{rx}, \mathbf{r}_{tx} \in s)$.

Noise analysis and phase encoding

Let us now consider the errors introduced by the approximation of the Hessian in equation 21. When the data Hessian is constructed according to equation 7, each term contributing is computed from the Green's functions associated with one receiver and one source position. The same terms appear when computing the low-rank approximation (equation 21), but in addition, the approximation introduces terms (crosstalk terms), which are not part of the Gauss-Newton data Hessian. This is illustrated by the diagrams in Figure 6, where Figure 6a shows a low-rank approximation contribution from one source position, which is the same as that which also appears in a standard Gauss-Newton data Hessian computation, and Figure 6b shows one of the additional terms introduced through the low-rank approximation.

The terms of Figure 6a involve only one source position and one receiver position associated with a single measurement.

Due to the summation over source positions in $\tilde{G}$, the approximation will also include terms involving two different source positions (but the same receiver channel, e.g., $E_x(\mathbf{r}_{rx}|\mathbf{r}_{tx}, f)$) in the product in equation 21. These terms are shown as Figure 6b. We will refer to these diagrams as crosstalk and denote their contribution η, such that

$$\tilde{\mathbf{H}}_D = \mathbf{H}_D + \eta. \tag{28}$$

Though the crosstalk terms will represent an approximation error, we note that η has specific matrix properties. The matrix η will be symmetric, and from the definition, we have that

$$\text{rank}(\tilde{\mathbf{H}}_D - \eta) = \text{rank}(\mathbf{H}_D). \tag{29}$$

These properties could be used in schemes to reduce the effect of η in the approximation. In Figure 7, we show the diagonal (i.e., matrix elements for indices $\sigma_V(\mathbf{r}), \sigma_V(\mathbf{r})$ and $\sigma_H(\mathbf{r}), \sigma_H(\mathbf{r})$, where subscripts V and H pertain to vertical and horizontal component) of the matrices $\mathbf{H}_D$ and η for two different number of groups. These results were computed using a half-space model. Comparing Figure 7b and 7c, we note that the crosstalk (approximation noise) is reduced when we increase the number of source groups from one group to three. An important fact is that the Hessian magnitude exceeds the magnitude of the crosstalk. We study these two phenomena later in this section.

The large-scale structure of the crosstalk seen in Figure 7 could be detrimental to the inversion, convergence but it is hard to suppress. The synthetic data inversion results in Appendix A show that an inversion result of similar quality as in the Gauss-Newton approach can be achieved despite the crosstalk noise. As an additional observation, we note that the diagonal crosstalk noise was reduced more efficiently by increasing the number of source groups in the horizontal rather than the vertical components as seen in Figure 7b and 7c.

The phase encoding in the supershots in equation 18 is a way to reduce the errors due to crosstalk. Noise reduction approaches based on random phase encoding have also been applied in seismic modeling (see, e.g., Bansal et al., 2013). In equation 21, the random phase factors $e^{i\phi_\kappa}$ will cancel in the terms, where a single source position is involved, as in the terms for the Gauss-Newton Hessian (Figure 6a). This is because the source Green's functions appearing in $\tilde{\mathbf{H}}_D$ are included in combinations:

$$G_{p,q}^{E,J}(\mathbf{r}|\mathbf{r}_{\mathrm{tx}},f)e^{i\phi_{E,i,\mathbf{r}_{\mathrm{tx}}}}\ [G_{p,q}^{E,J}(\mathbf{r}'|\mathbf{r}_{\mathrm{tx}},f)e^{i\phi_{E,i,\mathbf{r}_{\mathrm{tx}}}}]^*. \quad (30)$$

For these terms, the random phase cancels, and phase encoding does not affect the diagrams corresponding to the Gauss-Newton data Hessian.

The crosstalk terms result in contributions, where the source Green's functions for two different positions $\mathbf{r}_{\mathrm{tx}}$ and $\mathbf{r}'_{\mathrm{tx}}$ appear in combinations of the type,

$$G_{p,q}^{E,J}(\mathbf{r}|\mathbf{r}_{\mathrm{tx}},f)[G_{p,q}^{E,J}(\mathbf{r}'|\mathbf{r}'_{\mathrm{tx}},f)]^*\ e^{i\phi_{E,i,\mathbf{r}_{\mathrm{tx}}}-i\phi_{E,i,\mathbf{r}'_{\mathrm{tx}}}}. \quad (31)$$

The random phase factor in this expression will act to reduce the total contribution of the crosstalk terms. To analyze this, let us first assume that the magnitudes of the crosstalk terms in η are Gaussian distributed. In such a case, the resulting magnitude of the sum of the crosstalk terms will be governed by the properties of a Gaussian random walk. If we assume that there are A terms contributing, then the magnitude $|\eta|$ will scale as $\sim\sqrt{A}$ instead of linear scaling due to the random phases in equation 31.

The number of terms (Figure 6a) contributing to $\mathbf{H}_D$ is equal to the different number of combinations (equation 11) of equation 30 that appear in equation 21; i.e., it is proportional to N_{tx}. The number of crosstalk terms (Figure 6b), that contribute to η, scales with the number of all possible combinations of two different sources (equation 31) included in a group simulation that appear in equation 21, i.e., scales as N_{tx}^2 (assuming a single source group $N_s = 1$). However, the magnitude $|\eta|$ should still scale by N_{tx} by analogy to a Gaussian random walk as described above, i.e., $|\eta| \sim N_{\mathrm{tx}}$.

Let us now consider the details of the distribution of crosstalk magnitudes. The physics of the problem indicates that the magnitude distribution of the crosstalk terms should be more centered than a Gaussian distribution. The exponential decay of the magnitude of the Green's functions will result in negligible contributions from the

product of Green's functions in equation 31, when the distance $|\mathbf{r}_{\mathrm{tx}} - \mathbf{r}'_{\mathrm{tx}}|$ is large. This will be the case for many of the terms making up η. The spatial decay of the Green's functions is shown in Figure 8.

Thus, the attenuation of the magnitude for the contributions makes $|\eta|$ scale by the number of sources in a source group such as $(N_{\mathrm{tx}})^\alpha$ with $\alpha < 1$, and the asymptotic behavior of the approximation is given as $\lim_{N_{\mathrm{tx}}\to\infty}|\eta|/|\mathbf{H}_D| = 0$.

Next, we turn our attention to the magnitude of the individual terms contributing to $\mathbf{H}_D$ relative to the crosstalk terms in η. First, we note that for each crosstalk contribution, there is a contribution to the data Hessian with larger magnitude. The elements of the Jacobian are Fréchet derivatives computed from a product of a Green's function associated with a receiver position and a source position (see equation 20). Magnitudes of Fréchet derivatives (formed by the Green's functions shown in Figure 8) are shown in Figure 9. Contributions to $\tilde{\mathbf{H}}_D$ are products of two such Fréchet derivatives. A crosstalk contribution is the product of two Frechet derivatives (with complex conjugation on one factor) corresponding to two different source positions, e.g., the product of Rx1-Tx5 by Rx1-Tx3 of Figure 9. The terms contributing to $\mathbf{H}_D$ are products of a Fréchet derivative by its complex conjugate, e.g., the product of Rx1-Tx3 by the complex conjugate. Using the magnitudes represented in Figure 9 as a reference, we see that this contribution to $\mathbf{H}_D$ will dominate over the crosstalk contribution from Rx1-Tx5 by Rx1-Tx3. This is true in general: a term with a single source will dominate over the crosstalk terms sharing a source.
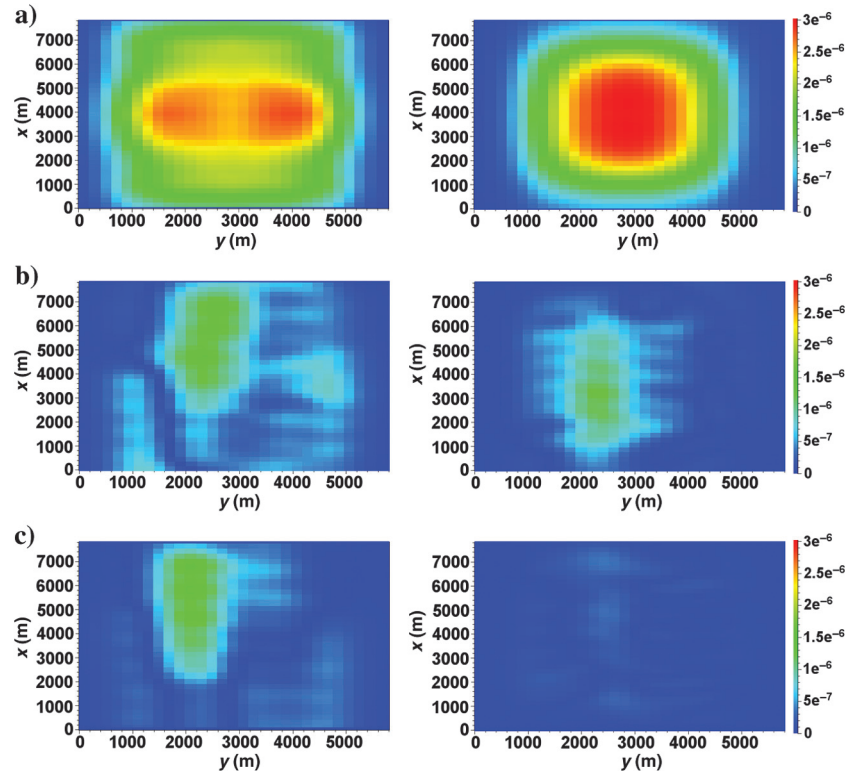

Figure 7. Plots of the Hessian and crosstalk matrix diagonals at depth 1500 m for the survey shown in Figure 2. Vertical conductivity component on the left and horizontal conductivity component on the right. These results were computed in a half-space model.

We can reduce the error of the approximation by increasing the number of source groups N_s. In fact, as we showed above, using the maximum $N_s = N_{tx}$ makes $\tilde{\mathbf{H}}_D$ identical to $\mathbf{H}_D$, but in this case, there is no reduction in computational cost. We also demonstrated that we can optimize the approximation by constructing the source groups with maximum separation between the spatial locations of sources.

This supports the numerical results, where source groups based on the maximum distance between sources were found to give the highest accuracy, as is discussed in the "Results" section. The distance between sources determines the number of groups, and can

therefore be considered as the tuning parameter for the accuracy of the approximation.

The approach described in this paper will retain the same cost function ϵ and gradient $\mathbf{g}$ as is used in a standard Gauss-Newton scheme. The approximation only introduces error into the Hessian in the Gauss-Newton equation, which results in an approximate search direction $\tilde{\boldsymbol{\Delta}}\boldsymbol{\sigma}$. However, at early iterations of the inversion, when the cost function is large and far from the minimum, the parabolic assumption of the Gauss-Newton scheme will be poor and the accuracy of the computed search direction need not be very good. Treating the number of source groups as a tunable parameter in our approach, we may then improve the accuracy of the approximation dynamically as the inversion gets closer to the minimum and improved search direction accuracy is required to maintain good convergence rate.

Data Hessian matrix-free recursive direct solver

As we mentioned in the "Introduction" section, the main challenge of implementing a 3D CSEM Gauss-Newton is the computational cost. The low-rank approximation to the data Hessian matrix described above can reduce the number of forward solutions, as well as the size of the Jacobian by a factor of 10–100. The size of the linear equation system for the model update in equation 15 will, however, scale quadratically by the number of free parameters of the inversion if the Hessian is constructed explicitly, and this will be a dense matrix. For a typical 3D CSEM problem, the number of parameters considered in quasi-Newton inversion schemes can be of order 10^7. In such a case, a parameter compression strategy that can reduce the order of magnitude of the parameter number by 10–100 is required to solve the numerical linear algebra problem. For the low-rank approximation, we can however make use of the implicit matrix representation in equation 21 to construct a data Hessian matrix-free solver. The feasibility of the resulting direct solver is determined by the sparsity of the regularization Hessian matrix and the smaller size of the Jacobian in the low-rank approach.

From the Gauss-Newton equation 15, the update at each iteration is obtained as $\tilde{\boldsymbol{\Delta}}\boldsymbol{\sigma} = -\tilde{\mathbf{H}}^{-1}\mathbf{g}$, where the inverted matrix has the following structure:
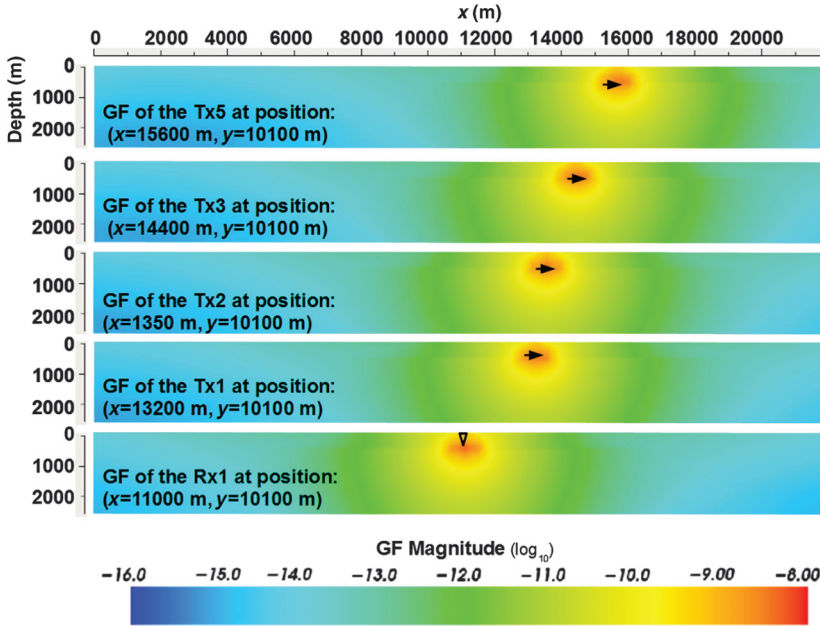

Figure 8. Green's function amplitude at different source (Tx) and receiver (Rx) positions, for the electric field E_x at frequency $f = 0.25$ Hz, when using the survey layout represented at Figure 2 for a larger model with the same background and water conductivity but without the 50 ohm-m target.
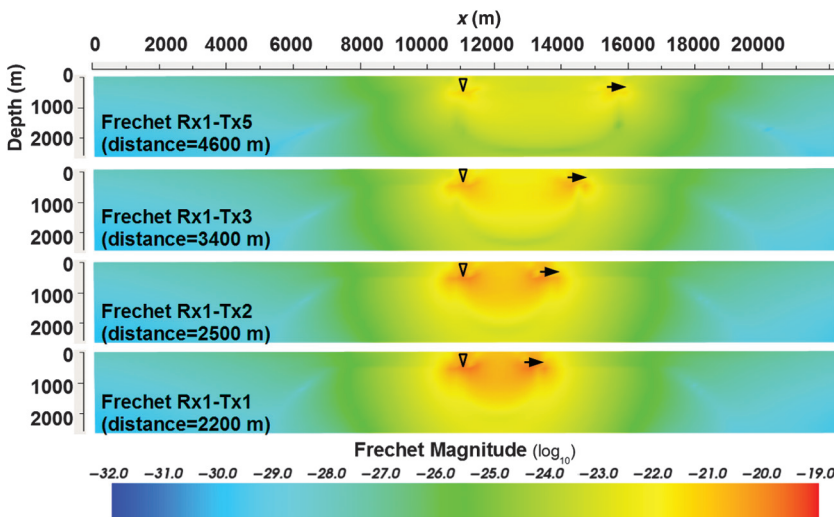
$$\tilde{\mathbf{H}}^{-1} = \left(\mathbf{H}_R + \sum_{m=1}^{\tilde{M}} (\tilde{\mathbf{J}}_m \tilde{\mathbf{J}}_m^\dagger + \text{c.c.}) \right)^{-1},$$

(32)

where $\tilde{M}$ is the number of approximated Jacobian columns as in equation 22, and $\tilde{\mathbf{J}}_m$ represents the mth column. The superscript $\dagger$ symbol represents the conjugate transpose. Introducing the shorthand notation

Figure 9. Fréchet derivatives computed from the Green's functions in Figure 8.

$$\tilde{\mathbf{J}}_{m_{\text{Re}}} = \text{Re}(\sqrt{2}\tilde{\mathbf{J}}_m), \qquad \tilde{\mathbf{J}}_{m_{\text{Im}}} = \text{Im}(\sqrt{2}\tilde{\mathbf{J}}_m), \qquad (33)$$

we can rewrite equation 32 in the following form:

$$\tilde{\mathbf{H}}^{-1} = \left(\mathbf{H}_R + \sum_{j=1}^{2\tilde{M}} (\tilde{\mathbf{J}}_j \tilde{\mathbf{J}}_j^T) \right)^{-1}, \qquad (34)$$

with $\tilde{\mathbf{J}}_1 = \tilde{\mathbf{J}}_{1_{\text{Re}}}$, $\tilde{\mathbf{J}}_2 = \tilde{\mathbf{J}}_{1_{\text{Im}}}$, $\tilde{\mathbf{J}}_3 = \tilde{\mathbf{J}}_{2_{\text{Re}}}$, and so on. As described in Press et al. (2002), the matrix inverse can be computed recursively applying the Sherman-Morrison formula:

$$(\mathbf{A} + \mathbf{u}\mathbf{u}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{u}^T\mathbf{A}^{-1}}{1 + \mathbf{u}^T\mathbf{A}^{-1}\mathbf{u}}. \qquad (35)$$

To compute $\tilde{\boldsymbol{\Delta}}\sigma$, we have implemented a recursive solver using the Sherman-Morrison formula, which is explained below. At each iteration, we incorporate a single approximated Jacobian column $\tilde{\mathbf{J}}_j$ that has been computed on the fly to avoid storing the whole $\tilde{\mathbf{J}}$ matrix. At iteration j, it is only necessary to keep in memory the sparse $\mathbf{H}_R$, the so-called iteration vectors $\mathbf{w}_j$ and the scalars c_j. The algorithm is described below in three steps.

Initialization,

$$\tilde{\mathbf{H}}_{(0)}^{-1} = \mathbf{H}_R^{-1}. \qquad (36)$$

First iteration ($j = 1$),

$$\begin{aligned}
\tilde{\mathbf{H}}_{(1)}^{-1} &= (\tilde{\mathbf{H}}_{(0)} + \tilde{\mathbf{J}}_1\tilde{\mathbf{J}}_1^T)^{-1} \\
&= \tilde{\mathbf{H}}_R^{-1} - \frac{\tilde{\mathbf{H}}_R^{-1}\tilde{\mathbf{J}}_1\tilde{\mathbf{J}}_1^T\tilde{\mathbf{H}}_R^{-1}}{1 + \tilde{\mathbf{J}}_1^T\tilde{\mathbf{H}}_R^{-1}\tilde{\mathbf{J}}_1} \\
&= \tilde{\mathbf{H}}_R^{-1} - c_1\mathbf{w}_1\mathbf{w}_1^T, \qquad (37)
\end{aligned}$$

$$\mathbf{w}_1 = \tilde{\mathbf{H}}_R^{-1}\tilde{\mathbf{J}}_1, \qquad (38)$$

$$c_1 = (1 + \tilde{\mathbf{J}}_1^T\mathbf{w}_1)^{-1}. \qquad (39)$$

The Sherman-Morrison formula, equation 35, is applied in the second step of equation 37. Second iteration ($j = 2$),

$$\begin{aligned}
\tilde{\mathbf{H}}_{(2)}^{-1} &= (\tilde{\mathbf{H}}_{(1)} + \tilde{\mathbf{J}}_2\tilde{\mathbf{J}}_2^T)^{-1} \\
&= \tilde{\mathbf{H}}_{(1)}^{-1} - \frac{\tilde{\mathbf{H}}_{(1)}^{-1}\tilde{\mathbf{J}}_2\tilde{\mathbf{J}}_2^T\tilde{\mathbf{H}}_{(1)}^{-1}}{1 + \tilde{\mathbf{J}}_2^T\tilde{\mathbf{H}}_{(1)}^{-1}\tilde{\mathbf{J}}_2} \\
&= \tilde{\mathbf{H}}_{(1)}^{-1} - c_2\mathbf{w}_2\mathbf{w}_2^T \\
&= \tilde{\mathbf{H}}_R^{-1} - c_1\mathbf{w}_1\mathbf{w}_1^T - c_2\mathbf{w}_2\mathbf{w}_2^T, \qquad (40)
\end{aligned}$$

$$\mathbf{w}_2 = \tilde{\mathbf{H}}_{(1)}^{-1}\tilde{\mathbf{J}}_2 = \tilde{\mathbf{H}}_R^{-1}\tilde{\mathbf{J}}_2 - c_1\mathbf{w}_1\mathbf{w}_1^T\tilde{\mathbf{J}}_2, \qquad (41)$$

$$c_2 = (1 + \tilde{\mathbf{J}}_2^T\mathbf{w}_2)^{-1}. \qquad (42)$$

When generalizing for iteration j,

$$\tilde{\mathbf{H}}_{(j)}^{-1} = \tilde{\mathbf{H}}_{(j-1)}^{-1} - c_j\mathbf{w}_j\mathbf{w}_j^T = \tilde{\mathbf{H}}_{(0)}^{-1} - \sum_{q=1}^{j} c_q\mathbf{w}_q\mathbf{w}_q^T, \qquad (43)$$

$$\mathbf{w}_j = \tilde{\mathbf{H}}_{(j-1)}^{-1}\tilde{\mathbf{J}}_j = \tilde{\mathbf{H}}_{(0)}^{-1}\tilde{\mathbf{J}}_j - \sum_{q=1}^{j-1} c_q\mathbf{w}_q\mathbf{w}_q^T\tilde{\mathbf{J}}_q, \qquad (44)$$

$$c_j = (1 + \tilde{\mathbf{J}}_j^T\mathbf{w}_j)^{-1}. \qquad (45)$$

At the final iteration $j = 2\tilde{M}$, we obtain the solution

$$\tilde{\mathbf{H}}^{-1} = \tilde{\mathbf{H}}_{(2\tilde{M})}^{-1}, \qquad (46)$$

and then the approximated model update is obtained as

$$\tilde{\boldsymbol{\Delta}}\sigma = -\left(\tilde{\mathbf{H}}_{(0)}^{-1}\mathbf{g} - \sum_{j=1}^{2\tilde{M}} c_j\mathbf{w}_j\mathbf{w}_j^T\mathbf{g} \right). \qquad (47)$$

We note that it is also possible to omit the explicit construction of the inverse regularization Hessian matrix, $\mathbf{H}_R^{-1} = \tilde{\mathbf{H}}_{(0)}^{-1}$. If we provide a method to compute solutions of the system $\mathbf{H}_R\mathbf{x} = \mathbf{v}$, then all matrix-vector products involving $\tilde{\mathbf{H}}_{(0)}^{-1}$, such as in equations 44 and 47 can be obtained without the explicit inverse. Because the regularization will be a sparse matrix (being a banded matrix for the case of the regularization described in equation 13) and these solutions will be required many times, a direct solver will be efficient: If a factorization is carried out at the initialization step, all subsequent computations of $\mathbf{H}_{(0)}^{-1}\mathbf{v}$ can be obtained very quickly from the direct solver.

The main benefit of solving the linear system in equation 15 as described here is that the memory complexity is smaller than by keeping the dense $\tilde{\mathbf{H}}$ matrix in memory. The memory complexity for this Sherman-Morrison formula-based recursive solver is $\mathcal{O}(N\tilde{M})$, where $\tilde{M} \ll N$ for a realistic example with a small N_s. Compared with approaches where the Hessian is constructed explicitly, it becomes feasible to solve linear systems in equation 15 with the same number of parameters as is typically used in quasi-Newton approaches.

Approximate inverse regularization Hessian contribution

In the example shown in this paper, we introduced a further approximation to simplify the data Hessian matrix-free solver described above. We suppress off-diagonal elements of the sparse regularization Hessian matrix,

$$\tilde{\mathbf{H}}_R = \text{diag}(\mathbf{H}_R), \qquad (48)$$

and use the Hessian

$$\hat{\mathbf{H}} = \tilde{\mathbf{H}}_D + \tilde{\mathbf{H}}_R. \qquad (49)$$

The approximated regularization Hessian matrix $\tilde{\mathbf{H}}_R$ allows us to explicitly construct the inverse of the regularization contribution to the Hessian for the linear equation solver in equation 36.

The a priori model regularization matrix $\mathbf{H}_{apmod}$ is not modified when it is used in the approximated regularization Hessian matrix (equation 48) because $\mathbf{H}_{ap\,mod}$ is diagonal by definition. For the gradient smoothness $\mathbf{H}_{grad}$, the approximation will affect the resulting update, but because we are carrying out iterations of the nonlinear optimization, the lack of information can be compensated at later iterations.

INVERSION RESULTS

In this section, we present and compare the inversion results from a recent 3D CSEM field data set. The geologic complexity and the scale of the acquisition of the field data make this a challenging and interesting survey to use to test the performance of inverse schemes. We consider four different schemes: (1) quasi-Newton limited-memory Broyden-Fletcher-Golfarb-Shanno with boundaries (L-BFGS-B) (Zhu et al., 1997), (2) 3D Gauss-Newton using model parameter compression and a CG solver for equation 4, (3) the proposed low-rank approach using model parameter compression and a CG solver for equation 15, and (4) the low-rank approach using the presented recursive direct solver with the approximation in equation 48. Approaches 1 and 2 are in the class of optimization methods most commonly applied for inversion of marine CSEM data today. For all the tests, we use ℓ_2-norm model smoothness regularization. In Appendix A, we include inversion results from the synthetic survey in Figure 2.

The survey data considered here were acquired by EMGS ASA in 2014. Due to show right restrictions, we cannot describe the precise location of the survey or the details of the survey layout. We consider a subset of the source towlines and receiver deployments from the survey, and the spatial extent of the area covered by receivers and towlines is shown in Figure 10. The water depth in this area is approximately 2.5 km, and multiple salt bodies are located at the boundaries of the area covered by receivers. For the examples in this section, we included data from 256 receivers recording E_x and E_y from sources along 11 towlines. The receiver and towline spacing was 2 km. Note that there are receivers without a source towline crossing over their position in this data selection, i.e.,
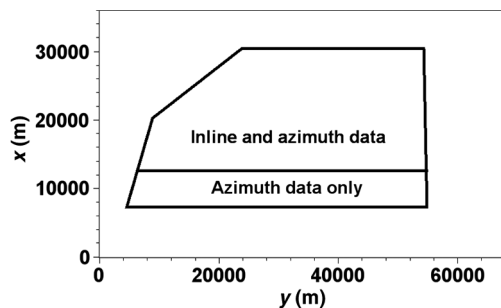
receivers recording only azimuth data. There are 1802 source positions in the data subset considered here. The survey frequencies were 0.125, 0.25, 0.5, 1.0, and 1.875 Hz. In total, the number of data samples is approximately 0.55×10^6 in these examples.

All of the inversions used a half-space initial guess resistivity model, with the bathymetry and water conductivity determined from survey data. We note that typically some structural information, such as salt outlines from seismic, is used to obtain the best inversion results. However, in this case, we instead use a simple initial model to highlight the performance of different inversion schemes. We consider anisotropic inversion with transverse isotropy defined by a vertical axis, i.e., a vertical transverse isotropy (VTI) model, in which we invert for horizontal and vertical resistivity. For some of the inversion tests, the convergence became slow after the misfit reached a root-mean-square (rms) value of 3.9, and we therefore use this value to have a fair comparison between methods. We define the rms value as $\sqrt{\varepsilon/N}$, where ε is defined in equation 2 and N is the number of observations. Although results with misfit this large are not suited for geologic interpretation, they allow us to effectively compare the performance and numerical cost of the different inversion schemes that we consider in this paper.

The model parameter compression scheme used for the inversion schemes 2 and 3 defined previously is based on creating a coarser discretization of the model at depth and away from the area covered by the receivers. The coarsening is determined by the expected resolution of the data. This scheme was used to obtain a reduction of the number of parameters on the order of a factor of 200 compared with a homogeneous discretization of the entire domain, and makes it feasible to construct the Gauss-Newton Hessian matrix for such a large model.

The results from the L-BFGS-B optimizer 1 are shown in Figure 11. To achieve this result we carried out 60 iterations of the inversion, where each iteration required 1004 forward solutions

Figure 10. Subset of the survey data considered in the inversion examples. Receivers were positioned in a 2×2 km grid inside the area shown by the polygons. There are nine towlines with direction west–east, and two towlines with direction north–south. The average lateral separation between towlines is 1.5 km, and the average maximum source-receiver offset is 12 km. The source towing was focused on the top region, where receivers recorded inline and azimuth data. The receivers in the lower region only recorded azimuth data.
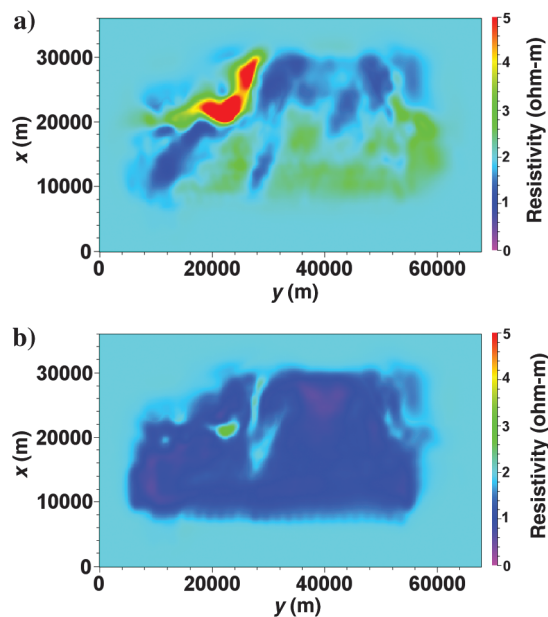

Figure 11. Survey scheme 1 inversion result: L-BFGS-B optimizer result; depth slices at 4200 m for (a) the vertical resistivity model and (b) the horizontal resistivity model; the total number of free parameters was 10^7.

(60,240 solutions in total). The memory usage to compute the update was approximately 10 GB for 10^7 free parameters.

In Figure 12, we show results from the Gauss-Newton scheme 2, where we used the model parameter compression and used a CG method to solve equation 4. This result was obtained after 13 iter-
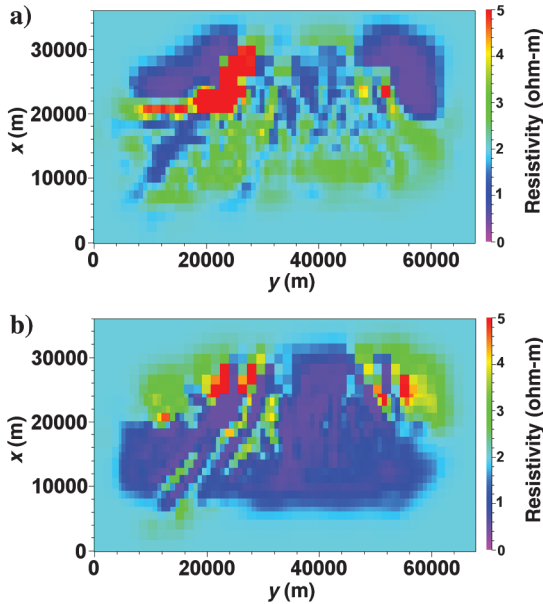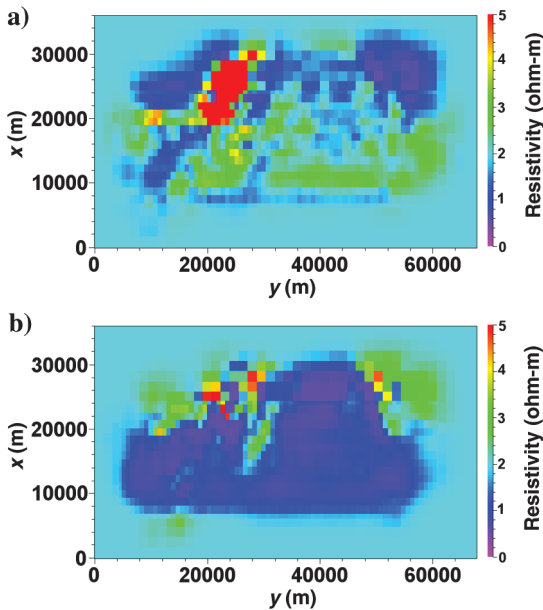

Figure 12. Survey scheme 2 inversion result: 3D Gauss-Newton optimizer result ($\mathbf{H} = \mathbf{H}_D + \mathbf{H}_R$) using CG solver and model parameter compression; depth slices at 4200 m for (a) the vertical resistivity model and (b) the horizontal resistivity model; the total number of free parameters was 5×10^4.

Figure 13. Survey scheme 3 inversion result: low-rank optimizer ($\tilde{\mathbf{H}} = \tilde{\mathbf{H}}_D + \mathbf{H}_R$) using CG solver and model parameter compression; depth slices at 4200 m for (a) the vertical resistivity model and (b) the horizontal resistivity model; the total number of free parameters was 5×10^4 and the number of source groups was $N_s = 1$.

ations, using 2806 forward solutions per iteration (36,478 simulations in total) and a memory usage of 12 GB for 5×10^4 parameters.

Figure 13 shows results obtained using the low-rank approximation 3 to the data part of the Hessian matrix in equation 15. For this case, we used the same model compression as in case 2. The results were obtained after 15 iterations, and the number of forward solutions per iteration was 1506 (22,590 forward solutions in total) using $N_s = 1$ source groups. The maximum memory usage was similar to case 2.

In Figure 14, we compare the eigenvalue distribution of the Hessian matrices for the first iteration in cases 2 and 3, with the eigenvalue distribution of their data Hessian matrices (without
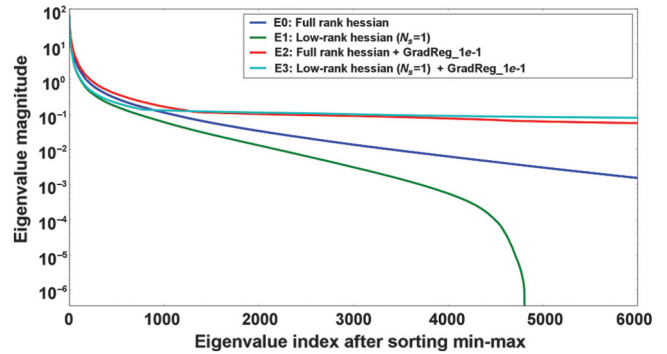

Figure 14. Eigenvalue decomposition for the first-iteration Hessian matrices (with and without regularization) of the inversions 2 and 3. E0: data Hessian matrix $\mathbf{H}_D$ of inversion 2. E1: data Hessian matrix $\tilde{\mathbf{H}}_D$ of inversion 3. E2: Hessian matrix $\mathbf{H} = \mathbf{H}_D + \mathbf{H}_{grad}$, with $W_{grad}(\sigma) = 1e^{-1}$, of inversion 2. E3: Hessian matrix $\hat{\mathbf{H}} = \tilde{\mathbf{H}}_D + \mathbf{H}_{grad}$, with $W_{grad}(\sigma) = 1e^{-1}$, of inversion 3.
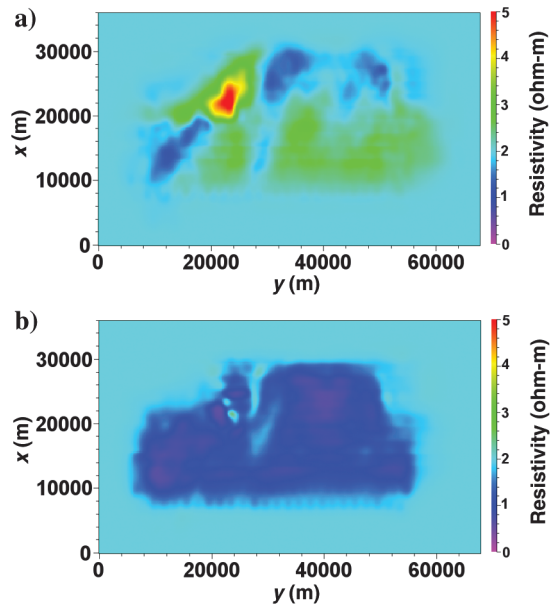

Figure 15. Survey scheme 4 inversion result: low-rank optimizer using the recursive solver with the diagonal of the regularization Hessian matrix ($\hat{\mathbf{H}} = \tilde{\mathbf{H}}_D + \tilde{\mathbf{H}}_R$); depth slices at 4200 m for (a) the vertical resistivity model and (b) the horizontal resistivity model; the total number of free parameters was 10^7 and the number of source groups was $N_s = 1$.

regularization). We see that when we incorporate the regularization, the eigenvalue distributions of both cases (compare E2 and E3) are similar. This is in agreement with the discussion of the results as shown in Figure 5.

Finally, in Figure 15 we show inversion results for case 4, where we use the low-rank approximation to the data part of the Hessian ($N_s = 1$), as well as the matrix-free recursive solver summarized in equation 47. For this case, we executed 21 iterations of the inversion, and the number of forward solutions per iteration was 1506 (31,584 simulations in total). We used the same number of free
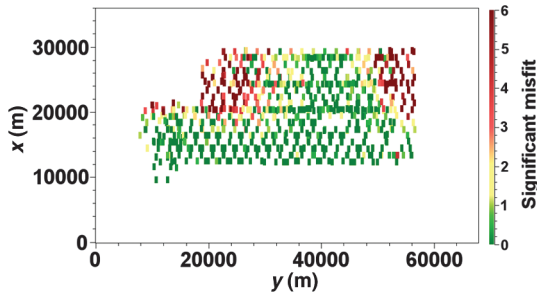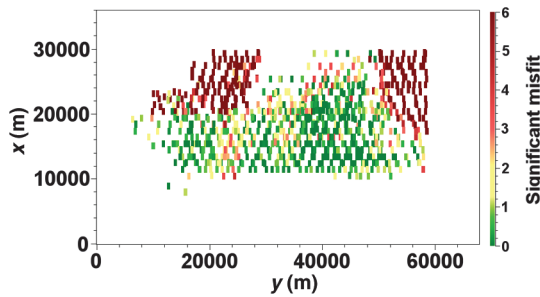

Figure 16. Significant misfit (equation 50) for survey scheme (4) inversion results: low-rank optimizer using the recursive solver with the diagonal of the regularization Hessian matrix ($\hat{\mathbf{H}} = \tilde{\mathbf{H}}_D + \tilde{\mathbf{H}}_R$). Significant misfit for all the data at offset 4 km and frequency 0.5 Hz.

Figure 17. Significant misfit (equation 50) for survey scheme (4) inversion results: low-rank optimizer using the recursive solver with the diagonal of the regularization Hessian matrix ($\hat{\mathbf{H}} = \tilde{\mathbf{H}}_D + \tilde{\mathbf{H}}_R$). Significant misfit for all the data at offset 8 km and frequency 0.5 Hz.

parameters, 10^7, as in case 1 with no compression of model parameters. The memory usage to construct the update was 180 GB. As explained above, for this particular case, we neglected the off-diagonal parts of the regularization Hessian matrix. This approximation may have impaired the inversion convergence (see the effects of this approximation in the synthetic results shown in Figure 17).

The large-scale structures in the recovered resistivity models are similar for the four inversion approaches. There is a difference in scale between cases 1 and 4 and 2 and 3 due to the model parameter compression scheme. The similarities appear to be stronger in the vertical than in the horizontal resistivity models. The resistive regions in the top left and top right corners are due to the presence of salt. The structures seen in the center of the survey area correspond to structures identified in seismic data. The magnitude of the resistivity contrast varies in the four cases, but note that the misfit is approximately the same for these results.

Let us now analyze the data fit. Our measure for the misfit is computed according to the following expression:

$$\chi = \frac{|F^{\text{obs}} - F^{\text{synth}}|}{\delta F^{\text{obs}}}. \tag{50}$$

This quantity, which we call a "significant misfit," gives the discrepancy between the observed and synthetic data in units of the estimated observed data measurement uncertainty, denoted by δF^{obs}. Figures 16 and 17 show the common midpoint sorted data misfit χ for offsets 4 and 8 km, respectively, for the case of the low-rank approximation with the recursive solver 4. The misfit distribution is similar for the other inversion cases considered in this paper. The largest variations of the residuals between different approaches are found in the northwest and northeast corners of the survey area, where salt bodies are present. The data coverage for these structures is poor due to a lack of receivers over these regions. Only the extension of source towlines out of the receiver grid gives data that sample this part of the model. The resulting lack of information makes it difficult for the inversion to fit these data.

DISCUSSION

In this section, we compare the computational complexity of the low-rank approach introduced in this paper to that of the quasi-Newton and Gauss-Newton approaches.

The L-BFGS-B optimizer is of quasi-Newton type and requires us to compute the gradient **g** for each update. The required number

---

Table 1. Computational complexity for the different schemes: (1) L-BFGS-B, (2) 3D Gauss-Newton using CG (3DGN + CG), (3) low-rank approach using CG (LR + CG), and (4) low-rank approach using the proposed recursive direct solver but just using the diagonal of the regularization matrix (LR + DS diag). The value k is the condition number of the Hessian matrix that in general depends on the size of the matrix.

	(1) L-BFGS-B $N = 10^7$	(2) 3DGN + CG $N_s = 1, N = 5 \times 10^4$	(3) LR + CG $N_s = 1, N = 5 \times 10^4$	(4) LR + DS diag $N_s = 1, N = 10^7$
Number of simulations per iteration	$\mathcal{O}(N_{\text{rx}})$	$\mathcal{O}(N_{\text{tx}})$	$\mathcal{O}(N_{\text{rx}}N_s)$	$\mathcal{O}(N_{\text{rx}}N_s)$
Jacobian columns (M or $\tilde{M}$)	Five updates	$\mathcal{O}(N_{\text{tx}})$	$\mathcal{O}(N_s)$	$\mathcal{O}(N_s)$
Solver time complexity	$\mathcal{O}(N)$	$\mathcal{O}(N^2 k^{1/2})$	$\mathcal{O}(N^2 k^{1/2})$	$\mathcal{O}(N\tilde{M}^2)$
Solver memory complexity	$\mathcal{O}(N)$	$\mathcal{O}(N^2)$	$\mathcal{O}(N^2)$	$\mathcal{O}(N\tilde{M})$

of forward solutions scales by the number of receivers and not the number of sources. The computation of the update has a relatively small numerical cost. The L-BFGS-B builds up an approximation to the Hessian matrix using past gradients and updates. The Gauss-Newton scheme has a larger computational cost because it involves an explicit computation of a second derivative from the Jacobian.

In Table 1, we show a comparison between the computational cost of the different schemes in which we showed inversion results above.

Table 1 shows that the number of simulations for the low-rank approach is closer to that of the quasi-Newton scheme when the number of source groups is small (the presented results use $N_s = 1$). The computational load from forward solutions following from Gauss-Newton is very large.

A further benefit of the low-rank approach is that the data samples (Jacobian columns) required to construct the Hessian scales with the number of source groups N_s rather than the number of sources N_{tx}, such as the Gauss-Newton scheme. For the L-BFGS-B optimizer, it is only necessary to store the last few model updates and gradients for a few iterations (five in our case) to estimate the Hessian matrix.

In the previous section, we showed how a recursive direct solver can avoid the construction of the data part of the Hessian matrix. The use of the low-rank approach in conjunction with this solver results in a very small memory complexity. Thus, we can solve large systems when the memory resources are limited (e.g., requires GB instead of TB when using 10^6 unknowns). The time complexity may however be similar for cases 3 and 4 described above.

In Table 2, we summarize the computational complexity of each of the four tests.

The 3D Gauss-Newton scheme 2 and low-rank schemes 3 and 4 require fewer iterations than the quasi-Newton case 1 to achieve the same misfit. This is an indication that the low-rank schemes give better updates than the quasi-Newton case.

In the case of the low-rank inversion (schemes 3 and 4), the total number of simulations demonstrates that fewer forward solutions are required than in case of the 3D Gauss-Newton scheme 2 to reach the same misfit. We expect that this difference can be larger for surveys with a larger number of source positions (in this case, $N_{tx}/N_{rx} = 4.6$). The difference between the number of simulations for the low-rank cases 3 and 4 is probably due to the approximated regularization Hessian matrix, which leads to the requirement for more iterations to reach the same misfit.

The low-rank scheme results in a significant compression of the Jacobian matrix used to form the Hessian. Because computational complexity is proportional to the number of Jacobian columns, this is an important factor to take into consideration. In implementations in which a large Jacobian matrix is stored, the memory complexity can be a challenge, too. The low-rank scheme can reduce this complexity significantly as we showed in Tables 1 and 2, and in our example, leads to reduction by a factor $M/\tilde{M} = 550,000/2510 \asymp 200$.

CONCLUSIONS

In this paper, we have introduced a low-rank approximation to the 3D CSEM Gauss-Newton data Hessian matrix, and a matrix-free recursive direct solver. When the approximation is used together with the direct solver, inversion results similar to standard 3D Gauss-Newton optimization are obtained, but at a cost similar to that of quasi-Newton methods. This is achieved without modifying the cost function or the gradient, preserving the main structure of the Gauss-Newton scheme. Moreover, the direct solver allowed us to use a much finer discretization of the model with an associated larger number of inversion parameters.

The low-rank data Hessian matrix approach can reduce the number of simulations per iteration in the order of the ratio of the number of sources to the number of receivers in a survey. Moreover, the matrix-free recursive direct solver reduces the memory complexity when using large number of inversion parameters.

In our results, the low-rank approach reaches the same residual misfit as full rank 3D Gauss-Newton when starting from a half-space, but with a lower total number of simulations. This is due to the fact that the low-rank scheme needs approximately the same number of iterations as the full rank 3D Gauss-Newton, but a number of simulations per iteration closer to a quasi-Newton scheme. The recursive direct solver allows using a larger number of inversion parameters than with the 3D Gauss-Newton scheme due to the smaller use of memory.

The recursive direct solver benefits when using a narrow-banded regularization Hessian matrix. For this case, it is feasible to factorize the regularization Hessian matrix, and reuse it in every step of the direct solver. This avoids the approximation based on neglecting the off-diagonal parts of the regularization Hessian matrix, which may impair the inversion.

Table 2. Computational cost to obtain inversion results: (1) L-BFGS-B; (2) 3D Gauss-Newton using CG (3DGN + CG), (3) low-rank approach using CG (LR + CG), and (4) low-rank approach using the proposed recursive direct solver but just using the diagonal of the regularization matrix (LR + DS diag). Note: The numbers in square brackets for schemes (2) and (3) show the cost in the case in which the model parameter compression scheme is not applied, i.e., in the case of using 10^7 free parameters.

	(1) L-BFGS-B	(2) 3DGN + CG	(3) LR + CG	(4) LR + DS diag
Number of free parameters (N)	10^7	5×10^4	5×10^4	10^7
Number of simulations per iteration	1004	2806	1506	1506
Jacobian columns (M or $\tilde{M}$)	Five updates	550000	2510	2510
Solver time complexity (FLOPs)	10^9	2.5×10^{12} [10^{17}]	2.5×10^{12} [10^{17}]	6.3×10^{12}
Solver memory complexity (GB)	0.149	12 [93,132]	12 [93,132]	180
Total number of iterations	60	13	15	21
Total number of simulations	60,240	36,478	22,590	31,584

APPENDIX A

SYNTHETIC SURVEY RESULTS

In this appendix, we will present inversion results for the synthetic survey shown in Figure 2. This survey example was used to illustrate certain theoretical aspects of the low-rank approximation in section "Theory." The numerical complexity involved in the inversion of this synthetic data set is not very large, but the results are useful to validate the different schemes considered.
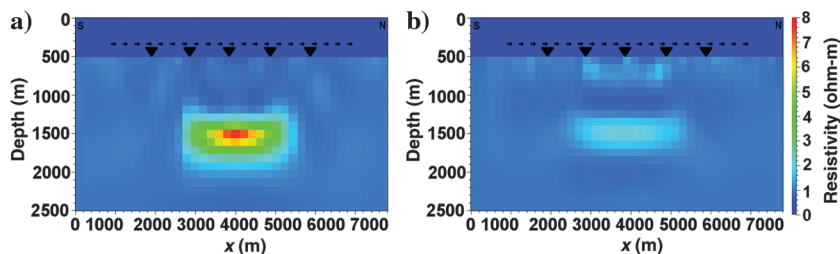

Figure A-1. Synthetic-survey scheme 1_s inversion result: L-BFGS-B optimizer result; vertical slice at $y = 2900$ m for (a) the vertical resistivity model and (b) the horizontal resistivity model; iteration 74 with misfit $= 0.382$ rms.
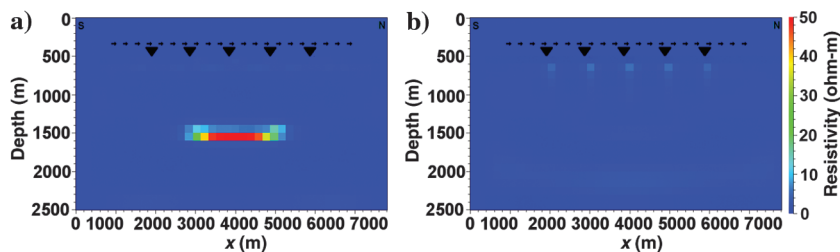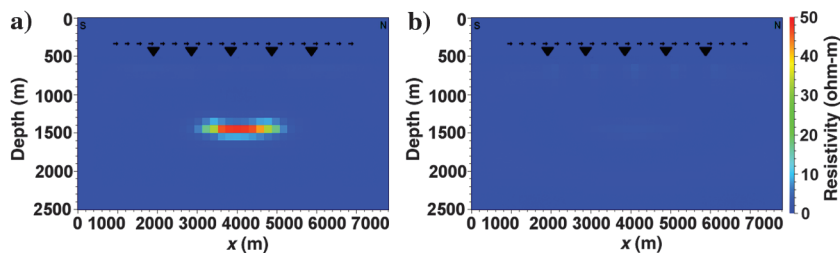

Figure A-2. Synthetic-survey scheme 2_s inversion result: 3D Gauss-Newton optimizer result ($\mathbf{H} = \mathbf{H}_D + \mathbf{H}_R$) using CG solver; vertical slice at $y = 2900$ m for (a) the vertical resistivity model and (b) the horizontal resistivity model; iteration 53 with misfit $= 0.125$ rms.

Figure A-3. Synthetic-survey scheme 3_s inversion result: low-rank optimizer ($\tilde{\mathbf{H}} = \tilde{\mathbf{H}}_D + \mathbf{H}_R$); vertical slice at $y = 2900$ m for (a) the vertical resistivity model and (b) the horizontal resistivity model; iteration 76 with misfit $= 0.137$ rms.

We consider four different inversion methodologies: (1_s) quasi-Newton L-BFGS-B (Zhu et al., 1997), (2_s) 3D Gauss-Newton with a CG solver for equation 4, (3_s) the proposed low-rank approach with a CG solver for equation 15, and (4_s) the low-rank approach using the presented recursive direct solver with the approximation in equation 48. Due to the small size of the synthetic survey considered, and in contrast to the field data example studied in section "Inversion results," we do not need to use a model parameter compression. The inversion parametrization is a regular grid with cell size $200 \times 200 \times 100$ m. We consider a VTI model, resulting in 5×10^4 free parameters. The initial guess model was a half-space for all the inversions considered in this appendix.

This initial model has a background resistivity of 1 ohm-m, and a water resistivity of 0.25 ohm-m. We terminated the iterations when the convergence rate became too slow (misfit change in new iteration smaller than 1%).

In Figure A-1, we show the inversion results for the scheme 1_s. The inversion took 74 iterations, where each iteration required 50 forward solutions (3700 solutions in total). The memory usage to compute the update was approximately 10 MB. Though the transverse resistivity of the target is of the same order as in the true model, the inversion does not recover the correct target shape. We expect that the geometry of the reconstruction could be improved by tuning the regularization settings to favor a model with resistivity contrasts rather than smooth variations. Note that the target is reconstructed in the vertical resistivity component because the vertical resistivity governs the target response, and we did not apply any anisotropy regularization. This is common to all the results in this appendix.

The inversion result for the scheme 2_s is shown in Figure A-2. This result was obtained after 53 iterations, using 645 forward solutions per iteration (34,185 simulations in total) and a memory usage of 12 GB for the calculation of the update. The target is recovered in the vertical model with the approximate correct shape and resistivity. In the horizontal model, we can observe weak artifacts at positions correlating to the receiver positions.

Figure A-3 shows the inversion results for the scheme 3_s. This result was achieved after 53 iterations, with 150 forward solutions per iteration (7950 simulations in total), and a memory usage of 12 GB for the calculation of the update. We observe that these results are similar to those achieved with the scheme 2_s. However, the target is situated 100 m above its position in the true model. Considering the low frequencies involved, we believe that this difference (on the scale of the discretization) is below the resolution of the data.

For the synthetic survey that we considered in this appendix, the reduced memory usage from the recursive direct solver is not relevant because the low number of model parameters makes the CG solver feasible even without model parameter
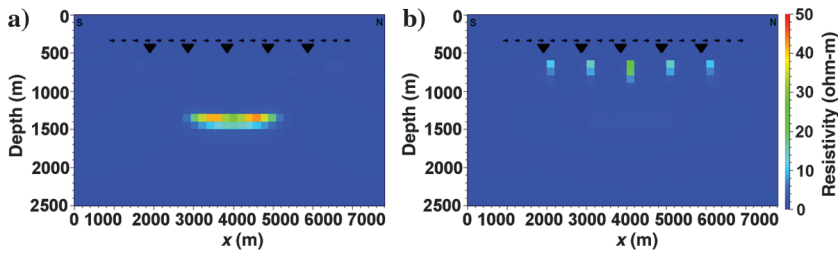
Figure A-4. Synthetic-survey 4_s inversion result: low-rank optimizer using the recursive solver with the diagonal of the regularization Hessian matrix ($\tilde{\mathbf{H}} = \tilde{\mathbf{H}}_D + \tilde{\mathbf{H}}_R$); vertical slice at $y = 2900$ m for (a) the vertical resistivity model and (b) the horizontal resistivity model; iteration 119 with misfit = 0.139 rms.

compression. However, to validate the different approaches, we show in Figure A-4, the inversion results for scheme 4_s. This result was achieved after 119 iterations, using 150 forward solutions per iteration (17,850 simulations in total). The memory usage to compute the update was approximately 20 MB. In these inversion results, we observe detrimental effects following from the approximation of the regularization Hessian matrix as a diagonal matrix. The target is not as focused as in the result from the previous schemes, and in the horizontal model the artifacts are stronger. We expect that these aspects would improve if a better approximation to the inverse of the regularization Hessian contribution was used when constructing the update in the data Hessian matrix-free solver.

REFERENCES

Abubakar, A., T. M. Habashy, V. Druskin, D. Alumbaugh, A. Zerelli, and L. Knizhnerman, 2006, Two-and-half-dimensional forward and inverse modeling for marine CSEM: 76th Annual International Meeting, SEG, Expanded Abstracts, 750–754.

Abubakar, A., T. M. Habashy, M. Li, and J. Liu, 2009, Inversion algorithms for large-scale geophysical electromagnetic measurements: Inverse Problems, 25, 123012, doi: 10.1088/0266-5611/25/12/123012.

Bansal, R., J. Krebs, P. Routh, S. Lee, J. Anderson, A. Baumstein, A. Mullur, S. Lazaratos, I. Chikichev, and D. McAdow, 2013, Simultaneous-source full-wavefield inversion: The Leading Edge, 32, 1100–1108, doi: 10.1190/tle32091100.1.

Ben-Hadj-Ali, H., S. Operto, and J. Virieux, 2011, An efficient frequency-domain full waveform inversion method using simultaneous encoded sources: Geophysics, 76, no. 4, R109–R124, doi: 10.1190/1.3581357.

Boonyasiriwat, C., and G. T. Schuster, 2010, 3D multisource full-waveform inversion using dynamic random phase encoding: 80th Annual International Meeting, SEG, Expanded Abstracts, 1044–1049.

de Groot-Hedlin, C., and S. Constable, 1990, Occam's inversion to generate smooth, two-dimensional models from magnetotelluric data: Geophysics, 55, 1613–1624, doi: 10.1190/1.1442813.

Grayver, A. V., R. Streich, and O. Ritter, 2013, Three-dimensional parallel distributed inversion of CSEM data using a direct forward solver: Geophysical Journal International, 193, 1432–1446, doi: 10.1093/gji/ggt055.

Hesthammer, J., A. Stefatos, M. Boulaenko, S. Fanavoll, and J. Danielsen, 2010, CSEM performance in light of well results: The Leading Edge, 29, 34–41, doi: 10.1190/1.3284051.

Hoversten, G., D. Myer, K. Key, O. Hermann, R. Hobbet, and D. Alumbaugh, 2013, CSEM and MMT base basalt imaging: 75th Annual International Conference and Exhibition, EAGE incorporating SPE EUROPEC 2013.

Jing, X., C. J. Finn, T. A. Dickens, and D. E. Willen, 2000, Encoding multiple shot gathers in prestack migration: 70th Annual International Meeting, SEG, Expanded Abstracts, 786–789.

Krebs, J. R., J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D. Lacasse, 2009, Fast full-wavefield seismic inversion using encoded sources: Geophysics, 74, no. 6, WCC177–WCC188, doi: 10.1190/1.3230502.

Li, M., A. Abubakar, J. Liu, G. Pan, and T. M. Habashy, 2011, A compressed implicit Jacobian scheme for 3D electromagnetic data inversion: Geophysics, 76, no. 3, F173–F183, doi: 10.1190/1.3569482.

Li, Y., and K. Key, 2007, 2D marine controlled-source electromagnetic modeling: Part 1 — An adaptive finite-element algorithm: Geophysics, 72, no. 2, WA51–WA62, doi: 10.1190/1.2432262.

Lin, Y., M. Li, A. Abubakar, and T. M. Habashy, 2013, A wavelet-based model compression method for three-dimensional electromagnetic data inversion: 83rd Annual International Meeting, SEG, Expanded Abstracts, 707–712, doi: 10.1190/segam2013-0395.1.

Loke, M., and T. Dahlin, 2002, A comparison of the Gauss-Newton and quasi-Newton methods in resistivity imaging inversion: Journal of Applied Geophysics, 49, 149–162, doi: 10.1016/S0926-9851(01)00106-9.

Mackie, R., M. Watts, and W. Rodi, 2007, Joint 3D inversion of marine CSEM and MT data: 77th Annual International Meeting, SEG, Expanded Abstracts, 574–578.

Mittet, R., H. Maulana, K. Brauti, and T. A. Wicklund, 2007, CMP inversion of marine CSEM data: Presented at the Electromagnetic, Gravity, and Magnetic (EGM) 2007 International Workshop, http://www.earthdoc.org/publication/publicationdetails/?publication=41234.

Morten, J. P., F. Roth, D. Timko, C. Pacurar, A. K. Nguyen, and P. A. Olsen, 2011, 3D reservoir characterization of a North Sea oil field using quantitative seismic and CSEM interpretation: 81st Annual International Meeting, SEG, Expanded Abstracts, 1903–1907, doi: 10.3997/2214-4609.20130139.

Morten, J. P., C. Twarz, V. Ricoy-Paramo, and S. Sun, 2013, Improved resolution salt imaging from 3D CSEM anisotropic inversion: 75th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: 10.3997/2214-4609.20130139.

Morton, S. A., and C. C. Ober, 1998, Faster shot-record depth migrations using phase encoding: 68th Annual International Meeting, SEG, Expanded Abstracts, 1131–1134.

Nocedal, J., and S. J. Wright, 2006, Numerical optimization, 2nd ed.: Springer.

Pratt, R. G., C. Shin, and G. J. Hicks, 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: Geophysical Journal International, 133, 341–362, doi: 10.1046/j.1365-246X.1998.00498.x.

Press, W. H., S. A. Teukolosky, W. T. Vetterling, and B. P. Flannery, 2002, Numerical recipes in C, 2nd ed.: Cambridge University Press.

Romero, L. A., D. C. Ghiglia, C. C. Ober, and S. A. Morton, 2000, Phase encoding of shot records in prestack migration: Geophysics, 65, 426–436, doi: 10.1190/1.1444737.

Schiemenz, A., and H. Igel, 2013, Accelerated 3-D full-waveform inversion using simultaneously encoded sources in the time domain: Application to Valhall ocean-bottom cable data: Geophysical Journal International, 195, 1970–1988, doi: 10.1093/gji/ggt362.

Støren, T., J. Zach, and F. Maaø, 2008, Gradient calculations for 3D inversion of CSEM data using a fast finite-difference time-domain modeling code: 70th Annual International Conference and Exhibition, EAGE, Extended Abstracts, P194.

Zhdanov, M. S., 2009, Geophysical electromagnetic theory and methods: Elsevier Science.

Zhdanov, M. S., A. Gribenko, and M. Cuma, 2007, Regularized focusing inversion of marine CSEM data using minimum vertical-support stabilizer: 77th Annual International Meeting, SEG, Expanded Abstracts, 579–583.

Zhu, C., H. Byrd, and J. Nocedal, 1997, L-BFGS-B: Algorithm 778: L-BFGS-B, FORTRAN routines for large scale bound constrained optimization: ACM Transactions on Mathematical Software, 23, 550–560, doi: 10.1145/279232.279236.