Manuel Amaya

High-order optimization methods for large-scale 3D CSEM data inversion

Manuel Amaya

Doctoral Thesis

NTNU
Norwegian University of
Science and Technology
Faculty of Information Technology,
Mathematics and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

NTNU

NTNU
Norwegian University of
Science and Technology

Manuel Amaya

High-order optimization methods for large-scale 3D CSEM data inversion

Thesis for the degree of Philosophiae Doctor

Trondheim, October 2015

Norwegian University of Science and Technology
Faculty of Information Technology,
Mathematics and Electrical Engineering
Department of Mathematical Sciences

◙ NTNU
Norwegian University of
Science and Technology

NTNU

Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Information Technology,
Mathematics and Electrical Engineering
Department of Mathematical Sciences

To my family.

This page is intentionally left blank.

Abstract

Marine controlled-source electromagnetic (CSEM) method is a non-invasive off-shore technique used, in association with magnetotelluric and seismic data, for the study of the oceanographic lithosphere and hydrocarbon reservoir exploration. CSEM data are often used in optimization processes that produce an electrical resistivity imaging of the subsurface. CSEM research shows interest for developing high-order optimization methods, able to achieve faster convergences without investing too much manual effort building initial inversion models. As a result, 3D CSEM industry has started a transition from quasi-Newton to Gauss-Newton methods.

The large numerical complexity is a limiting factor when applying the Gauss-Newton method for the 3D inversion of CSEM data. These problems can involve $\mathcal{O}(10^6)$ inversion parameters and $\mathcal{O}(10^5)$ forward simulations, resulting in a Jacobian matrix of $\mathcal{O}(100 \text{ TB})$ and a Gauss-Newton Hessian matrix of $\mathcal{O}(1 \text{ TB})$. There are some papers that propose methods to reduce the memory complexity and others that present schemes to reduce the time complexity. However there is not a proposal to significantly reduce the total numerical complexity of the 3D Gauss-Newton optimization method without affecting the parameterization of the problem.

The first main contribution of this thesis is a method for obtaining a low-rank approximation of the Gauss-Newton Hessian matrix that dramatically reduces the numerical complexity of the 3D CSEM Gauss-Newton optimization without altering the parameterization of the resistivity models. For large-scale surveys, it can reduce the number of forward simulations between 10-100 times, and it also reduces the memory complexity, from $\mathcal{O}(\text{TB})$ to $\mathcal{O}(\text{GB})$. It is based on simulating groups of distant phase-encoded sources, instead of single-source simulations. The resultant small number of simulations motivated the development of a matrix free recursive direct solver to obtain the model updates at each iteration with a reduced memory usage. A study of the associated cross-talk noise and inversion results validates this proposal.

The second main contribution of this thesis is the introduction, apparently for the first time in 3D CSEM, of the Newton and the Halley class methods. This opens the state-of-the-art frontiers to higher-order methods where the computation of a Green function per model parameter is required. Initially, the numerical complexity of these methods makes their use unapproachable. In this research it is concluded that it is possible to apply these methods with the same memory complexity as in a Gauss-Newton method, and with a contained time complexity. It is proposed the use of a finite-difference frequency-domain direct solver for on-the-fly computations of the Green functions, a reduced memory construction of the systems matrices and the modification of a trust-region solver to handle the indefiniteness of these matrices. Synthetic 3D CSEM survey inversion results demonstrate the feasibility of this method.

This page is intentionally left blank.

Preface and publications

This thesis is submitted in partial fulfilment of the requirements for the degree of *philosophiae doctor* (PhD) at the Norwegian University of Science and Tecnology (NTNU). Starting the 11th of October of 2012, this research was funded for three years by the Research Council of Norway (PETROMAKS project 217223). While keeping the contact with the department, the research was mainly carried out at EMGS ASA company headquarters (Trondheim, Norway). During October 2014 I was visiting the Institute of Geophysics and Planetary Physics (IGPP) of SCRIPPS (UC San Diego, US).

Professor Brynjulf Owren has been my supervisor for this thesis. He is professor in the Department of Mathematical Sciences at the Faculty of Information Technology, Mathematics and Electrical Engineering (IME) at the Norwegian University of Science and Technology (NTNU) in Trondheim, Norway.

PhD. Jan Petter Morten has been my co-supervisor. He works as Principal Scientist at the company EMGS ASA in Trondheim (Norway).

Electromagnetic Geoservices (EMGS) ASA is an oil-industry services company. In 2002 EMGS was the first company to successfully apply controlled-source electromagnetic (CSEM) technology to hydrocarbons exploration. This company provides services for the acquisition, processing, inversion, integration and interpretation of CSEM data.

During this research, were written two extended abstracts (chapter 2 and chapter 3) that are already presented and published at relevant international conferences, and two journal papers (chapter 4 and chapter 5) submitted for publication. I am the main author of all of them. The layout, bibliography and typography of these papers have been unified to be included in this thesis. The developed software solutions and tools are based on EMGS proprietary code. I have participated in the development of the Gauss-Newton inversion code, and I am the sole author of the implementation of the different algorithms/methods presented in the four papers. EMGS ASA collected the CSEM field data that I have used for the inversions which results are shown in chapter 4. I have obtained all the inversion results shown in the papers, with the exception of the CSEM field data inversion result based on the quasi-Newton method L-BFGS-B in chapter 4. This inversion was carried out by EMGS employees and it is used for comparing the rest of the methods presented in the paper.

Manuel Amaya

This page is intentionally left blank.

Acknowledgements

This page is intentionally left blank.

Table of Contents

Chapter 1

Introduction

The aim of this thesis is to develop efficient high-order optimization methods for the large-scale inversion of three-dimensional (3D) controlled-source electromagnetic (CSEM) data. The first part of this chapter is an introduction to the main concepts, motivations and challenges that this PhD research involves. This part is divided into six subsections: The first three, intended for a general audience, introduce common concepts of geophysical inversion data, explain the 3D CSEM survey technique, and summarize the main challenges that a 3D CSEM optimization problem involves; the following two subsections are focused on the 3D CSEM Gauss-Newton optimization, and the main papers that present different approaches to reduce its computational cost; in the last subsection it is discussed the motivation for developing higher order optimization methods than Gauss-Newton for the inversion of 3D CSEM data. The last two parts of this chapter contain an explanation of the scope of this thesis, and its outline.

1.1 Background and motivation

1.1.1 Geophysical data inversion

An important part of geophysics is to understand the subsurface of the earth by making inferences based on different types of measurements (observed data). These measurements can be based on direct digging (or drilling), or can be based on non-destructive methods. The data from these last methods depend on the interaction of a physical field (e.g. gravity, magnetic field, seismic wave-field or electromagnetic field), generated by a natural or an artificial source, with the physical properties of the subsurface where it propagates through. In some cases, a simple representation of observed data is sufficient, but when more detailed information is needed, quantitative models of the earth need to be estimated. Calculating the (synthetic) data, given a source and a media, is called the forward problem. Obtaining the media parameters, given the

observed data and the source, is known as the inverse problem.

In general, the objective of a geophysical inverse problems is to find a parameter model of the earth that is consistent with the observed data. These inverse problems are formulated as a mathematical optimization problem, where the function to be optimized (objective function or misfit function) typically considers the differences between the observed data and the synthetic data calculated for the trial model. These problems are often ill-posed due to: the number of observed data is limited and usually contaminated by noise; the mathematical models are simplifications of the true geophysical phenomena; and the solution of these problems is solved with finite precision. According to the definition given by Jacques Hadamard, a well-posed problem has a solution, this solution is unique, and it changes continuously with the initial condition. Therefore, the objective function often incorporates an additional regularization term, based on a priori information of the geophysical phenomena in the subsurface, that reduces the ambiguity and increase the stability of the solution. The foundations of the regularization theory were developed by Andrei N. Tikhonov. An overview of the different optimization methods for geophysics data inversion can be found in Zhdanov (2002), Tarantola (2005) and Zhdanov (2009).

1.1.2 3D CSEM survey technique

Measurements of the sub-seafloor electrical resistivity have an important function in the study of the oceanographic lithosphere, the hydrocarbon (oil and gas) exploration and reservoir assessment. These data were traditionally collected using wire-line logging of boreholes, a technique successfully introduced in the oil and gas industry by the Schlumberger brothers (Johnson, 1962). This method involves the significant cost of drilling test wells into structures about which a priori information is not always available. The use of a non-invasive geophysical electromagnetic (EM) method represents a major economic advantage although it is not able to provide the same vertical resolution, due to the diffusive nature of EM energy propagation (Løseth et al., 2006).

A marine controlled-source electromagnetic (CSEM) survey is an offshore technique used to map the electrical resistivity of the sub-seafloor. Initially this technique was used for the study of the oceanographic lithosphere (Constable and Srnka, 2007). Eidesmo et al. (2002) introduced a technique for the application of CSEM to determine the presence of hydrocarbon reservoirs. The method is based on the resistivity contrast between the resistivity (tens of Ωm) of hydrocarbon-saturated layers, and the surrounding low resistivity (few Ωm) sedimentary layers saturated with aqueous saline fluids. Nowadays, CSEM data is used in conjunction with magnetotelluric (MT) data and seismic data (e.g. Hu et al. (2009) and Brown et al. (2012)), demonstrating the potential to significantly increase drilling success rate (e.g. of the study of this benefit in Hesthammer et al. (2010) and Fanavoll et al. (2010)). It has also shown to be effective in field appraisal (e.g. Morten et al. (2011) and Ziolkowski and Wright (2012)) and in structural imaging applications (e.g. Hoversten et al. (2013) and Morten et al. (2013)).

In a CSEM survey, a horizontal electric dipole (antenna) is towed close to the seabed, transmitting a high-power low-frequency signal that penetrates in the subsurface (figure 1.1). The EM field is attenuated when it propagates through conductive sedimentary layers, but the presence of resistive layers, like the ones produced by hydrocarbon deposits, results in a wave-guide EM propagation (an explanation of this effect can be found in Mittet and Morten (2013)). This effect makes the detection of hydrocarbon reservoir possible despite the inherent low resolution of CSEM data. A grid of receivers deployed on the seafloor measures the EM fields.
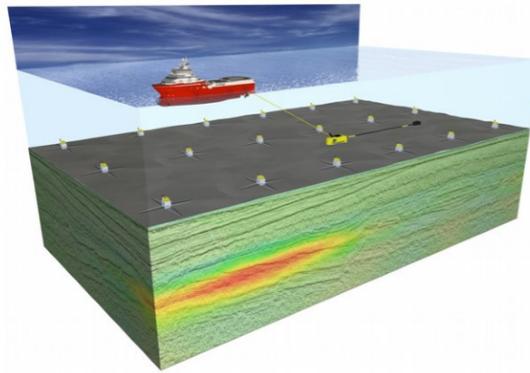
Figure 1.1: Representation of a CSEM survey. A vessel is towing an electric dipole (in yellow and black) close to the seabed while a grid of receivers (white boxes) are recording electromagnetic measurements. As a pseudo-example of CSEM data inversion results, on the subsurface section the electrical resistivity information (in light colours) is superimposed to the geological strata (seismic information in grey lines).

Currently, the most advanced techniques are the full-azimuth three-dimensional (3D) CSEM surveys (e.g. Fanavoll et al. (2014)). Using the collected data in an inversion process, it is possible to generate a 3D resistivity map of the subsurface, with the benefits over one-dimensional (1D) and two-dimensional (2D) techniques, of imaging the lateral resistivity variations and the resistors situated between towlines. In full-azimuth 3D CSEM surveys the receivers are taking EM measurements while the antenna is towed through the whole survey (azimuth data), and not only for those moments when the antenna is being towed over the receiver line (inline data) like in lower-dimensional methods. These are measurements of two orthogonal components of the electric and magnetic fields, for several frequencies (mostly between 0.1 Hz and 10 Hz). As an example, a realistic 3D CSEM survey can cover an area of 40 km $\times$ 40 km using hundreds of receivers. For this example the total amount of data measurements is about 10^7 samples (survey data or observed data), when considering frequency-domain data.

Due to the collaboration with EMGS ASA, this PhD research is focused on the inversion of 3D CSEM data related to hydrocarbon explorations. However, all the methods contained in this thesis can also be applied in the inversion of 3D CSEM

data for its applicability in the study of the oceanographic lithosphere. In some studies (e.g. Hölz et al. (2015) and Myer et al. (2006)) the use of 3D CSEM data inversion is essential for having a correct interpretation of the geology.

1.1.3 The optimization problem in 3D CSEM

3D CSEM data inversion is formulated as a local iterative optimization problem due to the numerical complexity associated when considering regular size surveys. By comparing full-wave forward modelling simulation results (synthetic data) with the observed data, an initial resistivity model is iteratively modified with inversion model updates until the differences between both types of data are sufficiently reduced (figure 1.2). In this local optimization scheme, the diagonal of the covariance matrix is commonly used to estimate the confidence in the inverted parameters. Currently, stochastic optimization of CSEM data, where a global minimum is recovered and an accurate uncertainty measurement is obtained, is only feasible in small 1D and 2D problems. In Trainor-Guitton and Hoversten (2010) some of the main practical challenges of stochastic methods in CSEM are shown.
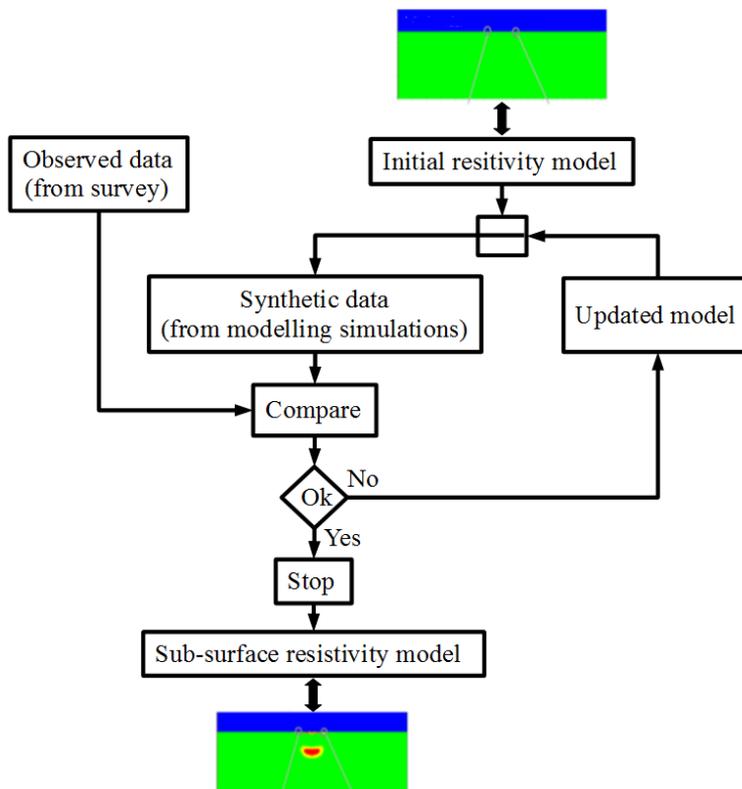
Figure 1.2: Local optimization scheme for CSEM data.

The first main reason of the numerical complexity associated with 3D CSEM data

inversion is the number of (forward) modelling simulations. For each iteration model in a local optimization method, and for each random sample model in a stochastic method, it is necessary to compute the synthetic data. In addition, at each iteration of a local optimization, it is necessary to compute the Green functions involved (an example in the subsection 1.1.4). In EM, a Green function is a unit dipole moment source solution of Maxwell's equations for a given conductivity model.

For performing EM modelling simulations there are examples of solution methods based on finite-difference frequency-domain (FDFD), finite-difference time-domain (FDTD), finite-element (FE), finite-volume (FV) and integral equation (IE). A review of these methods can be found in e.g. Avdeev (2005), Zhdanov (2009) and Börner (2010). Except for IE, the results for Maxwell equations are obtained by solving large sparse linear systems. In EM modelling, iterative solvers are typically used to solve the linear systems due to their moderate computational complexity, though their convergence behaviour and their iteration accuracy are deteriorated in the presence of large resistivity contrast and non-uniform simulation meshes. On the other hand, direct solvers are robust methods that factorize the system matrix and then efficiently obtain the right-hand side solutions (RHS) with high accuracy. Due to their large computational complexity and limited scalability, direct solvers have not been considered for 3D EM simulations. However, as is studied in Grayber and Streich (2012), the evolution of the direct solvers starts to be feasible for use in moderate size problems (smaller than typical industrial use).

The other main reason for the numerical complexity of the 3D CSEM data inversion is due to the large number of inversion parameters. As is explained in the following paragraphs, this number is determined by the model discretization, the model anisotropy considered and the grid decoupling technique applied between the modelling discretization and the inversion parameterization.

The model discretization depends on the dimensions of the survey area, the heterogeneity of the survey layout and the source frequencies. These frequencies are determined based on the maximum depth bellow the mud-line that need to be represented, the maximum EM wave penetration (the electromagnetic skin-depth for the lowest frequency propagating at the largest background resistivity), and the smallest length-scale that need to be describe considering the EM maximum sensitivity (Mittet and Morten, 2012). As an example, for an area of 40 km x 40 km, a maximum depth of 5 km bellow mud-line (considering highest background resistivity of 10 Ω m at 0.1 Hz), using regular box-basis functions of 200 m x 200 m x 100 m (for a pixel-based or cell-based model representation), the model would be discretized into $2 \cdot 10^6$ cells. In addition, for the modelling simulations it would be necessary to include the cells corresponding to the water layer and the air layer. These layers are not considered in the inversion due to their conductivity values are typically known (measured during the survey).

The anisotropy of the geological formations related to hydrocarbons is commonly linked to the layers of sedimentary materials. For example, the electrical conductivity in the parallel direction to a sedimentary layer can be different than in the vertical

direction, due to the influence of the gravity on the deposition. In EM, it is common to use a tensor of six independent elements (e.g. Weiss and Newman (2002)) to describe an anisotropic conductivity model $\sigma(\mathbf{r})$ at position $\mathbf{r}$:

$$\sigma(\mathbf{r}) = \begin{bmatrix} \sigma_{\mathtt{xx}}(\mathbf{r}) & \sigma_{\mathtt{xy}}(\mathbf{r}) & \sigma_{\mathtt{xz}}(\mathbf{r}) \\ \sigma_{\mathtt{yx}}(\mathbf{r}) & \sigma_{\mathtt{yy}}(\mathbf{r}) & \sigma_{\mathtt{yz}}(\mathbf{r}) \\ \sigma_{\mathtt{zx}}(\mathbf{r}) & \sigma_{\mathtt{zy}}(\mathbf{r}) & \sigma_{\mathtt{zz}}(\mathbf{r}) \end{bmatrix}. \tag{1.1}$$

As it is shown in Tompkins (2004), Tompkins (2005) and Ellis et al. (2011), it is important to consider the electrical anisotropy to avoid wrong interpretation of CSEM data. In many situations of CSEM hydrocarbon explorations, horizontally layered sediments that can exhibit vertical transversal isotropy (VTI) on a macroscopic scale appear. This allows us to use the simplification that only considers three elements in eq. 1.1, $\sigma_{xx}(\mathbf{r}) = \sigma_{yy}(\mathbf{r}) = \sigma_H(\mathbf{r})$ and $\sigma_{zz}(\mathbf{r}) = \sigma_V(\mathbf{r})$. For the previous paragraph model example, discretized into $2 \cdot 10^6$ cells, this simplification involves two conductivity models in the optimization ($\sigma_H(\mathbf{r})$ and $\sigma_V(\mathbf{r})$). In the case of using the same mesh for the parameterization, that example results in a total of $4 \cdot 10^6$ inversion parameters (unknowns). Note that another grid decoupling technique could be applied, for example combining several discretization cells in a single inversion parameter, based on the expected resolution of the data.

1.1.4 3D CSEM Gauss-Newton optimization

The state-of-the-art of CSEM data inversion has experienced an evolution from gradient based optimization methods, like non-linear conjugate gradient (e.g. Gribenko and Zhdanov (2007), Commer and Newman (2008)) and quasi-Newton techniques (e.g. of the use of L-BFGS in Plessix and Mulder (2008)), to Gauss-Newton methods (e.g. Habashy and Abubakar (2004), Liu et al. (2008) and Abubakar et al. (2009)). The main interest for developing higher-order inversion methods is to achieve faster convergence rates without the need of too much manual effort building complex initial models. High-order methods provide more accurate and efficient inversion paths to the local minima. In EM problems, it is easy to find examples of the use of the Gauss-Newton scheme when having a small number of unknowns as when using 2D models (e.g. Abubakar et al. (2005)) or when solving small size 3D problems (e.g. of a magnetotelluric application in Chen et al. (2002), and a CSEM application in Sasaki (2013)). Still, the use of the Gauss-Newton scheme in industrial-survey size 3D CSEM problems is not so common due to the numerical complexity that it involves, as it is shown in the next paragraphs (details in the following chapters).

The inversion of 3D CSEM data is formulated as a constrained non-linear least-squares optimization problem

$$\boldsymbol{\sigma}^\star = \arg \min_{\boldsymbol{\sigma} \in \mathcal{M}} \varepsilon(\boldsymbol{\sigma}), \tag{1.2}$$

where $\boldsymbol{\sigma}$ is a discrete parameterization of the 3D conductivity model, $\mathcal{M}$ represents

the set of possible models compatible with a priori information, and

$$\varepsilon(\boldsymbol{\sigma}) = \varepsilon_{\text{D}}(\boldsymbol{\sigma}) + \lambda\, \varepsilon_{\text{R}}(\boldsymbol{\sigma}), \tag{1.3}$$

is the cost function. This cost function is formed by the regularization misfit term ε_{R}, the "trade-off factor" λ, and the data misfit term

$$\varepsilon_{\text{D}}(\boldsymbol{\sigma}) = \sum_{\kappa} d_{\kappa} d_{\kappa}^{*} = \sum_{\kappa} W_{\kappa} \Delta F_{\kappa} \left(W_{\kappa} \Delta F_{\kappa} \right)^{*}. \tag{1.4}$$

The short hand notation $\kappa = (F, i, f, \mathbf{r_{rx}}, \mathbf{r_{tx}})$ uniquely labels a measurement, * is the complex conjugate, F represents a field ($F = E$ for electric and $F = H$ for magnetic), i are the spatial components (x, y) of the field recordings, f are the frequencies, $\mathbf{r_{rx}}$ is a receiver position, and $\mathbf{r_{tx}}$ is a source position. Here $\boldsymbol{d}$ is a vector that contains N_{κ} data-misfit residuals. W_{κ} is a datum weight (typically inverse standard deviation) and $\Delta F_{\kappa} = F_{\kappa}^{\text{Synth}}(\boldsymbol{\sigma}) - F_{\kappa}^{\text{Obs}}$ represents the difference between synthetic and observed fields.

The regularization term $\varepsilon_{\text{R}}(\boldsymbol{\sigma})$ introduces a priori information about the solution model $\boldsymbol{\sigma}^{\star}$ (see more details in Zhdanov (2009) and in Portniaguine and Zhdanov (1999)). For example, the regularization can incorporate information on model conductivity ($\varepsilon_{\text{ap model}}$), model smoothness ($\varepsilon_{\text{grad}}$) and conductivity anisotropy ($\varepsilon_{\text{ap aniso}}$),

$$\varepsilon_{\text{R}}(\boldsymbol{\sigma}) = \varepsilon_{\text{ap model}}(\boldsymbol{\sigma}) + \varepsilon_{\text{grad}}(\boldsymbol{\sigma}) + \varepsilon_{\text{ap aniso}}(\boldsymbol{\sigma}). \tag{1.5}$$

In a local optimization scheme (figure 1.2), the solution $\boldsymbol{\sigma}^{*}$ is obtained iteratively, starting from an initial guess $\boldsymbol{\sigma}^{(0)}$ and updating the model with a new step $\boldsymbol{\Delta\sigma}^{(k)}$ at each iteration ($\boldsymbol{\sigma}^{(k+1)} = \boldsymbol{\sigma}^{(k)} + \boldsymbol{\Delta\sigma}^{(k)}$). The equations (details in chapter 5) for getting these model updates in a Gauss-Newton scheme are:

$$\boldsymbol{\mathcal{H}}^{(k)}\, \boldsymbol{\Delta\sigma}^{(k)} = -\boldsymbol{g}^{(k)}, \tag{1.6}$$

$$\boldsymbol{\mathcal{H}}^{(k)} = \boldsymbol{\mathcal{H}}_{\text{D}}^{(k)} + \boldsymbol{\mathcal{H}}_{\text{R}}^{(k)}, \tag{1.7}$$

$$\boldsymbol{g}^{(k)} = \boldsymbol{g}_{\text{D}}^{(k)} + \boldsymbol{g}_{\text{R}}^{(k)}. \tag{1.8}$$

$\boldsymbol{\mathcal{H}}_{\text{D}}$ is the Gauss-Newton data Hessian matrix, $\boldsymbol{\mathcal{H}}_{\text{R}}$ is the Gauss-Newton regularization Hessian matrix, $\boldsymbol{g}_{\text{D}}$ is the data gradient vector and $\boldsymbol{g}_{\text{R}}$ is the regularization gradient vector. The Gauss-Newton Hessian matrix $\boldsymbol{\mathcal{H}} \approx \boldsymbol{\nabla^2}\varepsilon$ is a real $N \times N$ matrix, and the gradient $\boldsymbol{g} = \boldsymbol{\nabla}\varepsilon$ is a real vector of N elements. In this scheme, $\boldsymbol{\mathcal{H}}_{\text{D}}$ involves the largest computational complexity since it is based on the Jacobian matrix:

$$\mathcal{H}_{\text{D}\,n\,n'} = \sum_{\kappa} [\mathcal{J}_{n\,\kappa}\mathcal{J}_{n'\,\kappa}^{*} + \mathcal{J}_{n\,\kappa}^{*}\mathcal{J}_{n'\,\kappa}], \tag{1.9}$$

$$\mathcal{J}_{\kappa\,n} = W_{\kappa}\frac{\partial F_{\kappa}^{\text{Synth}}(\boldsymbol{\sigma})}{\partial \sigma_{n}}. \tag{1.10}$$

The Jacobian $\mathcal{J}$ is a $N_{\kappa} \times N$ complex matrix. As is explained in the following chapters, its straightforward computation requires many Green functions. They are

obtained in $N_{\tt Sim}$ forward modelling simulations, one per each receiver and source position, for the different field components:

$$N_{\tt Sim} = \underbrace{N_F\,N_i\,N_{\tt rx}}_{N_{\tt Sim_{Rx}}} + \underbrace{N_{\tt tx}}_{N_{\tt Sim_{Tx}}} \qquad \text{in case of using a FDTD solver;} \qquad (1.11)$$

$$N_{\tt Sim} = \underbrace{N_F\,N_i\,N_{\tt rx}\,N_f}_{N_{\tt Sim_{Rx}}} + \underbrace{N_{\tt tx}\,N_f}_{N_{\tt Sim_{Tx}}} \qquad \text{in case of using a FDFD solver.} \quad (1.12)$$

The notation N_a denotes the number of unique elements of index a. Note that the number of source positions $N_{\tt tx}$ can be several order of magnitude larger than the number of receiver positions $N_{\tt rx}$.

The data gradient also involves the computation of Green functions, however it can be efficiently computed in $2 \cdot N_{\tt Sim_{Rx}}$ using the adjoint scheme described in Støren et al. (2008). Therefore the computational cost per iteration when applying a gradient based method in a 3D CSEM problem is much smaller than when using the Gauss-Newton method.

The computation of the regularization terms $\mathcal{H}_{\rm R}$ and $g_{\rm R}$ does not need modelling simulations. This cost is really insignificant in comparison to the one associated to the data terms (see Zhdanov (2009) for details).

In addition to the number of forward modelling simulations, another main challenge of applying Gauss-Newton in a large-scale 3D CSEM problem is the memory requirement associated with a large number of Green functions, and the sizes of $\mathcal{J}$ and $\mathcal{H}$. For the previous subsections 3D CSEM survey example, where the number of inversion parameters is $4 \cdot 10^6$, and the number of data residuals is 10^7, an explicit computation of the Jacobian requires 291 TB (using 4 bytes precision), while the storage of the Hessian matrix requires 29.1 TB (using 4 bytes precision and taking into account the matrix symmetry). Moreover, the solution of a dense matrix linear-system (eq. 1.6) when using a large number of inversion parameters, precludes the application of a direct solver and involves a significant time complexity.

1.1.5 Reducing the computational cost of 3D CSEM Gauss-Newton

The interest in applying higher order methods, for the inversion of regular-size survey 3D CSEM data, has motivated the research on techniques to reduce the high numerical complexity associated with the application of the Gauss-Newton method. The following paragraphs is an overview of the main works that have been done in relation with this topic.

In Li et al. (2011), the authors propose a compressed implicit Jacobian calculation that reduces the memory complexity involved in a 3D CSEM Gauss-Newton optimization, at the cost of increasing the computational time. It uses the conjugate-gradient line-search (CGLS) method for solving eq. 1.6, but avoids the explicit construction of the Hessian or the Jacobian matrix. It performs the products $\mathcal{J}\boldsymbol{u} = \boldsymbol{v}$ and $\mathcal{J}^T\boldsymbol{v}$ in an efficient way. Firstly it multiplies the Green functions relative to receiver positions by

the iteration vector a, and then reusing these intermediate results in the product $\mathcal{J}^T \boldsymbol{v}$. The memory required is $\mathcal{O}((N_{\text{Sim}_{\text{Rx}}} + N_{\text{Sim}_{\text{Tx}}}) \cdot N)$, which is less than in the case of storing the Hessian or Jacobian matrix when $N_\kappa < N$. In addition this paper applies the adaptive cross approximation (ACA, Bedendorf (2000)) to improve the efficiency of the proposed method. Based on the smoothness of the fields, the ACA approximation allows a compress representation of the field matrices that form the Jacobian matrix.

Abubakar et al. (2012) presents a model-compression scheme, extended for 3D models in Lin et al. (2013), that is able to reduce the computational complexity (time and memory) of the Gauss-Newton method by at least one order of magnitude. This proposal consists of transforming the cell (pixel) basis representation of the resistivity model into another discrete-basis function representation with compressive capabilities. Based on the limited resolution of CSEM, it reduces the number of inversion parameters by truncating these basis functions without significantly affecting the inversion results. The authors consider the discrete Fourier transform (DFT) and the discrete wavelet transform (DWT) for the Haar wavelet (Haar, 1910) and the Daubechies 4 wavelet (Daubechies, 1992). A comparison of synthetic and real data inversion results is presented, where the wavelet transform shows the best performance.

CSEM data can be acquired in the frequency-domain (fCSEM) or in the time domain (tCSEM). Although both responses can be related to each other using the Fourier transform, a transformation of tCSEM data to the frequency domain requires an appropriated preprocessing (e.g. Zach et al. (2008)) that ensures a correct data interpretation. When trying to invert tCSEM data in the time-domain (i.e. without transforming the data to the frequency-domain), the number of data samples increases several order of magnitude with respect to fCSEM data. Zaslavsky et al. (2013) presents a method that allows to perform a 3D tCSEM data Gauss-Newton inversion with a similar cost of inverting fCSEM data. This method uses a space-time data compression in association with a fast 3D forward solver that is based on the rational Krylov subspace reduction (RKSR) algorithm. RKSR is a method for the solution of non-dispersive Maxwell's systems, based on the projection of a system of time-convolution equations onto a small subspace of the Laplace domain solutions.

These previous works are good alternatives to reduce the computational cost of 3D CSEM data inversion when using the Gauss-Newton method. However these improvements are not always enough when trying to run a large-scale 3D CSEM inversion. For the example survey described in this introduction, with $4 \cdot 10^6$ inversion parameters and 10^7 data residuals (in the frequency-domain), the implicit Jacobian representation (Li et al., 2011) significantly reduces the required memory but it would also increase the time complexity several orders of magnitude. For the same survey example, if using the model-compression described in Lin et al. (2013), a reduction of one order of magnitude of the memory complexity would result in a Hessian matrix of 2.91 TB, far beyond a common computer capacity (not considering clusters). Moreover, the solution of the equation system (eq. 1.6) at each iteration involves an important computational cost too. The development of a simple method for obtaining

a pre-conditioner would help to reduce that cost.

As is explained in this introduction, the numerical complexity of a 3D CSEM data Gauss-Newton inversion is also due to the number of forward modelling simulations required. The development of faster modelling solvers tries to reduce this complexity. Still there are more options to explore as reducing the number of required simulations, for example based on a stochastic source selection of sources to be simulated at each iteration.

1.1.6   Higher order optimization methods than Gauss-Newton

The same arguments that motivate the development of the Gauss-Newton optimization method for the inversion of CSEM data could be used to justify the research on higher order methods. It is well known that higher order methods than Gauss-Newton are not typically efficient due to the relation cost versus convergence rate. However, there are problems where properties like system matrix (tensor) sparsity can make it affordable (Gundersen and Steihaug, 2011).

As it was previously commented, Gauss-Newton optimization is the highest order method typically in use for inversion of 3D CSEM data. It would be useful to explore the possibilities of a higher order method, in order to compare its inversion results with Gauss-Newton. Although these implementations initially might not have a direct application in the industry context, they would help to evaluate the accuracy of the Gauss-Newton convergence paths. In addition, it would contribute to the understanding of the numerical complexity and performance of higher order methods in 3D CSEM.

Beyond the Gauss-Newton method, it is necessary to compute higher order derivatives of the residuals with respect to the model parameters. These derivatives require the computation of tensor Green functions for the entire domain, i.e. compute the electromagnetic field when situating a unit dipole moment source at any spatial position in the model. Note that for these methods, these calculations represent a challenge since the number of Green functions is several order of magnitude larger than in the Gauss-Newton method where the Green functions are computed placing a unit dipole moment source only at each source and receiver spatial position. However, the performance achieved by current direct solvers (Grayber and Streich, 2012) starts to make computations of a large number of Green functions feasible.

As is detailed in chapter 5, another challenge is the memory cost required to store the large number of Green functions and the dense multi-dimensional matrices that appear in higher order methods than 3D CSEM Gauss-Newton. Moreover, these methods have the difficulty of solving linear equation systems with indefinite matrices.

1.2 Scope

The forward modelling simulations involved in a 3D CSEM data Gauss-Newton inversion is one of the main contribution to its computational cost. The first part of this PhD research is focused on the inversion aspects of the problem and not specifically on the modelling solvers. There is an associated research project, also funded through PETROMAKS, that studies efficient modelling approaches.

The computational complexity of a large-scale 3D CSEM data Gauss-Newton inversion is due to: a) The large number of forward modelling simulations; b) The large number of inversion parameters; c) Its distributed computation.

The aim of this research is to achieve an efficient inversion algorithm that ideally reduces the resource requirements in these issues. This research is primarily focused on the improvement of those aspects related to a) and b).

The second part of this PhD research is to develop even higher-order methods than Gauss Newton for 3D CSEM problems, as Newton and Halley class optimization methods. These implementations represent a challenge because of the large numerical complexity involved.

The development of high-order optimizations methods beyond the Gauss-Newton approximation is useful to have the possibility of comparing their results with the ones obtained with current implementations of Gauss-Newton.

1.3 Outline

This PhD thesis is divided into 6 parts: an introduction, four chapters, each one corresponding to an individual paper, and a future work description.

This introduction (chapter 1) includes an explanation of the topics that are going to be addressed, and the motivation of this PhD research, based on the interest for developing efficient high-order 3D CSEM data inversion schemes. It shows the large numerical complexity that a 3D CSEM data Gauss-Newton (GN) inversion involves, and a summary of the main previous papers that other authors have done to cope with it. Moreover, the research on higher order optimization methods than 3D CSEM data Gauss-Newton inversion is motivated. The scope of this thesis is also defined in this chapter.

In order to reduce the large numerical complexity that a 3D CSEM GN inversion involves, in chapter 2 a low-rank approximation to the GN data Hessian matrix is proposed. This approximation is based on the simulation of phase-encoded groups of sources instead of incorporating sources individually. The idea is similar to the "super-shot" technique that is applied in seismic full-wave inversion. The feasibility of the approach is demonstrated by numerical examples, showing the potential of reducing the number of forward simulations and memory use (Green functions and Jacobian storage) by two orders of magnitude.

In chapter 3 the presented low-rank approximation to the GN data Hessian matrix

is further researched. Several strategies to form the group of sources are studied and it is concluded that grouping based on maximum distance has an eigenvalue distribution closer to the standard GN data Hessian matrix one. It is also described how the number of groups of sources determinates the reduction of the computational cost, and how the approximation error can be reduced by increasing the number of source groups.

A large-scale 3D CSEM data GN inversion could potentially benefit the low-rank method introduced in the two previous chapters. For these purpose, additional research is necessary, mainly to determinate how to cope with the large number of inversion parameters involved in a realistic survey, and to validate this low-rank approach with synthetic and realistic CSEM data inversions. In chapter 4, it is studied all the aspects relative to the practical use of the low-rank approximation to the GN data Hessian matrix in a realistic problem. This chapter introduces a matrix free recursive direct solver that allows to reduce the memory complexity associated with a large number of inversion parameters. Moreover, an approximation of the regularization Hessian matrix is suggested, demonstrating a dramatic reduction on the memory use. This work includes an analysis of the influence of the number of distant-source groups in the eigenvalue decomposition of the GN data Hessian matrix, and a qualitative analysis of the introduced cross-talk error. Synthetic and realistic survey data results demonstrate the feasibility of this low-rank method.

To date, the 3D CSEM GN inversion is the highest-order method in use. The acquired knowledge and the results presented in the previous chapters, together with the interest for developing methods able to achieve faster convergence rates, drove this PhD research to explore beyond state-of-the-art frontiers of 3D CSEM inversion. The paper presented in chapter 5 is focused on the development of the Newton and Halley class methods that so far does not seem to have been applied in 3D CSEM. Higher-order methods than GN are not generally considered efficient due to the computational cost involved. As it is shown in this chapter, higher-order methods than GN require one forward modelling simulation when a unit dipole moment source is situated in every spatial position (tensor Green functions), and the construction of large system of equations. It is shown that the use of forward modelling direct solvers, an efficient construction of the different linear system elements, and a trust-region solver able to handle negative curvature of the system matrices, allow to apply the Newton and Halley class methods for a moderate size 3D CSEM data inversion (~ 50000 inversion parameters). Numerical simulation results show that the different approaches drives the convergence through different paths that may end in different minima.

Finally, the chapter 6 describes several directions for future extensions of this research. Though this work proposes a solution for the main challenges that a state-of-the-art large-scale 3D CSEM data inversion presents, there are some tests and validations, beyond the scope of this PhD research, that can be useful to explore.

1.4 References

Abubakar, A., Habashy, T. M., Druskin, V. L., Alumbaugh, D., Zhang, P., Wilt, M., Denaclara, H., Nichols, E., and Knizhnerman, L. A fast and rigorous 2.5d inversion algorithm for cross-well electromagnetic data. In *SEG Technical Program Expanded Abstracts*, pages 534–537, 2005. doi: 10.1190/1.2144374. URL http://dx.doi.org/10.1190/1.2144374.

Abubakar, A., Habashy, T. M., Li, M., and Liu, J. Inversion algorithms for large-scale geophysical electromagnetic measurements. *IOP Science Inverse Problems*, 25(12):123012, 2009. doi: 10.1088/0266-5611/25/12/123012. URL http://dx.doi.org/10.1088/0266-5611/25/12/123012.

Abubakar, A., Habashy, T. M., Lin, Y., and Li, M. A model-compression scheme for nonlinear electromagnetic inversions. *Geophysics*, 77(5):E379–E389, 2012. doi: 10.1190/geo2011-0494.1. URL http://dx.doi.org/10.1190/geo2011-0494.1.

Avdeev, D. B. Three-dimensional electromagnetic modelling and inversion from theory to application. *Surveys in Geophysics*, 26(6):767–799, 2005. ISSN 0169-3298. doi: 10.1007/s10712-005-1836-x. URL http://dx.doi.org/10.1007/s10712-005-1836-x.

Bedendorf, M. Approximation of boundary element matrices. *Numerische Mathematik*, 86(4):565–589, 2000. doi: 10.1007/PL00005410. URL http://dx.doi.org/10.1007/PL00005410.

Börner, R.-U. Numerical modelling in geo-electromagnetics: Advances and challenges. *Surveys in Geophysics*, 31(2):225–245, 2010. ISSN 0169-3298. doi: 10.1007/s10712-009-9087-x. URL http://dx.doi.org/10.1007/s10712-009-9087-x.

Brown, V., Key, K., and Singh, S. Seismically regularized controlled-source electromagnetic inversion. *Geophysics*, 77(1):E57–E65, 2012. doi: 10.1190/GEO2011-0081.1. URL http://dx.doi.org/10.1190/GEO2011-0081.1.

Chen, J., Haber, E., and Oldenburg, D. W. Three-dimensional numerical modelling and inversion of magnetometric resistivity data. *Geophysical Journal International*, 149(3):679–697, 2002. doi: 10.1046/j.1365-246X.2002.01688.x. URL http://dx.doi.org/10.1046/j.1365-246X.2002.01688.x.

Commer, M. and Newman, G. A. New advances in three-dimensional controlled-source electromagnetic inversion. *Geophysical Journal International*, 172(2):513–535, 2008. doi: 10.1111/j.1365-246X.2007.03663.x. URL http://dx.doi.org/10.1111/j.1365-246X.2007.03663.x.

Constable, S. and Srnka, L. An introduction to marine controlled-source electromagnetic methods for hydrocarbon exploration. *Geophysics*, 72(2):WA3–WA12, 2007. doi: 10.1190/1.2432483. URL http://dx.doi.org/10.1190/1.2432483.

Daubechies, I. *Ten Lectures on Wavelet*. CBMS-NSF Regional Conference Series in Applied Mathematics, 1992. ISBN 978-0-89871-274-2. URL http://dx.doi.org/10.1137/1.9781611970104.

Eidesmo, T., Ellingsrud, T., MacGregor, L., Constable, S., Sinha, M., Johansen, S. E., Kong, F., and Westerdahl, H. Sea bed logging (sbl), a new method for remote and direct identification of hydrocarbon filled layers in deepwater areas. *First Break*, 20(3):144–152, 2002. URL http://fb.eage.org/publication/content?id=25008.

Ellis, M., Ruis, F., Nanduri, S., Keirstead, R., Azizov, I., Frenkel, M., and MacGregor, L. Importance of anisotropic rock physics modelling in integrated seismic and csem interpretation. *First Break*, 29(6):87–95, 2011. URL http://fb.eage.org/publication/content?id=51279.

Fanavoll, S., Hesthammer, J., Danielsen, J., and Stefatos, A. Controlled source electromagnetic technology and hydrocarbon exploration efficiency. *First Break*, 28(5):61–69, 2010. URL http://fb.eage.org/publication/content?id=38985.

Fanavoll, S., Gabrielsen, P., and Ellingsrud, S. The impact of CSEM on exploration decisions and seismic: two case studies from the barents sea. *First Break*, 32(11):105–110, 2014. URL http://fb.eage.org/publication/content?id=78587.

Grayber, A. and Streich, R. Comparison of iterative and direct solvers for 3D CSEM modeling. In *SEG Technical Program Expanded Abstracts*, pages 1–6, 2012. doi: 10.1190/segam2012-0727.1. URL http://dx.doi.org/10.1190/segam2012-0727.1.

Gribenko, A. and Zhdanov, M. Rigorous 3D inversion of marine CSEM data based on the integral equation method. *Geophysics*, 72(2):WA73–WA84, 2007. doi: 10.1190/1.2435712. URL http://dx.doi.org/10.1190/1.2435712.

Gundersen, G. and Steihaug, T. On large-scale scale unconstrained optimization problems and higher order methods. *Optimization Methods and Software*, 35(3):337–358, 2011. doi: 10.1080/10556780903239071. URL http://doi.org/10.1080/10556780903239071.

Haar, A. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69(3):331–371, 1910. ISSN 0025-5831. doi: 10.1007/BF01456326. URL http://dx.doi.org/10.1007/BF01456326.

Habashy, T. M. and Abubakar, A. A general framework for constraint minimization for the inversion of electromagnetic measurements. *Inverse Problems*, 46:265–312, 2004. doi: 10.2528/PIER03100702. URL http://dx.doi.org/10.2528/PIER03100702.

Hesthammer, J., Stefatos, A., Boulaenko, M., Fanavoll, S., and Danielsen, J. CSEM performance in light of well results. *The Leading Edge*, 29(1):34–41, 2010. doi: 10.1190/1.3284051. URL http://dx.doi.org/10.1190/1.3284051.

Hölz, S., Swidinsky, A., Sommer, M., Jegen, M., and Bialas, J. The use of rotational invariants for the interpretation of marine csem data with a case study from the north alex mud volcano. *Geophysical Journal International*, 201(1):224–245, 2015. URL http://dx.doi.org/10.1093/gji/ggv015.

Hoversten, G., Myer, D., Key, K., Hermann, O., Hobbet, R., and Alumbaugh, D. CSEM & MMT base basalt imaging. In *75th EAGE Conference & Exhibition incorporating SPE EUROPEC*, 2013. doi: 10.3997/2214-4609.20130137. URL http://dx.doi.org/10.3997/2214-4609.20130137.

Hu, W., Abubakar, A., and Habashy, T. M. Joint electromagnetic and seismic inversion using structural constraints. *Geophysics*, 74(6):R99–R1090, 2009. doi: 10.1190/1.3246586. URL http://dx.doi.org/10.1190/1.3246586.

Johnson, H. M. A history of well logging. *Geophysics*, 27(4):507–527, 1962. doi: 10.1190/1.1439054. URL http://dx.doi.org/10.1190/1.1439054.

Li, M., Abubakar, A., Liu, J., Pan, G., and Habashy, T. M. A compressed implicit jacobian scheme for 3D electromagnetic data inversion. *Geophysics*, 76(3):F173–F183, 2011. doi: 10.1190/1.3569482. URL http://dx.doi.org/10.1190/1.3569482.

Lin, Y., Li, M., Abubakar, A., and Habashy, T. A wavelet-based model compression method for three-dimensional electromagnetic data inversion. In *SEG Technical Program Expanded Abstracts*, pages 707–712, 2013. doi: 10.1190/segam2013-0395.1. URL http://dx.doi.org/10.1190/segam2013-0395.1.

Liu, J., Habashy, T., Zaslavsky, M., and Druskin, V. A three-dimensional multiplicative regularized non-linear inversion algorithm for cross-well electromagnetic and controlled-source electromagnetic applications. In *SEG Technical Program Expanded Abstracts*, pages 584–588, 2008. doi: 10.1190/1.3063720. URL http://dx.doi.org/10.1190/1.3063720.

Løseth, L. O., Pedersen, H. M., Ursin, B., Amundsen, L., and Ellingsrud, S. Low-frequency electromagnetic fields in applied geophysics: Waves or diffusion? *Geophysics*, 71(4):W29–W40, 2006. doi: 10.1190/1.2208275. URL http://dx.doi.org/10.1190/1.2208275.

Mittet, R. and Morten, J. P. The marine controlled-source electromagnetic method in shallow water. *Geophysics*, 78(2):E67–E77, 2013. doi: 10.1190/GEO2012-0112.1. URL http://dx.doi.org/10.1190/GEO2012-0112.1.

Mittet, R. and Morten, J. P. Detection and imaging sensitivity of the marine csem method. *Geophysics*, 77(6):E411–E425, 2012. doi: 10.1190/GEO2012-0016.1. URL http://dx.doi.org/10.1190/GEO2012-0016.1.

Morten, J. P., Roth, F., Timko, D., Pacurar, C., Nguyen, A. K., and Olsen, P. A. 3D reservoir characterization of a North Sea oil field using quantitative seismic & CSEM interpretation. In *SEG Technical Program Expanded Abstracts*, pages 1903–1907, 2011. doi: 10.1190/1.3627578. URL http://library.seg.org/doi/abs/10.1190/1.3627578.

Morten, J., Twarz, C., Ricoy-Paramo, V., and Sun, S. Improved resolution salt imaging from 3D CSEM anisotropic inversion. In *75th EAGE Conference & Exhibition*, 2013. doi: 10.3997/2214-4609.20130139. URL http://dx.doi.org/110.3997/2214-4609.20130139.

Myer, D., Constable, S., and Key, K. Electrical impedance tomography of the loihi underwater volcano. cruise report. Unpublished, 2006. URL http://marineemlab.ucsd.edu/Projects/Loihi2006/LoihiCruiseReport.pdf.

Plessix, R.-. and Mulder, W. A. Resistivity imaging with controlled-source electromagnetic data: depth and data weighting. *IOP Science Inverse Problems*, 24(3): 034012, 2008. doi: 10.1088/0266-5611/24/3/034012. URL http://dx.doi.org/10.1088/0266-5611/24/3/034012.

Portniaguine, O. and Zhdanov, S. Focusing geophysical inversion images. *Geophysics*, 64(3):874–887, 1999. doi: 10.1190/1.1444596. URL http://dx.doi.org/10.1190/1.1444596.

Sasaki, Y. 3D inversion of marine CSEM and MT data: An approach to shallow-water problem. *Geophysics*, 78(1):E59–E65, 2013. doi: 10.1190/geo2012-0094.1. URL http://dx.doi.org/10.1190/geo2012-0094.1.

Støren, T., Zach, J., and Maaø, F. Gradient calculations for 3D inversion of CSEM data using a fast finite-difference time-domain modelling code. In *70th EAGE Conference & Exhibition, P194*, 2008. doi: 10.3997/2214-4609.20147963. URL http://dx.doi.org/10.3997/2214-4609.20147963.

Tarantola, A. *Inverse problem theory and methods for model parameter estimation*. SIAM, 1st edition, 2005. ISBN 978-0898715729. URL http://www.ipgp.jussieu.fr/~tarantola/Files/Professional/Books/.

Tompkins, M. J. Effects of vertical anisotropy on marine active source electromagnetic data and inversions. In *66th EAGE Technical Program Expanded Abstracts*, 2004. URL http://earthdoc.eage.org/publication/publicationdetails/?publication=2041.

Tompkins, M. J. The role of vertical anisotropy in interpreting marine controlled-source electromagnetic data. In *SEG Technical Program Expanded Abstracts*, pages 514–517, 2005. doi: 10.1190/1.2144369. URL http://dx.doi.org/10.1190/1.2144369.

Trainor-Guitton, W. and Hoversten, G. M. Stochastic inversion for electromagnetic geophysics: Practical challenges and improving convergence efficiency. *Geophysics*, 76(6):F373–F386, 2010. doi: 10.1190/geo2010-0223.1. URL http://dx.doi.org/10.1190/geo2010-0223.1.

Weiss, C. J. and Newman, G. A. Electromagnetic induction in fully 3-D anisotropic earth. *Geophysics*, 67(4):1104–1114, 2002. doi: 10.1190/1.1500371. URL http://dx.doi.org/10.1190/1.1500371.

Zach, J. J., Roth, F., and Yuan, H. Preprocessing of marine CSEM data and model preparation for frequency-domain 3D. In *PIERS proceedings*, pages 144–148, 2008. URL http://piers.org/piersproceedings/download.php?file=cGllcnMyMDA4Q2FtYnJpZGdllfDJQMl8wMTQ0LnBkZnwwNzEyMTkwNjQzMjU=.

Zaslavsky, M., Druskin, V., Abubakar, A., Habashy, T., and Simoncini, V. Large-scale gauss-newton inversion of transient controlled-source electromagnetic measurement data using the model reduction framework. *Geophysics*, 78(4):E161–E171, 2013. doi: 10.1190/geo2012-0257.1. URL http://dx.doi.org/10.1190/geo2012-0257.1.

Zhdanov, M. S. *Geophysical inverse theory and regularization problems*. Elsevier Science, 1st edition, 2002. ISBN 978-0-444-51089-1. URL http://www.elsevier.com/books/geophysical-inverse-theory-and-regularization-problems/zhdanov/978-0-444-51089-1.

Zhdanov, M. S. *Geophysical Electromagnetic Theory and Methods*. Elsevier Science, 1st edition, 2009. ISBN 9780444529633. URL http://www.elsevier.com/books/geophysical-electromagnetic-theory-and-methods/zhdanov/978-0-444-52963-3.

Ziolkowski, A. and Wright, D. The potential of the controlled source electromagnetic method. *IEEE Signal Processing Magazine*, 29(4):36–52, 2012. doi: 10.1109/MSP.2012.2192529. URL http://dx.doi.org/10.1109/MSP.2012.2192529.

This page is intentionally left blank.

Chapter 2

A low-rank approximation to the Hessian for 3D CSEM Gauss-Newton Inversion

As it is explained in chapter 1, the main challenge when addressing a 3D CSEM data Gauss-Newton inversion is the large numerical complexity involved. In this chapter it is introduced a low-rank approximation to the Gauss-Newton data Hessian matrix that has the potential of significantly reduce that complexity.

Paper #1. Published and presented at 76th EAGE Conference and Exhibition, 16th-19th June 2014, Amsterdam (Netherlands).

Authors: M. Amaya* (Norwegian University of Science and Technology), L. Boman (Electromagnetic Geoservices ASA) and J.P. Morten (Electromagnetic Geoservices ASA)

Abstract: Use of controlled-source electromagnetics in increasingly challenging exploration applications has led to the requirement for more powerful 3D inversion approaches. For 3D cases, application of Gauss-Newton algorithms is limited by the computational cost required to compute the Hessian matrix and solve for the model update. We consider a low-rank approximation to the Hessian matrix, which has the potential to reduce the numerical complexity drastically. The scheme is based on phase encoding groups of sources instead of incorporating sources individually. We demonstrate the feasibility of the approach by numerical examples and present an analysis of the errors introduced by the approximation.

2.1 Introduction

The application of 3D controlled-source electromagnetics (CSEM) to image and characterize targets in increasingly challenging environments has motivated the development of more powerful inversion methods. The target response often represents a small perturbation of the measured signal response from complex background resistivity variations. When the geological understanding is limited, we require the 3D CSEM inversion to reconstruct from the data not only the potential hydrocarbon reservoir target, but also an accurate representation of the background resistivity variation and the structural framework.

The Gauss–Newton optimization algorithm is known to work well for inversion of CSEM data when assumptions of lower spatial dimensionality can be applied (Abubakar et al., 2006; Mittet et al., 2007). When a 3D model description is required, and when the input from state-of-the-art 3D acquisition is to be used, the numerical complexity of the Gauss–Newton algorithm can be very large (Abubakar et al., 2009; Sasaki, 2011). The large size of the Jacobian and Hessian matrices, as well as the number of 3D forward simulations can be a severe limitation. This has been addressed by several authors considering schemes to reduce the numerical cost by e.g. model reparameterization (Lin et al., 2013), and input data decimation (Schwarzbach and Haber, 2011).

Gradient-based approaches to 3D CSEM inversion, like conjugate-gradient and quasi-Newton (Mackie et al., 2007; Støren et al., 2008), are less computationally demanding, and are now commonly used. However, these approaches are most accurate when a good background model has been built. The construction of the background model can be a demanding task if the geology is complex and if little other geophysical data is available.

In this paper we present a Hessian approximation based on the superposition of phase-encoded sources. This approach leads to a low-rank representation of the Hessian matrix, and alleviates the computational cost of constructing and storing this matrix as well as the solution of the Gauss–Newton equation. We show by numerical examples how the approximation is able to capture important features of the Hessian, at a numerical cost that is up to two orders of magnitude smaller than the exact calculation.

2.2 Gauss–Newton optimization and Hessian approximation

The inversion of CSEM data is formulated as an optimization problem $\boldsymbol{\sigma} = \arg\min_{\boldsymbol{\sigma} \in \mathcal{M}} \varepsilon(\boldsymbol{\sigma})$, where $\boldsymbol{\sigma}$ is a 3D conductivity model in the set $\mathcal{M}$ of models compatible with a priori information, and ε is the cost function. The cost function includes both regularization terms and a data misfit term,

$$\varepsilon_{\text{Data}}(\boldsymbol{\sigma}) = \sum_{F,i,f,\mathbf{r}_{\text{rx}},\mathbf{r}_{\text{tx}}} \left| W_i^F(\mathbf{r}_{\text{rx}}|\mathbf{r}_{\text{tx}}, f) \, \Delta F_i(\mathbf{r}_{\text{rx}}|\mathbf{r}_{\text{tx}}, f; \boldsymbol{\sigma}) \right|^2. \qquad (2.1)$$

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

The non-linear optimization problem is solved by iteratively updating the conductivity model. The Gauss–Newton equation for model updates $\boldsymbol{\Delta\sigma}$ is $\mathbf{H}\,\boldsymbol{\Delta\sigma} = -\boldsymbol{g}$ where $\mathbf{H} = \mathbf{J}^{\dagger}\mathbf{J} + \text{c.c.}$ is the Hessian matrix constructed from the Jacobian matrix $\mathbf{J}$, and $\boldsymbol{g} = \sum_{\kappa} W_{\kappa}\Delta F_{\kappa}^{*}(\mathbf{J})_{\kappa} + \text{c.c.}$ is the model parameter gradient of the cost function. The c.c. denotes complex conjugated term. The Jacobian is a complex $N \times M$ matrix where N is the number of data samples, and M is the number of model parameters. The Jacobian can be constructed from Green functions,

$$(\mathbf{J})_{\kappa,\mathbf{r}} = W_i^F\left(\boldsymbol{r}_{\text{rx}}|\boldsymbol{r}_{\text{tx}}, f\right) \sum_m G_{i,m}^{F,J}\left(\mathbf{r}_{\text{rx}}|\mathbf{r}, f\right) \sum_n G_{m,n}^{E,J}\left(\mathbf{r}|\mathbf{r}_{\text{tx}}, f\right) J_n\left(\mathbf{r}_{\text{tx}}, f\right), \quad (2.2)$$

where $\mathbf{r}$ is the position in the model, $G_{m,n}^{F,J}$ denotes the Green function for field F, component m, from a unit electric current source in direction n, and J_n is a component of the source ($m, n = x, y, z$). It is straightforward to generalize the expression in eq. 2.2 to the anisotropic and discrete case. From this expression, we see that explicit construction of the Jacobian requires the Green function associated with every source position $\mathbf{r}_{\text{tx}}$ to be calculated. The Hessian matrix is a real $M \times M$ matrix, with rank given by the number of rows in the Jacobian, i.e. rank($\mathbf{H}$) $= 2\,N_F\,N_f\,N_i\,N_{\text{rx}}\,N_{\text{tx}}$ (Grayver et al., 2013). Here, and throughout, the notation N_a denotes the number of unique elements of index a. For a state-of-the-art 3D CSEM survey and with a realistic model representation, the numerical complexity involved with the construction of $\mathbf{H}$ and the solution for the model update can be very large. The number of forward solutions required can be of order 10^4, and the dense linear system for $\boldsymbol{\Delta\sigma}$ can be of size $10^6 \times 10^6$.

In this paper, we consider a low-rank approximation where sources in eq. 2.2 are combined after encoding with a random phase factor, i.e. we construct $\sum_{\kappa \in g} \mathrm{e}^{\mathrm{i}\phi_{\kappa}}(\mathbf{J})_{\kappa}$ for a group of sources g associated with a receiver channel, and where ϕ_{κ} are uniformly distributed random numbers in the interval $[0, 2\pi)$. The number of source groups N_g and the grouping scheme will be discussed below. Following this approach, the factors of $\mathbf{J}$ associated with the sources in a source group and a specific receiver channel can be calculated from a single simultaneous-source (super-shot) forward solution of the Maxwell equations. We denote the output of such simulation,

$$\tilde{G}_{i,m,\mathbf{r}_{\text{rx}},g}^{F}\left(\mathbf{r}, f\right) = \sum_{n,\,\mathbf{r}_{\text{tx}}\in g} W_i^F\left(\boldsymbol{r}_{\text{rx}}|\boldsymbol{r}_{\text{tx}}, f\right) J_n\left(\mathbf{r}_{\text{tx}}, f\right) \mathrm{e}^{\mathrm{i}\phi_{F,i,\mathbf{r}_{\text{tx}},f}}\, G_{m,n}^{E,J}\left(\mathbf{r}|\mathbf{r}_{\text{tx}}, f\right). \quad (2.3)$$

The approximate Hessian matrix $\tilde{\mathbf{H}}$ following from the Jacobian constructed in this

approach becomes,

$$\tilde{\mathbf{H}}(\mathbf{r}, \mathbf{r}') = \sum_{F,i,f,\mathbf{r}_{\mathrm{rx}},g} \left[\sum_m G_{i,m}^{F,J}\left(\mathbf{r}_{\mathrm{rx}}|\mathbf{r}, f\right) \tilde{G}_{i,m,\mathbf{r}_{\mathrm{rx}},g}^F\left(\mathbf{r}, f\right) \right] \times \qquad (2.4)$$

$$\left[\sum_n G_{i,n}^{F,J}\left(\mathbf{r}_{\mathrm{rx}}|\mathbf{r}', f\right) \tilde{G}_{i,n,\mathbf{r}_{\mathrm{rx}},g}^F\left(\mathbf{r}', f\right) \right]^* + \text{c.c.}$$

The rank of the approximation is given by the number of terms in the outer sum, i.e. $\mathrm{rank}(\tilde{\mathbf{H}}) = 2\, N_F\, N_i\, N_f\, N_{rx}\, N_g$. Note that both the rank and the storage requirement to construct $\tilde{\mathbf{H}}$ scale by N_g instead of N_{tx} as for $\mathbf{H}$. The reduction in numerical complexity from the approximation is described below, but it is realistic that the ratio N_{tx}/N_g can be of order $10 - 100$ meaning a dramatic decrease in complexity. Consider now the errors introduced by the approximation to the Hessian in eq. 2.4. Due to the summation over source positions in $\tilde{G}$, the approximation will include terms involving two different source positions. Such terms are not present in $\mathbf{H}$, see figure 2.1.
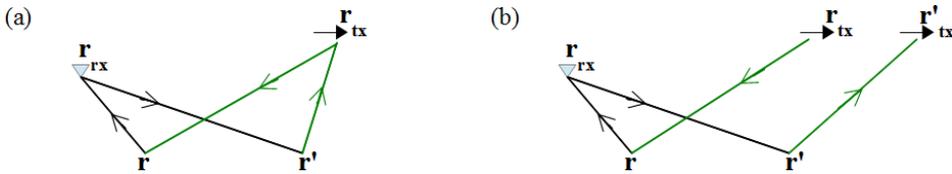
Figure 2.1: (a) Diagram representing terms included in the Hessian $\mathbf{H}$. Arrows correspond to Green functions, with reverse directions indicating complex conjugation. Current factors $W_i^F\left(\mathbf{r}_{\mathrm{rx}}|\mathbf{r}_{\mathrm{tx}}, f\right) J_n\left(\mathbf{r}_{\mathrm{tx}}, f\right) e^{i\phi_{F,i,\mathbf{r}_{\mathrm{tx}},f}}$ are associated with source positions. (b) Diagram representing the additional cross-talk terms introduced into $\tilde{\mathbf{H}}$, where two separate source positions give a contribution.

We refer to these errors as "cross-talk" and denote their contribution η, such that $\tilde{\mathbf{H}} = \mathbf{H} + \eta$. The source-diagonal terms, involving only a single $\mathbf{r}_{\mathbf{tx}}$, are the terms that make up the exact Gauss–Newton Hessian $\mathbf{H}$. The random phase-factors $e^{i\phi_\kappa}$ introduced in eq. 2.3 will cancel in the source-diagonal terms since they enter as an absolute value. However, for the cross-talk terms, where two different source points are involved, the phase factors remain and act to suppress the cross-talk in the outer summation in eq. 2.4. This is similar to applications of phase encoding in seismic modeling, see e.g. Bansal et al. (2013). The number of source-diagonal terms contributing to $\mathbf{H}$ is proportional to N_{tx}. The number of cross-talk terms contributing to η will scale with the number of sources as N_{tx}^2 (assuming $N_g = 1$). However, the magnitude $|\eta|$ should scale linearly in N_{tx} due to the random phase of the cross-talk terms from the $e^{i\phi_\kappa}$ factors, and by analogy to a Gaussian random walk. Further, the number of significant terms contributing to η should be less than N_{tx}^2 due to the exponential decay of the Green functions for cross-talk terms where $|\mathbf{r}_{\mathrm{tx}} - \mathbf{r}'_{\mathrm{tx}}|$ is large. Thus the magnitude $|\eta|$ scales by the number of sources in a source group as $(N_{\mathrm{tx}})^\alpha$ with

$\alpha < 1$. In summary, the asymptotic behaviour of the approximation is feasible since $\lim_{N_{\mathrm{tx}} \to \infty} |\boldsymbol{\eta}|/|\mathbf{H}| = 0$. We can reduce the error of the approximation by increasing the number of source groups N_g. In fact, using the maximum $N_g = N_{\mathrm{tx}}$ makes $\tilde{\mathbf{H}}$ identical to $\mathbf{H}$, but in this case there is no reduction in computational cost. We can optimize the approximation by constructing the source groups with maximum separation between the spatial locations of sources. In this case, each cross-talk contribution shown i figure 2.1 (b) will be smaller compared to a source-diagonal contribution in figure 2.1 (a) by the decay of the Green functions over distance $|\mathbf{r}_{\mathrm{tx}} - \mathbf{r}'_{\mathrm{tx}}|$.

The approximation $\tilde{\mathbf{H}}$ to the Hessian will be better for some parts of the matrix than others. A good implementation into an inverse scheme can use the approximation where the errors are below tolerance, and use the Gauss–Newton Hessian when accuracy is critical. In a standard Gauss–Newton implementation, the number of forward computations, N_{sim}, is mainly driven by the number of source Green functions needed, that is $N_{\mathrm{sim}} \sim N_{\mathrm{tx}}$. Likewise, the memory needed to store the Jacobian matrix scales linearly with N_{tx}. When the Hessian matrix is built using source groups and following the approach described above, both the number of simulations and the memory requirements for the Jacobian scale as N_g instead of N_{tx}. The low-rank approximation allows a Hessian representation using considerably less memory when N_g is small by storing the quantities in square brackets in eq. 2.4. In addition to the savings on number of forward solutions and memory requirements, we expect that solving the Gauss–Newton equation can be done very efficiently by exploiting the low-rank property of the Hessian matrix in eq. 2.4.

2.3 Results

In this section we show model updates computed for an example CSEM survey, using both the Gauss–Newton update and the approximation scheme described above. In the example, the inversion parameterization is a regular grid with cell size 200 m × 200 m × 100 m and the total number of cells is 28275. The survey layout is detailed in figure 2.2 with a total of 10500 data samples.
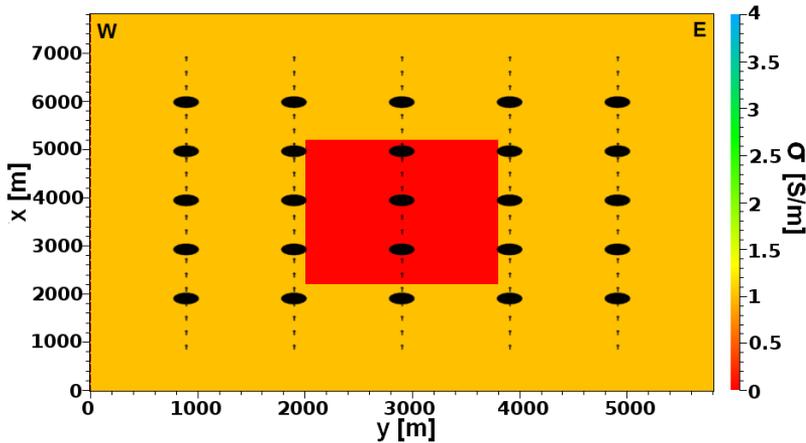
(a) $z = 1500$ m
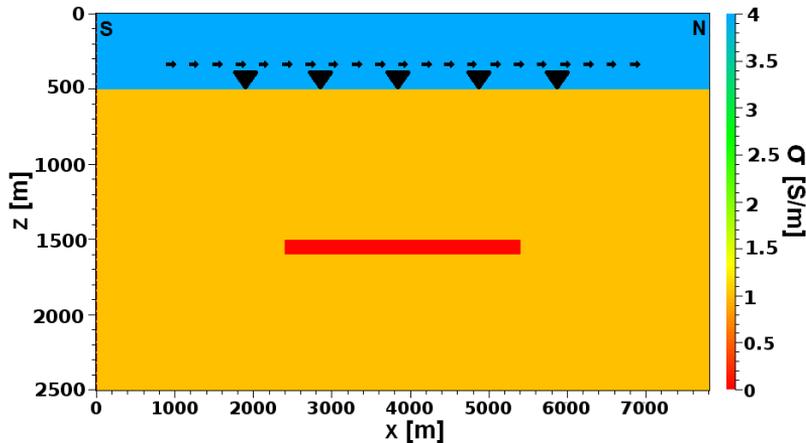
(b) $y = 3000$ m

Figure 2.2: True model and survey layout for the example CSEM survey. The water conductivity is 4 S/m, and the water depth is 500 m. A resistor is located at 1.5 km depth, with dimensions 3 km × 2 km × 0.1 km, and conductivity 0.02 S/m. The formation conductivity is 1 S/m. There are 5 towlines and 25 receivers recording E_x and E_y at 0.25 and 1.0 Hz. The source distance is 300 m along towlines.
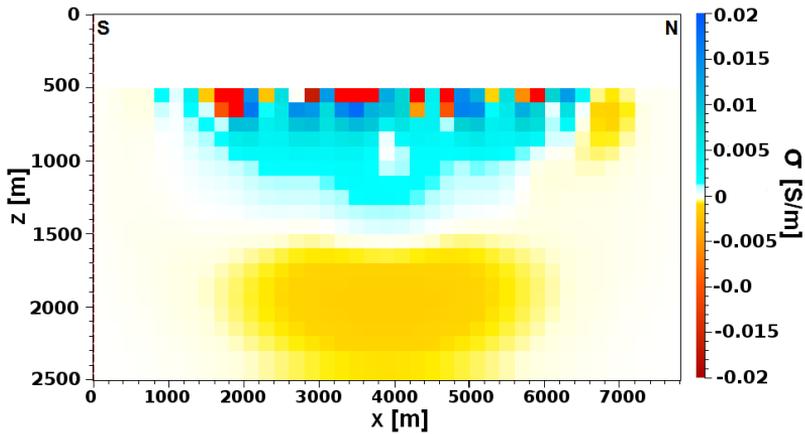
(a) $\tilde{\mathbf{H}}\,\Delta\boldsymbol{\sigma} = -\boldsymbol{g}$
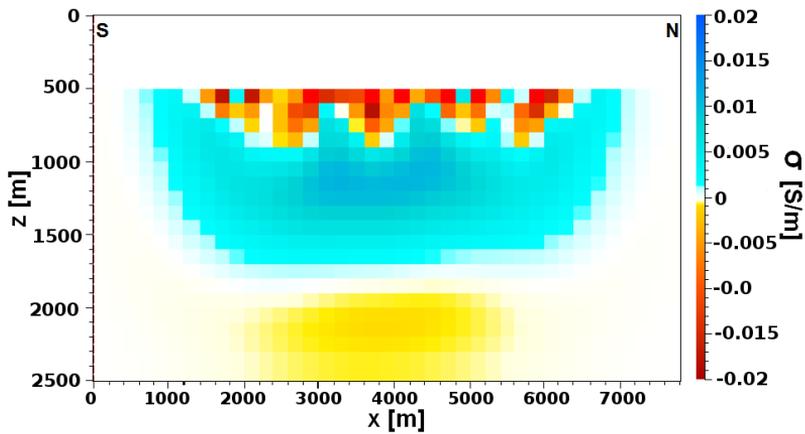
(b) $\mathbf{H}\,\Delta\boldsymbol{\sigma} = -\boldsymbol{g}$
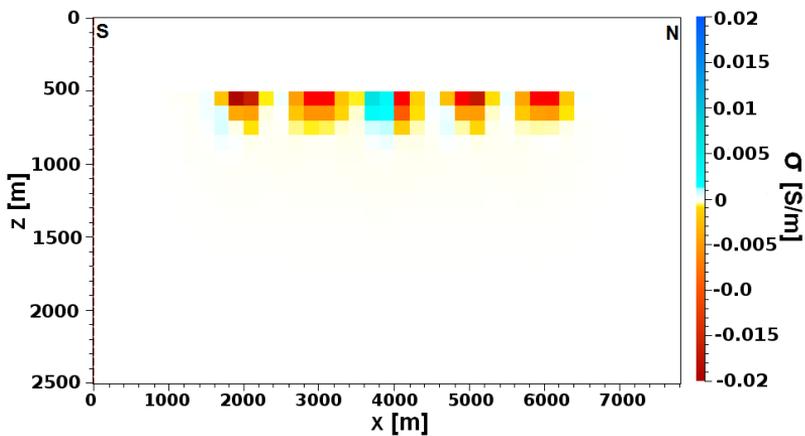
(c) $\Delta\boldsymbol{\sigma} = -\boldsymbol{g}$

Figure 2.3: Model updates $\Delta\boldsymbol{\sigma}$ at $y = 3000$ m.

The Gauss–Newton equation was solved using a conjugate gradient method with a small stabilizer and an initial model with the correct background conductivity. Figure 2.3 shows the solution for three cases using (a) the approximate Hessian matrix $\tilde{\mathbf{H}}$, (b) the exact Gauss–Newton Hessian matrix $\mathbf{H}$, and (c) steepest descent ($\mathbf{H} \rightarrow \mathbf{1}$). The approximate Hessian was obtained using one source group per receiver channel ($N_g = 1$). Comparing (a) with (b) and (c), we see that the solution to the Gauss–Newton equation with the approximate Hessian matrix is qualitatively more similar to a solution with the exact Hessian matrix than a steepest descent solution. In particular, much of the sensitivity information in $\mathbf{H}$ remains in $\tilde{\mathbf{H}}$ as seen at depth. Table 2.1 shows key characteristics for the computational cost of inversion for the example survey as well as for a larger, more realistically sized 3D CSEM survey. As is shown in the table, the number of forward simulations is reduced with a factor 2.6, however in a larger survey the reduction in computational cost can be much larger while keeping the error at the same level.

Case	Numerical cost	Standard GN	Approximate $\tilde{\mathbf{H}}$	Ratio
Example	Forward simulations	260	100	2.6
Example	Jacobian memory	4.4 GB	43.1 MB	105
Realistic	Forward simulations	11800	1000	11.8
Realistic	Jacobian memory	127.4 TB	68.7 GB	1900

Table 2.1: Computational cost comparison for exact Gauss–Newton and the approximation scheme, for the simple example case shown in figure 2.2 and figure 2.3, as well as a large-scale realistic survey.

The data for the realistic survey in table 2.1 were obtained from a survey area of 30 km×20 km×4 km, with 10 towlines at 2 km line separation and a source distance of 100 m along the towlines. In total we obtain 5700 source positions recorded at 100 receiver sites, measuring the horizontal components of electric and magnetic fields at 4 frequencies. The same discretization as in the smaller example is assumed, for two anisotropy components. The survey has a total of 14.59 million data samples. Using three source groups ($N_g = 3$) we keep the simultaneous source separation at 300 m such as in the smaller example survey discussed above. The size of the Hessian will be 2.6 TB, but the Jacobian representation in the approximation scheme offers a 40-fold reduction in size. The approximation error, η, could be reduced by increasing the number of source groups, but the computational cost would increase.

2.4   Conclusions

We have described a low-rank approximation to the Hessian for Gauss–Newton 3D inversion of CSEM data. The scheme is based on superposition of phase-encoded sources, and we have demonstrated the potential to significantly reduce both the number of forward simulations and memory requirements for inversion. We thank Re-

2.5 References

Abubakar, A., Habashy, T., Druskin, V., Alumbaugh, D., Zerelli, A., and Knizhnerman, L. Two-and-half-dimensional forward and inverse modeling for marine CSEM problems. In *SEG Expanded Abstracts*, volume 150, pages 750–754, 2006. doi: 10.1190/1.2370366. URL http://dx.doi.org/10.1190/1.2370366.

Abubakar, A., Habashy, T. M., Li, M., and Liu, J. Inversion algorithms for large-scale geophysical electromagnetic measurements. *IOP Science Inverse Problems*, 25: 123012, 2009. URL http://stacks.iop.org/0266-5611/25/i=12/a=123012.

Bansal, R., Krebs, J., Routh, P., Lee, S., Anderson, J., Baumstein, A., Mullur, A., Lazaratos, S., Chikichev, I., and McAdow, D. Simultaneous-source full-wavefield inversion. *The Leading Edge*, 32(9):1100–1108, 2013. doi: 10.1190/tle32091100. 1. URL http://dx.doi.org/10.1190/tle32091100.1.

Grayver, A. V., Streich, R., and Ritter, O. Three-dimensional parallel distributed inversion of CSEM data using a direct forward solver. *Geophysical Journal International*, 193(3):1432–1446, 2013. doi: 10.1093/gji/ggt055. URL http://dx.doi.org/10.1093/gji/ggt055.

Lin, Y., Li, M., Abubakar, A., and Habashy, T. A wavelet-based model compression method for three-dimensional electromagnetic data inversion. In *SEG Technical Program Expanded Abstracts*, volume 138, pages 707–712, 2013. doi: 10.1190/segam2013-0395.1. URL http://dx.doi.org/10.1190/segam2013-0395.1.

Mackie, R., Watts, M., and Rodi, W. Joint 3D inversion of marine CSEM and MT data. In *SEG Expanded Abstracts*, volume 116, pages 574–578, 2007. doi: 10.1190/1.2792486. URL http://dx.doi.org/10.1190/1.2792486.

Mittet, R., Maulana, H., Brauti, K., and Wicklund, T. A. CMP inversion of marine CSEM data. In *EGM 2007 International Workshop*, April 2007. URL http://www.earthdoc.org/publication/publicationdetails/?publication=41234.

Sasaki, Y. Gauss-newton-based 3D joint inversion of marine CSEM and MT data. In *73rd EAGE Conference & Exhibition, C027*, 2011. doi: 10.3997/2214-4609.20149043. URL http://dx.doi.org/10.3997/2214-4609.20149043.

Schwarzbach, C. and Haber, E. Finite element based inversion for electromagnetic problems using stochastic optimization. In *SEG Expanded Abstracts*, volume 110, pages 567–572, 2011. doi: 10.1190/1.3628145. URL http://dx.doi.org/10.1190/1.3628145.

Støren, T., Zach, J., and Maaø, F. Gradient calculations for 3D inversion of CSEM data using a fast finite-difference time-domain modelling code. In *70th EAGE Conference & Exhibition, P194*, 2008. doi: 10.3997/2214-4609.20147963. URL http://dx.doi.org/10.3997/2214-4609.20147963.

Chapter 3

Efficient computation of approximate low-rank Hessian for 3D CSEM inversion

This chapter is a continuation of chapter 2 where a low-rank approximation to the Gauss-Newton data Hessian matrix is introduced. In this case, several strategies to form the group of sources are studied and it is also described how the number of sources determinates the reduction of the computational cost and the approximation error.

Authors: Manuel Amaya*, Department of Mathematical Sciences, NTNU, Jan Petter Morten and Linus Boman, EMGS ASA

Abstract: Use of controlled-source electromagnetics for increasingly challenging exploration applications has led to the requirement for more powerful 3D inversion approaches. For 3D cases, application of Gauss-Newton algorithms is limited by the computational cost required to compute the Hessian matrix and solve for the model update. We consider a low-rank approximation to the Hessian matrix, which has the potential to reduce the numerical complexity drastically. The scheme is based on computing Green functions for phase-encoded groups of sources instead of incorporating sources individually. We describe the implementation and demonstrate the feasibility of the approach by numerical examples. We also give a theoretical analysis of the errors introduced by the approximation and how to mitigate them.

3.1 Introduction

The application of 3D controlled-source electromagnetics (CSEM) to image and characterize targets in increasingly challenging environments has motivated the development of more powerful inversion methods. The target response often represents a small perturbation of the measured signal response from complex background resistivity variations. When the geological understanding is limited, we require the 3D CSEM inversion to reconstruct from the data not only the potential hydrocarbon reservoir target, but also an accurate representation of the background resistivity variation and the structural framework.

The Gauss–Newton optimization algorithm is known to work well for inversion of CSEM data when assumptions of lower spatial dimensionality can be applied (Abubakar et al., 2006; Mittet et al., 2007). When a 3D model description is required, and when the input from state-of-the-art 3D acquisition is to be used, the numerical complexity of the Gauss–Newton algorithm can be very large (Abubakar et al., 2009; Sasaki, 2011). The large size of the Jacobian and Hessian matrices, as well as the number of 3D forward simulations can be a severe limitation. This has been addressed by several authors considering schemes to reduce the numerical cost by e.g. model reparameterization (Lin et al., 2013), and input data decimation (Schwarzbach and Haber, 2011). Gradient-based approaches to 3D CSEM inversion, like conjugate-gradient and quasi-Newton are less computationally demanding, and are now commonly used (Mackie et al., 2007; Støren et al., 2008). However, these approaches are most accurate when a good background model has been built. The construction of the background model can be a demanding task if the geology is complex and if little other geophysical data is available.

In this paper we present a Hessian approximation based on the superposition of phase-encoded sources. This approach leads to a low-rank representation of the Hessian matrix, and alleviates the computational cost of constructing and storing this matrix as well as the solution of the Gauss–Newton equation. We show by numerical examples how the approximation is able to capture important features of the Hessian, at a numerical cost that is up to two orders of magnitude smaller than the exact calculation. The low-rank approximation was introduced by Amaya et al. (2014) and is here expanded to include a more detailed consideration of the grouping of sources and the effect the grouping has on accuracy and on forward modeling.

3.2 Gauss–Newton optimization and Hessian approximation

The inversion of CSEM data is formulated as an optimization problem

$$\boldsymbol{\sigma} = \arg \min_{\boldsymbol{\sigma} \in \mathcal{M}} \varepsilon(\boldsymbol{\sigma}), \tag{3.1}$$

where $\boldsymbol{\sigma}$ is a 3D conductivity model in the set $\mathcal{M}$ of models compatible with a priori information, and ε is the cost function. This cost function includes both regularization

terms and data misfit terms that are dependent on the observed data,

$$\varepsilon_{\text{Data}}\left(\boldsymbol{\sigma}\right) = \sum_{F,i,f,\mathbf{r}_{\text{rx}},\mathbf{r}_{\text{tx}}} \left|W_i^F\left(\mathbf{r}_{\text{rx}}|\mathbf{r}_{\text{tx}},f\right)\Delta F_i\left(\mathbf{r}_{\text{rx}}|\mathbf{r}_{\text{tx}},f;\boldsymbol{\sigma}\right)\right|^2. \tag{3.2}$$

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

The non-linear optimization problem is solved by iteratively updating the conductivity model. The Gauss–Newton equation for model updates $\boldsymbol{\Delta\sigma}$ is $\mathbf{H}\,\boldsymbol{\Delta\sigma} = -\boldsymbol{g}$ where

$$\mathbf{H} = \mathbf{J}^\dagger\mathbf{J} + \text{c.c.} \tag{3.3}$$

is the Hessian matrix constructed from the Jacobian matrix $\mathbf{J}$, and

$$\boldsymbol{g} = \sum_{\kappa} W_\kappa \Delta F_\kappa^*(\mathbf{J})_\kappa + \text{c.c.} \tag{3.4}$$

is the model parameter gradient of the cost function. The c.c. denotes complex conjugated term. The Jacobian is a complex $N \times M$ matrix where N is the number of data samples, and M is the number of model parameters. The Jacobian can be constructed from Green functions,

$$(\mathbf{J})_{\kappa,\mathbf{r}} = W_i^F\left(\boldsymbol{r}_{\text{rx}}|\boldsymbol{r}_{\text{tx}},f\right)\sum_m G_{i,m}^{F,J}\left(\mathbf{r}_{\text{rx}}|\mathbf{r},f\right) \times \sum_n G_{m,n}^{E,J}\left(\mathbf{r}|\mathbf{r}_{\text{tx}},f\right)J_n\left(\mathbf{r}_{\text{tx}},f\right),$$

$$\tag{3.5}$$

where $\mathbf{r}$ is the position in the model, $G_{m,n}^{F,J}$ denotes the Green function for field F, component m, from a unit electric current source in direction n, and J_n is a component of the source ($m, n = x, y, z$). It is straightforward to generalize the expression in eq. 3.5 to the anisotropic and discrete case. From this expression, we see that explicit construction of the Jacobian requires the Green function associated with every source position $\mathbf{r}_{\text{tx}}$ to be calculated.

The gradient in eq. 3.4 can be computed efficiently and without explicitly constructing the Jacobian by the use of adjoint state modeling. In this case, the factors $W_\kappa \Delta F_\kappa^*$ are used as the source strength in simultaenous source simulations for each receiver component after the synthetic data has been calculated. An implementation of such approach is described in Støren et al. (2008).

The Hessian matrix is a real $M \times M$ matrix, with rank given by the number of rows in the Jacobian (Grayver et al., 2013), i.e.

$$\text{rank}(\mathbf{H}) = 2\,N_F\,N_f\,N_i\,N_{\text{rx}}\,N_{\text{tx}}. \tag{3.6}$$

Here, and throughout, the notation N_a denotes the number of unique elements of index a. For a state-of-the-art 3D CSEM survey and with a realistic model representation, the numerical complexity involved with the construction of $\mathbf{H}$ and the solution for the model update can be very large. The number of forward solutions required can be of order 10^4, and the dense linear system for $\Delta\sigma$ can be of size $10^6 \times 10^6$. In this paper, we consider a low-rank approximation where sources in eq. 3.5 are combined after encoding with a random phase factor, i.e. we construct $\sum_{\kappa \in s} e^{i\phi_\kappa}(\mathbf{J})_\kappa$ for a group of sources s associated with a receiver component, and where ϕ_κ are uniformly distributed random numbers in the interval $[0, 2\pi)$. The number of source groups N_s and the grouping scheme will be discussed below. Following this approach, the factors of $\mathbf{J}$ associated with the sources in a source group and a specific receiver component are calculated from a single simultaneous-source (super-shot) forward solution of the Maxwell equations. We denote the output of such simulation,

$$\tilde{G}^F_{i,m,\mathbf{r}_{\text{rx}},s}(\mathbf{r}, f) = \sum_{n,\,\mathbf{r}_{\text{tx}} \in s} W^F_i(\mathbf{r}_{\text{rx}}|\mathbf{r}_{\text{tx}}, f)\, J_n(\mathbf{r}_{\text{tx}}, f)\, e^{i\phi_{F,i,\mathbf{r}_{\text{tx}},f}}\, G^{E,J}_{m,n}(\mathbf{r}|\mathbf{r}_{\text{tx}}, f).$$

$$(3.7)$$

The approximate Hessian matrix $\tilde{\mathbf{H}}$ following from the Jacobian constructed in this approach becomes,

$$\tilde{\mathbf{H}}(\mathbf{r}, \mathbf{r}') = \sum_{F,i,f,\mathbf{r}_{\text{rx}},s} \left[\sum_m G^{F,J}_{i,m}(\mathbf{r}_{\text{rx}}|\mathbf{r}, f)\, \tilde{G}^F_{i,m,\mathbf{r}_{\text{rx}},s}(\mathbf{r}, f) \right]$$

$$\times \left[\sum_n G^{F,J}_{i,n}(\mathbf{r}_{\text{rx}}|\mathbf{r}', f)\, \tilde{G}^F_{i,n,\mathbf{r}_{\text{rx}},s}(\mathbf{r}', f) \right]^* + \text{c.c.} \qquad (3.8)$$

The rank of the approximation is given by the number of terms in the outer sum,

$$\text{rank}(\tilde{\mathbf{H}}) = 2\, N_F\, N_f\, N_i\, N_{\text{rx}}\, N_s. \qquad (3.9)$$

Note that both the rank and the storage requirement to construct $\tilde{\mathbf{H}}$ scale by N_s instead of N_{tx} as for $\mathbf{H}$. The reduction in numerical complexity from the approximation is described below, but it is realistic that the ratio N_{tx}/N_s can be of order $10 - 100$ meaning a dramatic decrease in complexity.

3.3 Source grouping strategies

Several strategies can be used in order to form the group of sources, as introduced in eq. 3.5. In this section we will explore three configurations, shown schematically in figure 3.1 below.
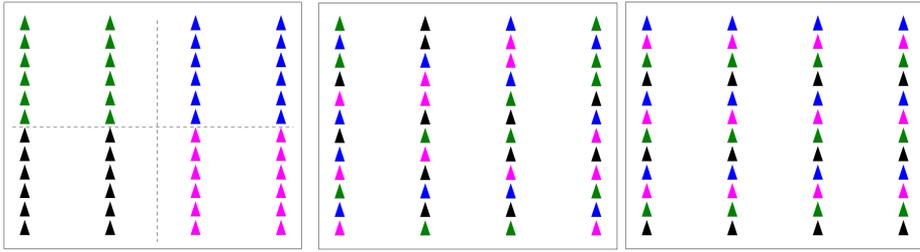
(a) Four groups of sources gathering closest ones in a single group simulation

(b) Four groups of sources gathering randomly the different sources in a single group simulation

(c) Four groups of sources gathering distant sources in a single group simulation

Figure 3.1: Examples of different strategies for source grouping. The triangles represent source positions, and the points with the same color are grouped.

In the example shown in figure 3.1 the sources are arranged in four groups. In figure 3.1(a) the sources are grouped such that each group covers a particular area. In figure 3.1(b) the sources are grouped randomly, and in figure 3.1(c) the groups are selected by maximizing the distance between the sources. Numerical studies described below have shown that using groups based on a largest distance, as in figure 3.1(c), 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 eq. 3.7.

Once the groups of sources have been established, one forward modeling per group of sources and receiver component is performed, using $W_i^F\left(\mathbf{r}_{\mathrm{rx}}|\mathbf{r}_{\mathrm{tx}}, f\right) \mathrm{e}^{\mathrm{i}\phi_{F,i,\mathbf{r}_{\mathrm{tx}},f}}$ as the strength for each source in a group. Linearity of the Maxwell equations implies that this is equivalent to the summation of individual terms in eq. 3.7. The random phase factors $\mathrm{e}^{\mathrm{i}\phi_{F,i,\mathbf{r}_{\mathrm{tx}},f}}$ are sampled independently for ϕ at each source position.

The number of forward simulations needed for building the Hessian matrix in a standard Gauss-Newton implementation, N_{Sim}, and with the low-rank approach, $\tilde{N}_{\mathrm{Sim}}$, are

$$N_{\mathrm{Sim}} = N_{\mathrm{rx}} \cdot N_c \cdot N_F + N_d \cdot N_{\mathrm{tx}} \tag{3.10a}$$

$$\tilde{N}_{\mathrm{Sim}} = N_{\mathrm{rx}} \cdot N_c \cdot N_F + N_d \cdot N_s \cdot N_{\mathrm{rx}}. \tag{3.10b}$$

With realistic values for a modern 3D CSEM survey, N_{Sim} is dominated by the number of sources N_{tx}. The ratio $\tilde{N}_{\mathrm{Sim}}/N_{\mathrm{Sim}}$ then displays a decrease in the number of forward modeling jobs whenever $N_s \cdot N_{\mathrm{rx}}/N_{\mathrm{tx}} < 1$.

It is not only the reduction in number of forward computations which will alleviate the computational complexity. For the Jacobian, or equivalently the Green functions required to construct it, the volume of data in standard Gauss–Newton inversion scales linearly with N_{tx}. Using the approximateion described here, this scaling is instead given by N_s. In the same way, the low-rank approximation allows a Hessian

representation using considerably less memory when N_s is small by storing the quantities in square brackets in eq. 3.8. In addition to the savings on number of forward solutions and memory requirements, we expect that solving the Gauss–Newton equation can be done very efficiently by exploiting the low-rank property of the Hessian matrix in eq. 3.8.

3.4   Qualitative analysis of the approximation accuracy

Consider now the errors introduced by the approximation of the Hessian in eq. 3.8. Due to the summation over source positions in $\tilde{G}$, the approximation will include terms involving two different source positions. Such terms are illustrated schematically in figure 3.2 where (a) shows contributions from one source position, as in a standard Hessian, and (b) shows the additional terms introduced through the low-rank approximation.
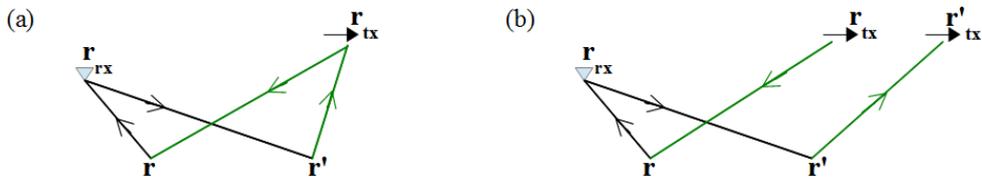
Figure 3.2: (a) Diagram representing terms included in the Hessian $\mathbf{H}$. Arrows correspond to Green functions, with reverse directions indicating complex conjugation. Source strength factors $W_i^F\left(\mathbf{r}_{\mathrm{rx}}|\mathbf{r}_{\mathbf{tx}}, f\right) J_n\left(\mathbf{r}_{\mathbf{tx}}, f\right) \mathrm{e}^{\mathrm{i}\phi_{F,i,\mathrm{r}_{\mathbf{tx}},f}}$ are associated with source positions. (b) Diagram representing the additional cross-talk terms introduced into $\tilde{\mathbf{H}}$, where two separate source positions give a contribution.

We refer to these errors as "cross-talk" and denote their contribution $\boldsymbol{\eta}$, such that $\tilde{\mathbf{H}} = \mathbf{H} + \boldsymbol{\eta}$. The source-diagonal terms, involving only a single $\mathbf{r}_{\mathbf{tx}}$, are the terms that make up the exact Gauss–Newton Hessian $\mathbf{H}$. The random phase-factors $\mathrm{e}^{\mathrm{i}\phi_\kappa}$ introduced in eq. 3.7 will cancel in the source-diagonal terms since they enter as an absolute value. However, for the cross-talk terms, where two different source points are involved, the phase factors remain and act to suppress the cross-talk in the outer summation in eq. 3.8. This is similar to applications of phase encoding in seismic modeling, see e.g. Bansal et al. (2013). The number of source-diagonal terms contributing to $\mathbf{H}$ is proportional to N_{tx}. The number of cross-talk terms contributing to $\boldsymbol{\eta}$ will scale with the number of sources as N_{tx}^2 (assuming $N_s = 1$). However, the magnitude $|\boldsymbol{\eta}|$ should scale by the square root of the number of terms. This is due to the cancellations from random phases $\mathrm{e}^{\mathrm{i}\phi_\kappa}$ of the cross-talk terms and by analogy to a Gaussian random walk. We therefore expect linear scaling $|\boldsymbol{\eta}| \sim N_{\mathrm{tx}}$. Further, the number of significant terms contributing to $\boldsymbol{\eta}$ should be even less due to the exponential decay of the Green functions. This means that for cross-talk terms where $|\mathbf{r}_{\mathrm{tx}} - \mathbf{r}'_{\mathrm{tx}}|$ is large the magnitude of the contribution to $|\boldsymbol{\eta}|$ is very small. Thus the magnitude $|\boldsymbol{\eta}|$

scales by the number of sources in a source group as $(N_{tx})^\alpha$ with $\alpha < 1$. In summary, the asymptotic behavior of the approximation is feasible since $\lim_{N_{tx} \to \infty} |\boldsymbol{\eta}|/|\mathbf{H}| = 0$.

We can reduce the error of the approximation by increasing the number of source groups N_s. In fact, using the maximum $N_s = N_{tx}$ makes $\tilde{\mathbf{H}}$ identical to $\mathbf{H}$, but in this case there is no reduction in computational cost. We can optimize the approximation by constructing the source groups with maximum separation between the spatial locations of sources In this case, each cross-talk contribution shown in figure 3.2 (b) will be smaller compared to a source-diagonal contribution in figure 3.2 (a) by the decay of the Green functions over distance $|\mathbf{r}_{tx} - \mathbf{r}'_{tx}|$. 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 is thus a tuning parameter for the accuracy of the approximation that determines the number of source groups N_s and the strength of the cross-talk noise.
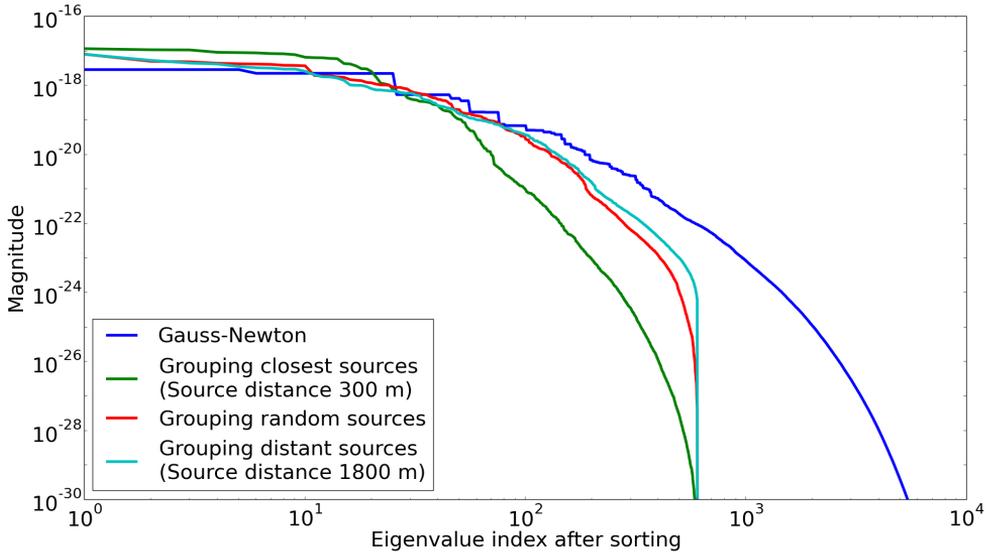
Figure 3.3: Eigenvalue distribution from various source grouping strategies, with $N_s = 12$ and rank $= 600$.

3.5 Results

In this section we will show numerical results for the low-rank Hessian approximation. Figure 3.3 shows the eigenvalue distributions for the three methods of grouping sources illustrated in figure 3.1, and also the distribution for the Hessian of standard Gauss–Newton. In this example the groups are chosen such that the rank of the approximate Hessians is 600. It is seen that source grouping based on maximum distance has an eigenvalue distribution that is closer to that of the standard Gauss–

Newton Hessian, than the other two methods of grouping. This is consistent with the discussion above that suggests that the grouping based on maximum distance is the most accurate. The following results show model updates computed for an example CSEM survey, using both the Gauss–Newton update and the approximation scheme described above. In the example, the inversion parameterization is a regular grid with cell size 200 m $\times$ 200 m $\times$ 100 m and the total number of cells is 28275. The survey layout is detailed in figure 3.4 with a total of 10500 data samples.
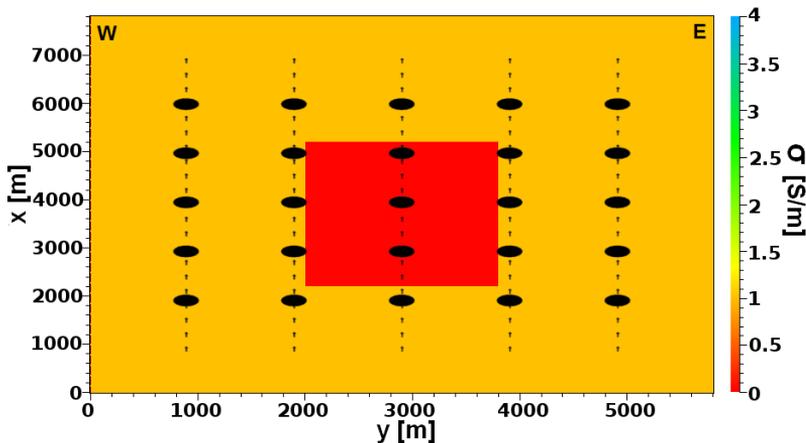
(a) $z = 1500$ m
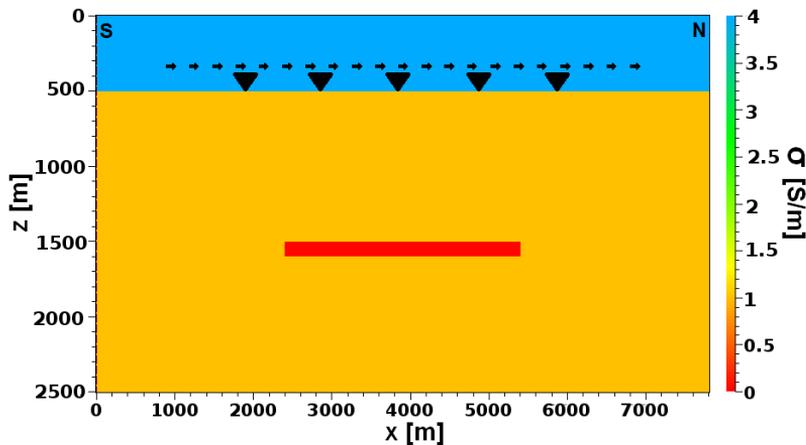
(b) $y = 3000$ m

Figure 3.4: True model and survey layout for the example CSEM survey. The water conductivity is 4 S/m, and the water depth is 500 m. A resistor is located at 1.5 km depth, with dimensions 3 km $\times$ 2 km $\times$ 0.1 km, and conductivity 0.02 S/m. The formation conductivity is 1 S/m. There are 5 towlines and 25 receivers recording E_x and E_y at 0.25 and 1.0 Hz. The source distance is 300 m along towlines.
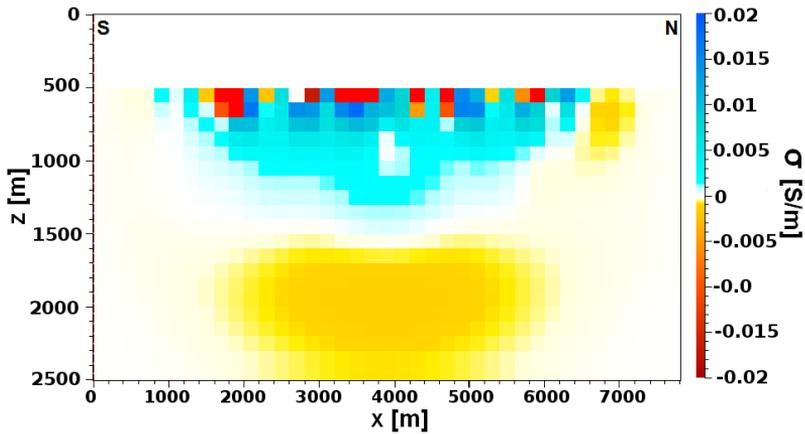
(a) $\tilde{\mathbf{H}}\,\Delta\boldsymbol{\sigma} = -\boldsymbol{g}$
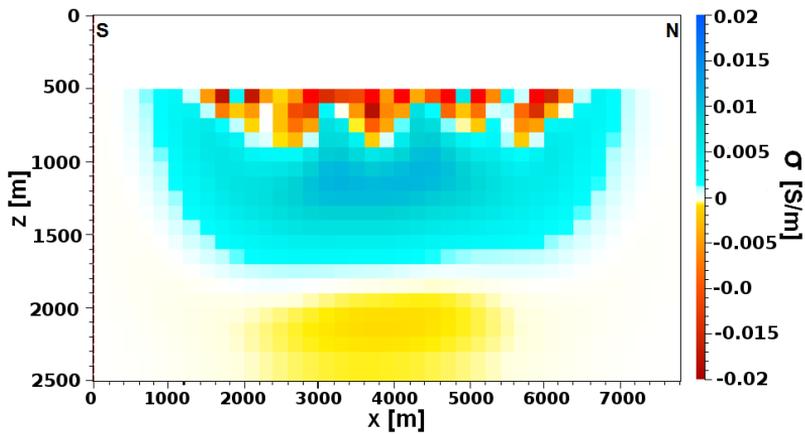
(b) $\mathbf{H}\,\Delta\boldsymbol{\sigma} = -\boldsymbol{g}$
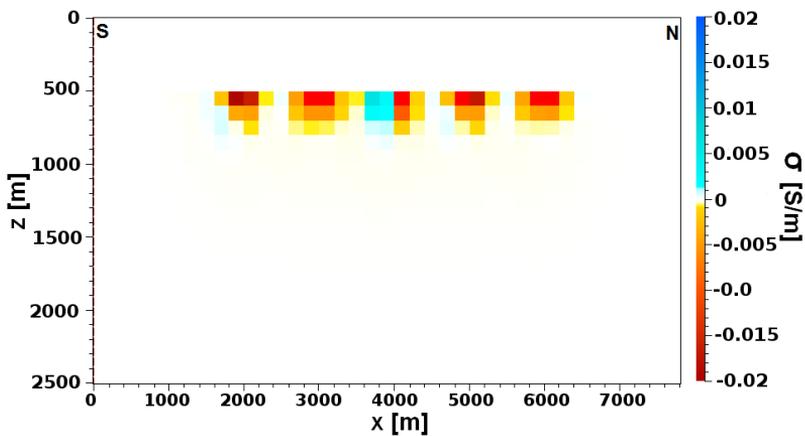
(c) $\Delta\boldsymbol{\sigma} = -\boldsymbol{g}$

Figure 3.5: Model updates $\Delta\boldsymbol{\sigma}$ at $y = 3000$ m.

The Gauss–Newton equation was solved using a conjugate gradient method with a small stabilizer and an initial model with the correct background conductivity. Figure 3.5 shows the solution for three cases using (a) the approximate Hessian matrix $\tilde{\mathbf{H}}$, (b) the exact Gauss–Newton Hessian matrix $\mathbf{H}$, and (c) steepest descent ($\mathbf{H} \rightarrow \mathbf{1}$). The approximate Hessian was obtained using one source group per receiver channel ($N_s = 1$). Comparing (a) with (b) and (c), we see that the solution to the Gauss–Newton equation with the approximate Hessian matrix is qualitatively more similar to a solution with the exact Hessian matrix than a steepest descent solution. In particular, much of the sensitivity information in $\mathbf{H}$ remains in $\tilde{\mathbf{H}}$ as seen at depth. Table 3.1 shows key characteristics for the computational cost of inversion for the example survey as well as for a larger, more realistically sized 3D CSEM survey. As is shown in the table, the number of forward simulations is reduced with a factor 2.6, however in a larger survey the reduction in computational cost can be much larger while keeping the error at the same level. The data for the realistic survey in table 3.1 were obtained

Case	Numerical cost	GN	Approx. $\tilde{\mathbf{H}}$	Ratio
a)	N_{Sim}	260	100	2.6
a)	$\mathbf{J}$ size	4.4 GB	43.1 MB	105
b)	N_{Sim}	11800	1000	11.8
b)	$\mathbf{J}$ size	127.4 TB	68.7 GB	1900

Table 3.1: Computational cost comparison for exact Gauss–Newton and the approximation scheme, for a) the simple example case shown in figure 3.4 and figure 3.5, as well as b) large-scale realistic survey.

from a survey area of 30 km×20 km×4 km, with 10 towlines at 2 km line separation and a source distance of 100 m along the towlines. In total we obtain 5700 source positions recorded at 100 receiver sites, measuring the horizontal components of electric and magnetic fields at 4 frequencies. The same discretization as in the smaller example is assumed, for two anisotropy components. The survey has a total of 14.59 million data samples. Using three source groups ($N_s = 3$) we keep the simultaneous source separation at 300 m such as in the smaller example survey discussed above. The size of the Hessian will be 2.6 TB, but the Jacobian representation in the approximation scheme offers a 40-fold reduction in size. The approximation error, η, could be reduced by increasing the number of source groups, but the computational cost would increase.

3.6 Conclusions

We have described a low-rank approximation to the Hessian for Gauss–Newton 3D inversion of CSEM data. The scheme is based on superposition of phase-encoded sources, and we have demonstrated the potential to significantly reduce both the number of forward simulations and memory requirements for inversion.

3.7   Acknowledgments

3.8   References

Abubakar, A., Habashy, T., Druskin, V., Alumbaugh, D., Zerelli, A., and Knizhnerman, L. Two-and-half-dimensional forward and inverse modeling for marine CSEM problems. In *SEG Expanded Abstracts 2006*, pages 750–754, 2006. doi: 10.1190/1.2370366. URL http://dx.doi.org/10.1190/1.2370366.

Abubakar, A., Habashy, T. M., Li, M., and Liu, J. Inversion algorithms for large-scale geophysical electromagnetic measurements. *IOP Science Inverse Problems*, 25: 123012, 2009. URL http://stacks.iop.org/0266-5611/25/i=12/a=123012.

Amaya, M., Morten, J. P., and Boman, L. A low-rank approximation to the Hessian for 3D CSEM Gauss-Newton inversion. In *76th EAGE Conference & Exhibition*, 2014. doi: 10.3997/2214-4609.20141099. URL http://dx.doi.org/10.3997/2214-4609.20141099.

Bansal, R., Krebs, J., Routh, P., Lee, S., Anderson, J., Baumstein, A., Mullur, A., Lazaratos, S., Chikichev, I., and McAdow, D. Simultaneous-source full-wavefield inversion. *The Leading Edge*, 32(9):1100–1108, 2013. doi: 10.1190/tle32091100.1. URL http://dx.doi.org/10.1190/tle32091100.1.

Grayver, A. V., Streich, R., and Ritter, O. Three-dimensional parallel distributed inversion of CSEM data using a direct forward solver. *Geophysical Journal International*, 193(3):1432–1446, 2013. doi: 10.1093/gji/ggt055. URL http://dx.doi.org/10.1093/gji/ggt055.

Lin, Y., Li, M., Abubakar, A., and Habashy, T. A wavelet-based model compression method for three-dimensional electromagnetic data inversion. In *SEG Technical Program Expanded Abstracts*, volume 138, pages 707–712, 2013. doi: 10.1190/segam2013-0395.1. URL http://dx.doi.org/10.1190/segam2013-0395.1.

Mackie, R., Watts, M., and Rodi, W. Joint 3D inversion of marine CSEM and MT data. In *SEG Expanded Abstracts*, volume 116, pages 574–578, 2007. doi: 10.1190/1.2792486. URL http://dx.doi.org/10.1190/1.2792486.

Mittet, R., Maulana, H., Brauti, K., and Wicklund, T. A. CMP inversion of marine CSEM data. In *EGM 2007 International Workshop*, April 2007. URL http://www.earthdoc.org/publication/publicationdetails/?publication=41234.

Sasaki, Y. Gauss-newton-based 3D joint inversion of marine CSEM and MT data. In *73rd EAGE Conference & Exhibition, C027*, 2011. doi: 10.3997/2214-4609. 20149043. URL http://dx.doi.org/10.3997/2214-4609.20149043.

Schwarzbach, C. and Haber, E. Finite element based inversion for electromagnetic problems using stochastic optimization. In *SEG Expanded Abstracts*, volume 110, pages 567–572, 2011. doi: 10.1190/1.3628145. URL http://dx.doi.org/10.1190/1. 3628145.

Støren, T., Zach, J., and Maaø, F. Gradient calculations for 3D inversion of CSEM data using a fast finite-difference time-domain modelling code. In *70th EAGE Conference & Exhibition, P194*, 2008. doi: 10.3997/2214-4609.20147963. URL http://dx.doi.org/10.3997/2214-4609.20147963.

Chapter 4

A low-rank approximation for large-scale 3D CSEM Gauss-Newton inversion

The low-rank approximation introduced in chapter 2 and chapter 3 could potentially reduce the computational cost of large-scale 3D CSEM data GN inversion. However there are still some practical aspects, as how to cope with a large number of inversion parameters, that need to be solved in order to apply this approximation to realistic size CSEM surveys. This is solved in this paper by introducing a matrix free recursive direct solver. A more detailed analysis of the approximation is also included.

Authors: Manuel Amaya, Jan Petter Morten, and Linus Boman.

Abstract: We consider an approximation to the Hessian for inversion of 3D controlled source electromagnetic data. The approach can considerably reduce the numerical complexity both 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 "super-shots" where 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 to use a large number of unknowns. The cross-talk introduced in the approximation is studied, and it is shown how the dissipative nature of marine electromagnetic field propagation reduces the impact of this noise. Inversion results from recent field data demonstrates the numerical and practical feasibility of the approach.

4.1   Introduction

The marine controlled-source electromagnetic (CSEM) method is an efficient tool for offshore hydrocarbon exploration with the potential to significantly increase 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 is today 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 the depth imaging and from quantitative resistivity information in the resulting subsurface models.

State-of-the-art 3D CSEM acquisition offers significant advantages by allowing to image the 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 both a hydrocarbon reservoir and other resistive structure that may be present, starting from a simple intial guess model. However, the Hessian-based inversion strategies that were originally devised and successfully applied for lower dimensional analysis employing a 2D assumption on model geometry (Abubakar et al., 2006; Mittet et al., 2007; Li and Key, 2007; Abubakar et al., 2009), will lead to very large computational complexity when scaled up for 3D appliccations.

One of the numerical complexities of inversion approaches that rely on second-derivative information like Gauss-Newton, Levenberg-Marquard, or Occam (deGroot Hedlin and Constable, 1990), arise 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 order $N \sim 10^6$, and the size of the Hessian will scale as N^2 making the matrix unpractical to handle even on large high-performance computer systems. Li et al. (2011) and Lin et al. (2013) introduced 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. In order to construct the Hessian matrix, it is necessary to compute the Green 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) studied 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.

In order 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 employed, 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 in order to achieve acceptable convergence rates and feasible models Loke and Dahlin (2002). Higher-order methods, like *e.g* Gauss-Newton, typically require less information in the initial models in order to achieve these two goals.

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 "super-shot" technique where the superposition of several Green functions is computed in single modeling jobs with simultaneously active sources. These constructs can be used to reduce the numerical load both from the simulation of Green 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 "super-shots" 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 cross-talk 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 utilized in seismic full waveform inversion, see *e.g.* Krebs et al. (2009); Boonyasiriwat and Schuster (2010); Ben-Hadj-Ali et al. (2011); Schiemenz and Igel (2013); Bansal et al. (2013). In this work we will consider a related phase encoding technique for electromagnetic data inversion. The approach based on "super-shots" can be particularly well suited for CSEM data due to the strong attentuation of signal amplitude from dissipative propagation, which limits cross-talk.

The proposed inversion scheme can be described as 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. Moreover, we do not make use of information from past iteration models to construct the Hessian approximation. If the model update is based on the models of past iterations, then the update may also be similar to previous updates. In comparision, when approximating the Hessian and computing model updates using only latest model data, we allow model changes in each iteration without dependence on past models. This allows a higher convergence rate. In the inversion results shown in this paper we will compare both 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 gradient computation, preserving the main structure of the Gauss-Newton scheme.

The Green 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 utilizing 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 to a quasi-Newton approach and to a Gauss-Newton scheme with model parameters compression. Finally we discuss the results and conclude.

4.2 Theory

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

$$\sigma(\mathbf{r}) = \arg \min_{\sigma \in \mathcal{M}} \varepsilon(\boldsymbol{\sigma}), \tag{4.1}$$

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

$$\varepsilon(\boldsymbol{\sigma}) = \varepsilon_{\mathrm{D}}(\boldsymbol{\sigma}) + \varepsilon_{\mathrm{R}}(\boldsymbol{\sigma}), \tag{4.2}$$

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,

$$\varepsilon_{\mathrm{D}}(\boldsymbol{\sigma}) = \sum_{F,i,f,\mathbf{r}_{\mathrm{rx}},\mathbf{r}_{\mathrm{tx}}} \left| W_i^F(\mathbf{r}_{\mathrm{rx}}|\mathbf{r}_{\mathrm{tx}}, f) \, \Delta F_i(\mathbf{r}_{\mathrm{rx}}|\mathbf{r}_{\mathrm{tx}}, f; \boldsymbol{\sigma}) \right|^2. \tag{4.3}$$

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

In this work, the non-linear 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 $\boldsymbol{\Delta\sigma}$ is obtained by solving the linear equation system

$$\mathbf{H}\,\boldsymbol{\Delta\sigma} = -\mathbf{g}, \tag{4.4}$$

where the Hessian is

$$\mathbf{H} = \mathbf{H}_{\mathrm{D}} + \mathbf{H}_{\mathrm{R}}, \tag{4.5}$$

and the gradient is

$$\mathbf{g} = \mathbf{g}_D + \mathbf{g}_R. \tag{4.6}$$

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

$$(\mathbf{H}_D)_{\mathbf{r},\mathbf{r}'} = \sum_{\kappa}(\mathbf{J})_{\mathbf{r},\kappa}(\mathbf{J})^*_{\mathbf{r}',\kappa} + \text{c.c.}, \tag{4.7}$$

$$(\mathbf{g}_D)_{\mathbf{r}} = \sum_{\kappa} W_\kappa \Delta F^*_\kappa (\mathbf{J})_{\mathbf{r},\kappa} + \text{c.c.} \tag{4.8}$$

The asterisk superscript denotes complex conjugation, and the notation c.c. denotes the complex conjugate of the preceeding 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 eq. 4.7 is approximated by neglecting a second-order derivative term. The Jacobian can be constructed from Green functions (Støren et al., 2008),

$$(\mathbf{J})_{\mathbf{r},\kappa} = W^F_i (\mathbf{r}_{\text{rx}}|\mathbf{r}_{\text{tx}}, f) \sum_{p} G^{F,J}_{p,i} (\mathbf{r}_{\text{rx}}|\mathbf{r}, f) \times \sum_{q} G^{E,J}_{p,q} (\mathbf{r}|\mathbf{r}_{\text{tx}}, f)\, \hat{j}_q (\mathbf{r}_{\text{tx}}, f), \tag{4.9}$$

where $\mathbf{r}$ is the position in the conductivity model, $G^{F,J}_{p,q}$ denotes the Green 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, $\hat{j}$ is a unit vector. It is straightforward to generalize the expression in eq. 4.9 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 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

$$N_{\text{Sim}} = \underbrace{N_F\, N_i\, N_{\text{rx}}}_{N_{\text{Sim}_{\text{Rx}}}} + \underbrace{N_{\text{tx}}}_{N_{\text{Sim}_{\text{Tx}}}}. \tag{4.10}$$

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 utilized computes the Green function at all survey frequencies from a single simulation. Therefore, N_{Sim} in eq. 4.10 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 semi-definite matrix. The number of independent data samples (number of Jacobian columns)

$$M = N_F \, N_i \, N_f \, N_{\text{rx}} \, N_{\text{tx}} \tag{4.11}$$

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

$$\text{rank}(\mathbf{H}_{\text{D}}) \leq \min(N, \, 2M). \tag{4.12}$$

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

The regularization $\varepsilon_{\text{R}}(\boldsymbol{\sigma})$ 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 eq. 4.4. In our implementation, the regularization can incorporate information on model conductivity ($\varepsilon_{\text{ap mod}}$), model smoothness ($\varepsilon_{\text{grad}}$) and conductivity anisotropy ($\varepsilon_{\text{ap aniso}}$),

$$\varepsilon_{\text{R}}(\boldsymbol{\sigma}) = \varepsilon_{\text{ap mod}}(\boldsymbol{\sigma}) + \varepsilon_{\text{grad}}(\boldsymbol{\sigma}) + \varepsilon_{\text{ap aniso}}(\boldsymbol{\sigma}). \tag{4.13}$$

In this scheme, the regularization $\varepsilon_{\text{R}}(\boldsymbol{\sigma})$ is normalized by the number of model parameters, and the contributions to the total misfit for the regularization terms in eq. 4.13 are balanced by the use of weights.

The regularization Hessian matrix derived from this scheme is

$$\mathbf{H}_{\text{R}} = \mathbf{H}_{\text{ap mod}} + \mathbf{H}_{\text{grad}} + \mathbf{H}_{\text{ap aniso}}, \tag{4.14}$$

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

4.2.1 Low-rank data Hessian matrix: super-shots 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 $\mathbf{H}_{\text{D}}$ 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 eq. 4.4, 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, $\tilde{\mathbf{H}}_{\text{D}}$, in a Gauss-Newton scheme, *i.e.*

$$\tilde{\mathbf{H}} \, \tilde{\Delta}\boldsymbol{\sigma} = -\mathbf{g}, \tag{4.15}$$

where the approximate Hessian is

$$\tilde{\mathbf{H}} = \tilde{\mathbf{H}}_{\text{D}} + \mathbf{H}_{\text{R}}. \tag{4.16}$$

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

To compute the matrix $\tilde{\mathbf{H}}_{\mathrm{D}}$ we consider a low-rank approximation where sources in eq. 4.9 are combined in superpositions, after being encoded with a random phase factor. The result is a contraction of Jacobian columns that can be written

$$(\tilde{\mathbf{J}})_{\mathbf{r},\tilde{\kappa}} = \sum_{\mathbf{r}_{\mathrm{tx}} \in s} e^{i\phi_\kappa} (\mathbf{J})_{\mathbf{r},\kappa}, \qquad (4.17)$$

for a group of source positions s, and where ϕ_κ are uniformly distributed random numbers in the interval $[0, 2\pi)$. The shorthand notation $\tilde{\kappa} = (F, i, f, \mathbf{r}_{\mathbf{rx}}, s)$ 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 eq. 4.17 associated with the source positions in a source group and a specific receiver field-component are calculated from a single simultaneous-source (super-shot) forward solution of the Maxwell equations. We denote the output of this simulation as

$$\tilde{G}_{p,i,\mathbf{r}_{\mathrm{rx}},s}^F (\mathbf{r}, f) = \sum_{q,\,\mathbf{r}_{\mathrm{tx}} \in s} G_{p,q}^{E,J} (\mathbf{r}|\mathbf{r}_{\mathrm{tx}}, f)\, W_i^F (\mathbf{r}_{\mathrm{rx}}|\mathbf{r}_{\mathrm{tx}}, f)\, e^{i\phi_\kappa} \hat{\jmath}_q (\mathbf{r}_{\mathrm{tx}}, f), \quad (4.18)$$

where $W_i^F (\mathbf{r}_{\mathrm{rx}}|\mathbf{r}_{\mathrm{tx}}, f)\, e^{i\phi_\kappa} \hat{\jmath}_q (\mathbf{r}_{\mathrm{tx}}, f)$ 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

$$N_{\mathrm{Sim}_s} = N_F\, N_i\, N_{\mathrm{rx}}\, N_s. \qquad (4.19)$$

With respect to the total number of simulations, the quantity N_{Sim_s} above will replace $N_{\mathrm{Sim}_{\mathrm{Tx}}}$ in eq. 4.10 when super-shots are applied. The super-shot simulations in eq. 4.18 define an approximate Jacobian $\tilde{\mathbf{J}}$ as

$$(\tilde{\mathbf{J}})_{\mathbf{r},\tilde{\kappa}} = \sum_p G_{p,i}^{F,J} (\mathbf{r}_{\mathrm{rx}}|\mathbf{r}, f)\, \tilde{G}_{p,i,\mathbf{r}_{\mathrm{rx}},s}^F (\mathbf{r}, f), \qquad (4.20)$$

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

$$(\tilde{\mathbf{H}}_{\mathrm{D}})_{\mathbf{r},\mathbf{r}'} = \sum_{\tilde{\kappa}} (\tilde{\mathbf{J}})_{\mathbf{r},\tilde{\kappa}} (\tilde{\mathbf{J}})_{\mathbf{r}',\tilde{\kappa}}^* + \mathrm{c.c.}, \qquad (4.21)$$

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

$$\tilde{M} = N_F\, N_i\, N_f\, N_{\mathrm{rx}}\, N_{\mathrm{s}}. \qquad (4.22)$$

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

$$\mathrm{rank}(\tilde{\mathbf{H}}_{\mathrm{D}}) \le 2\, \tilde{M}. \qquad (4.23)$$

Comparing eq. 4.23 with eq. 4.12 we see that this approach results in a low-rank data Hessian matrix approximation $\tilde{\mathbf{H}}_{\mathrm{D}}$. Note that both the number of forward solutions

(compare eq. 4.10 and eq. 4.19) and the storage requirement to construct $\tilde{\mathbf{H}}_D$ (compare eq. 4.11 and eq. 4.22) scale by N_s instead of N_{tx} as for $\mathbf{H}_D$. 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 to 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 $\mathbf{g}_D$ 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 eq. 4.8. Following this scheme we compute

$$(\mathbf{g}_D)_{\mathbf{r}} = \sum_{\upsilon=F,i,f,\mathbf{r}_{rx},p} \text{sign}(F) G_{p,i}^{F,J}(\mathbf{r}_{rx}|\mathbf{r},f)\, L_\upsilon(\mathbf{r}), \tag{4.24}$$

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

$$L_\upsilon(\mathbf{r}) = \sum_{q,\mathbf{r}_{tx}} G_{p,q}^{E,J}(\mathbf{r}|\mathbf{r}_{tx},f) \left[W_i^F(\mathbf{r}_{rx}|\mathbf{r}_{tx},f)\right]^2 \Delta F\, \hat{\jmath}_q(\mathbf{r}_{tx},f). \tag{4.25}$$

The Green function $G_{p,i}^{F,J}$ in eq. 4.24 is the same as the one used for the approximated Jacobian $\tilde{\mathbf{J}}$ in eq. 4.20. To compute $L_\upsilon(\mathbf{r})$ it is necessary to run additional simulations

$$N_{Sim_{Adj}} = N_F\, N_i\, N_{rx}, \tag{4.26}$$

since the source strength factors are different in the superpositions in eq. 4.18 and eq. 4.25.

The total number of forward solutions needed to compute the approximated data Hessian matrix in eq. 4.21 and the data gradient in eq. 4.24 in the low-rank approach is

$$\tilde{N}_{Sim} = N_{Sim_{Rx}} + N_{Sim_s} + N_{Sim_{Adj}} = N_F \cdot N_i \cdot N_{rx} \cdot (N_s + 2). \tag{4.27}$$

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 to Gauss-Newton inversion whenever $N_F\, N_i\, N_{rx}\,(N_s+2)/N_{tx} < 1$, where both factors N_F and N_i are of order unity. Note that the number N_{Sim_s} from eq. 4.19 corresponds to the number of additional forward solutions required by the low-rank approach compared to gradient-based approaches like quasi-Newton and non-linear conjugate gradients.

Source grouping strategies

We will now discuss the selection of source points for the groups introduced in eq. 4.17. We will consider the three different strategies shown schematically in figure

(a) Four groups of sources gathering closest ones in a single group simulation

(b) Four groups of sources gathering randomly the different sources in a single group simulation

(c) Four groups of sources gathering distant sources in a single group simulation
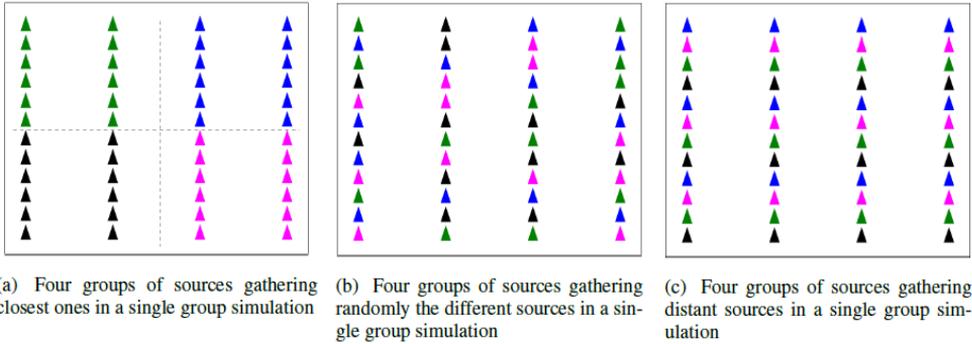
Figure 4.1: Examples of different strategies for source grouping. The triangles represent source positions, and the points with the same color are grouped.

4.1, where all source positions are arranged in four groups. In figure 4.1(a) the sources in each group are selected such that each group covers a contiguous area. In figure 4.1(b) the sources are grouped randomly, and in figure 4.1(c) 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 4.1(c), 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 eq. 4.18. Note that the specific grouping illustrated in the example figure 4.1(c) is only approximately optimal, as further increase in separation could be possible.
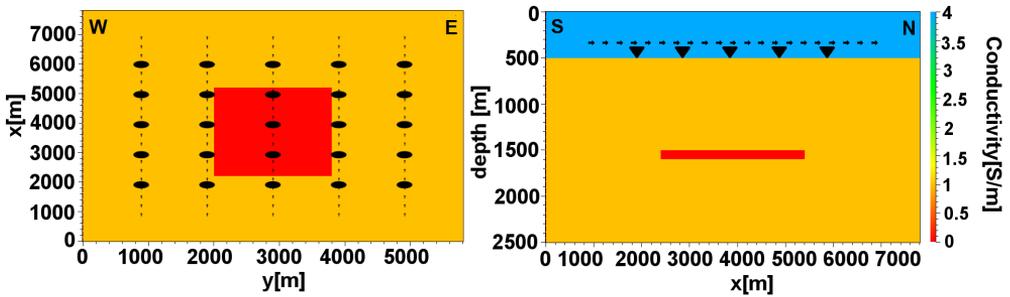
Figure 4.2: Survey layout for synthetic study.

In order to illustrate some properties of different grouping strategies we will consider synthetic data from the model and source-receiver layout shown in figure 4.2. The example survey includes 25 receivers, 5 towlines (sampling 330 source positions in total), and a thin resistor at 50 Ωm. The background resistivity is 1 Ωm and the water resistivity is 0.25 Ωm. The input synthetic data for the study were the E_x and E_y fields at frequencies 0.25 and 1.0 Hz, without noise.

In figure 4.3 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, $\text{rank}(\tilde{\mathbf{H}}_D) = 600$. The number of source groups was $N_s = 3$.
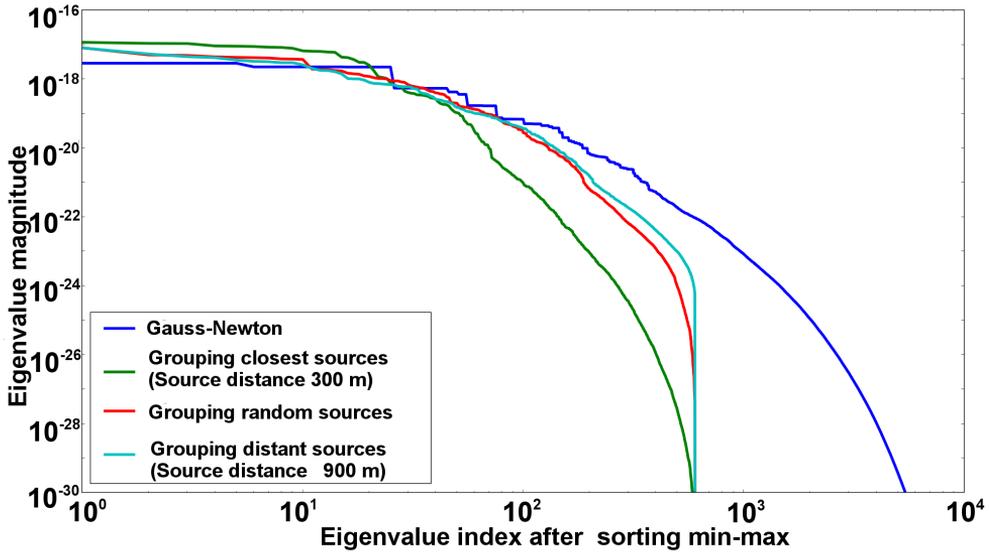
Figure 4.3: Example eigenvalue distribution for different source-grouping strategies. We use the same number of simulations in the three different grouping strategies. The eigenvalues are shown in sorted order from the largest to the smallest along the horizontal axis.

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 cross-talk noise (discussed below). Moreover, the rank of the approximate data Hessian $\tilde{\mathbf{H}}_D$ is proportional to the number of groups. In figure 4.4 we show an example, based on the survey shown in figure 4.2, of how increasing the number of groups increases the rank of the low-rank data Hessian in eq. 4.21. We use the grouping strategy based on distant sources described above. From figure 4.4 we see how increasing N_s leads to an increase in the number of non-zero eigenvalues, and an eigenvalue distribution closer to that of the Gauss-Newton data Hessian $\mathbf{H}_D$.
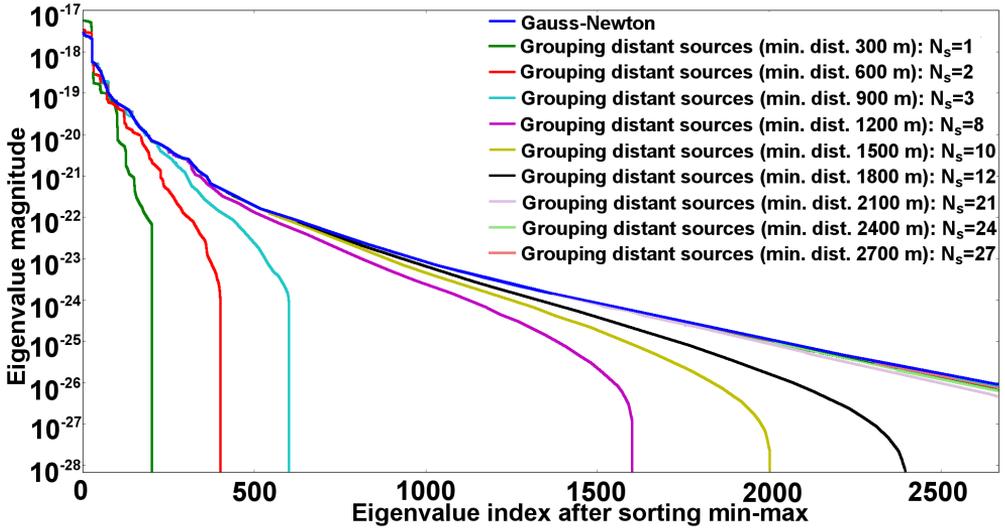
Figure 4.4: Example eigenvalue distribution for different number of distant-source groups N_s.

In a typical application of the 3D CSEM Gauss-Newton inversion scheme, the regularization terms in the cost function at eq. 4.13 will stabilize the solution of the linear system in eq. 4.4, and limit the condition number of the Hessian matrix **H**. Therefore, though the matrix $\tilde{\mathbf{H}}_D$ is typically rank deficient as shown in figure 4.3 and figure 4.4, the eigenvalue distribution of the matrix in eq. 4.4 that determines the update for the low-rank approximation can be close to that of the original Gauss-Newton scheme. Figure 4.5 demonstrates this for the survey shown in figure 4.2. Note that the magnitude scale of the eigenvalues in figure 4.5 is different than in figure 4.3 and figure 4.4 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 $N_F\,N_i\,N_{\mathrm{rx}}\,N_s$ forward solutions), using $W_i^F\left(\mathbf{r}_{\mathrm{rx}}|\mathbf{r}_{\mathrm{tx}}, f\right)\mathrm{e}^{i\phi_\kappa}$ 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 eq. 4.18. The random phase factors $\mathrm{e}^{i\phi_\kappa}$ are sampled independently for ϕ_κ in each group $\kappa_s = (F, i, f, \mathbf{r}_{\mathrm{rx}}, \mathbf{r}_{\mathrm{tx}} \in s)$.
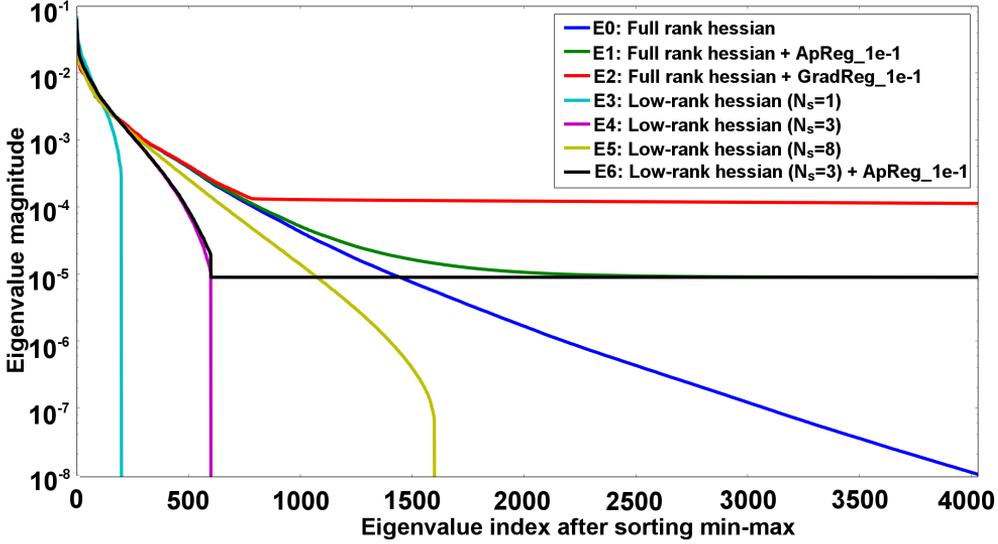
Figure 4.5: Hessian matrix **H** eigenvalue distributions. Blue curve E0: Gauss-Newton data Hessian $\mathbf{H}_D$ only. Green curve E1: Gauss-Newton data Hessian and regularization $\mathbf{H}_D + \mathbf{H}_{\text{ap mod}}$. Red curve E2: Gauss-Newton data Hessian and regularization $\mathbf{H}_D + \mathbf{H}_{\text{grad}}$. Cyan E3, purple E4, and yellow E5 curves: Low-rank data Hessian $\tilde{\mathbf{H}}_D$ only for 1,3, and 8 source groups. Black curve E6: Low-rank data Hessian and regularization $\tilde{\mathbf{H}}_D + \mathbf{H}_{\text{ap mod}}$.

Noise analysis and phase encoding

Let us now consider the errors introduced by the approximation of the Hessian in eq. 4.21. When the data Hessian is constructed according to eq. 4.7, each term contributing is computed from the Green functions associated with one receiver and one source position. The same terms appear when computing the low-rank approximation (eq. 4.21), but in addition the approximation introduces terms (cross-talk terms) which are not part of the Gauss-Newton data Hessian. This is illustrated by the diagrams in figure 4.6 where (a) shows a low-rank approximation contribution from one source position, which is the same that also appears in a standard Gauss-Newton data Hessian computation, and (b) shows one of the additional terms introduced through the low-rank approximation.

The terms of figure 4.6 (a) involve only one source and one receiver positions 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 eq. 4.21. These terms are shown as figure 4.6 (b). We will refer to these diagrams as "cross-talk" and denote their contribution $\boldsymbol{\eta}$, such that

$$\tilde{\mathbf{H}}_D = \mathbf{H}_D + \boldsymbol{\eta}. \tag{4.28}$$
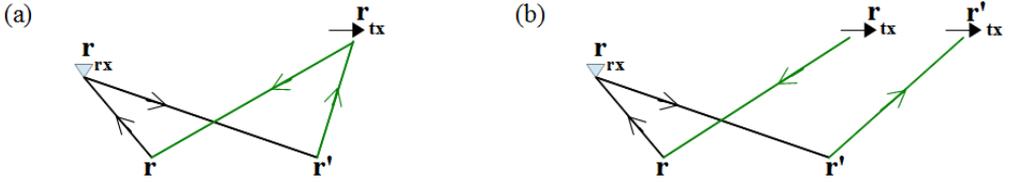
Figure 4.6: Diagrams representing the two types of terms included in the data Hessian matrix of a low-rank approximation (eq. 4.21). Each source position (subscript "tx") is associated with its corresponding source strength factors (see discussion following eq. 4.18). In these diagrams, an arrow from *e.g.* position $\mathbf{r}_{\text{tx}}$ to $\mathbf{r}$ represents a Green function $G_{p,q}^{F,J}(\mathbf{r}|\mathbf{r}_{\text{tx}}, f)$ 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 conjuation in the Hessian expressions. (a) Diagram representing those terms of $\tilde{\mathbf{H}}_{\text{D}}$ that also make up the data Gauss-Newton Hessian (eq. 4.7). (b) Diagram representing the additional cross-talk noise terms introduced into $\tilde{\mathbf{H}}_{\text{D}}$, where two different source positions contribute.

Though the cross-talk terms will represent an approximation error, we note that $\boldsymbol{\eta}$ has specific matrix properties. The matrix $\boldsymbol{\eta}$ will be symmetric, and from the definition we have that

$$\text{rank}\left(\tilde{\mathbf{H}}_{\text{D}} - \boldsymbol{\eta}\right) = \text{rank}\left(\mathbf{H}_{\text{D}}\right). \tag{4.29}$$

These properties could be utilized in schemes to reduce the effect of $\boldsymbol{\eta}$ in the approximation. In figure 4.7 we show the diagonal (*i.e.* matrix elements for indices $\sigma_{\text{V}}(\mathbf{r}), \sigma_{\text{V}}(\mathbf{r})$ and $\sigma_{\text{H}}(\mathbf{r}), \sigma_{\text{H}}(\mathbf{r})$ where subscripts V,H pertain to vertical and horizontal component) of the matrices $\mathbf{H}_{\text{D}}$ and $\boldsymbol{\eta}$ for two different number of groups. These results were computed using a half-space model. Comparing figure 4.7 (b) and figure 4.7 (c) we note that the cross-talk (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 cross-talk. We study these two phenomena later in this section. The large-scale structure of the cross-talk seen in figure 4.7 could be detrimental to the inversion convergence but is hard to suppress. The synthetic data inversion results in Appendix (section 4.7) show that an inversion result of similar quality as in the Gauss-Newton approach can be achieved despite the cross-talk noise. As an additional obseration, we note that the diagonal cross-talk noise was reduced more efficiently by increasing the number of source groups in the horizontal rather than the vertical components as seen in figure 4.7 (b) and figure 4.7 (c).

The phase encoding in the super-shots in eq. 4.18 is a way to reduce the errors due to cross-talk. Noise reduction approaches based on random phase encoding have also been applied in seismic modeling, see *e.g.* Bansal et al. (2013). In eq. 4.21, 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 4.6 (a). This is because

the source Green functions appearing in $\tilde{\mathbf{H}}_\mathrm{D}$ are included in combinations

$$G_{p,q}^{E,J}\left(\mathbf{r}|\mathbf{r}_\mathrm{tx},f\right)\mathrm{e}^{i\phi_{E,i,\mathbf{r}_\mathrm{tx}}} \times \left[G_{p,q}^{E,J}\left(\mathbf{r}'|\mathbf{r}_\mathrm{tx},f\right)\mathrm{e}^{i\phi_{E,i,\mathbf{r}_\mathrm{tx}}}\right]^*. \tag{4.30}$$

For these terms the random phase cancels, and phase encoding does not affect the diagrams corresponding to the Gauss-Newton data Hessian.
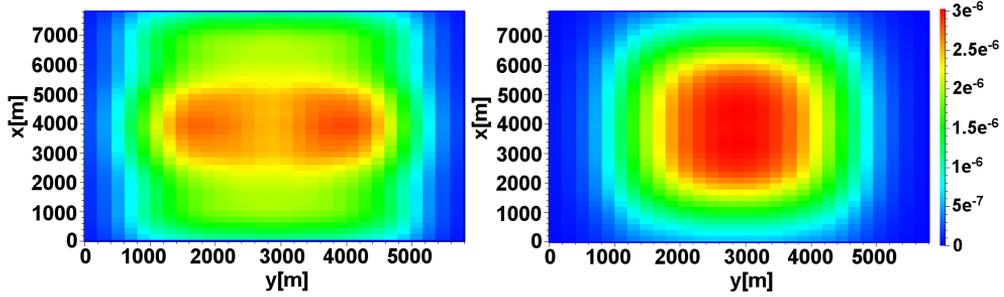
(a) Magnitude of the elements on the diagonal of $\mathbf{H}_\mathrm{D}$
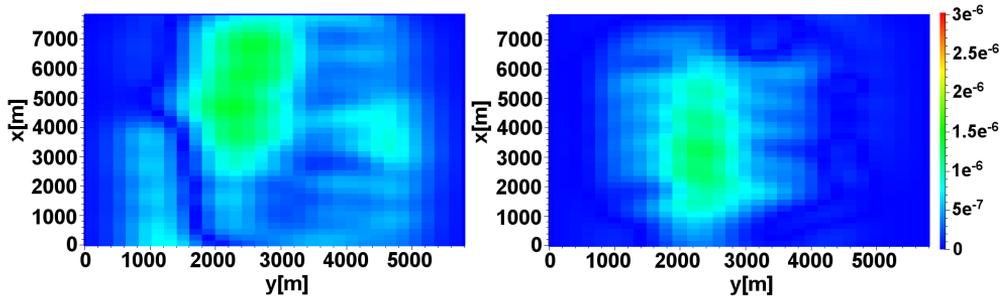
(b) Magnitude of the elements on the diagonal of $\boldsymbol{\eta}$, $N_s = 1$
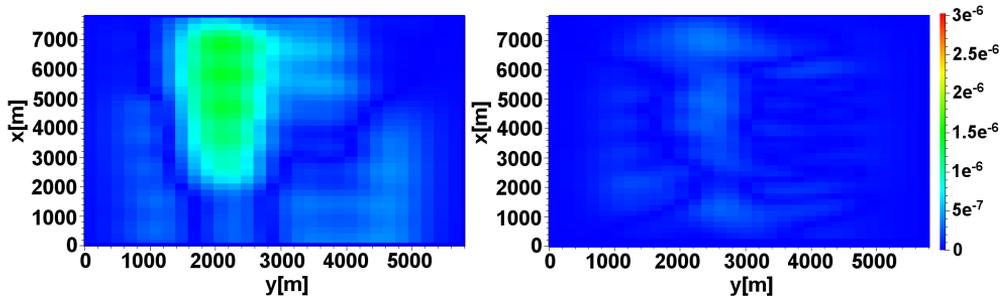
(c) Magnitude of the elements on the diagonal of $\boldsymbol{\eta}$, $N_s = 3$

Figure 4.7: Plots of the Hessian and cross-talk matrix diagonals at depth 1500 m for the survey shown in figure 4.2. Vertical conductivity component on the left and horizontal conductivity component on the right. These results were computed in a half-space model.

The cross-talk terms result in contributions where the source Green functions for

two different positions $\mathbf{r}_{\mathtt{tx}}$ and $\mathbf{r}'_{\mathtt{tx}}$ appear in combinations of the type

$$G_{p,q}^{E,J}\left(\mathbf{r}|\mathbf{r}_{\mathtt{tx}},f\right)\left[G_{p,q}^{E,J}\left(\mathbf{r}'|\mathbf{r}'_{\mathtt{tx}},f\right)\right]^{*}\times \mathrm{e}^{i\phi_{E,i,\mathbf{r}_{\mathtt{tx}}}-i\phi_{E,i,\mathbf{r}'_{\mathtt{tx}}}}. \tag{4.31}$$

The random phase factor in this expression will act to reduce the total contribution of the cross-talk terms. To analyze this, let us first assume that the magnitudes of the cross-talk terms in $\boldsymbol{\eta}$ are Gaussian distributed. In such case, the resulting magnitude of the sum of the cross-talk terms will be governed by the properties of a Gaussian random walk. If we assume that there are A terms contributing, then the magnitude, $|\boldsymbol{\eta}|$, will scale as $\sim \sqrt{A}$ instead of linear scaling due to the random phases in eq. 4.31.

The number of terms (figure 4.6 (a)) contributing to $\mathbf{H}_{\mathrm{D}}$ is equal to the different number of combinations (eq. 4.11) of the eq. 4.30 that appear in eq. 4.21, i.e. it is proportional to $N_{\mathtt{tx}}$. The number of cross-talk terms (figure 4.6 (b)) contributing to $\boldsymbol{\eta}$ scales with the number of all possible combinations of two different sources (eq. 4.31) included in a group simulation that appear in eq. 4.21, i.e. scales as $N_{\mathtt{tx}}^2$ (assuming a single source group $N_s = 1$). However, the magnitude $|\boldsymbol{\eta}|$ should still scale by $N_{\mathtt{tx}}$ by analogy to a Gaussian random walk as described above, *i.e.* $|\boldsymbol{\eta}| \sim N_{\mathtt{tx}}$.

Let us now consider the details of the distribution of cross-talk magnitudes. The physics of the problem indicates that the magnitude distribution of the cross-talk terms should be more centered than a Gaussian distribution. The exponential decay of the magnitude of the Green functions will result in negligible contributions from the product of Green functions in eq. 4.31 when the distance $|\mathbf{r}_{\mathtt{tx}} - \mathbf{r}'_{\mathtt{tx}}|$ is large. This will be the case for many of the terms making up $\boldsymbol{\eta}$. The spatial decay of the Green functions is shown in figure 4.8. Thus, the attenuation of the magnitude for the contributions makes $|\boldsymbol{\eta}|$ scale by the number of sources in a source group as $(N_{\mathtt{tx}})^\alpha$ with $\alpha < 1$, and the asymptotic behaviour of the approximation is given as $\lim_{N_{\mathtt{tx}} \to \infty} |\boldsymbol{\eta}|/|\mathbf{H}_{\mathrm{D}}| = 0$.

Next we turn our attention to the magnitude of the individual terms contributing to $\mathbf{H}_{\mathrm{D}}$ relative to the cross-talk terms in $\boldsymbol{\eta}$. First we note that for each cross-talk contribution there is a contribution to the data Hessian with larger magnitude. The elements of the Jacobian are Frechet derivatives computed from a product of a Green function associated with a receiver position and a source position, see eq. 4.20. Magnitudes of Frechet derivatives (formed by the Green functions shown in figure 4.8) are shown in figure 4.9. Contributions to $\tilde{\mathbf{H}}_{\mathrm{D}}$ are products of two such Frechet derivatives. A cross-talk 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 4.9. The terms contributing to $\mathbf{H}_{\mathrm{D}}$ are products of a Frechet derivative by its complex conjugate, *e.g.* the product of Rx1-Tx3 by the complex conjugate. Using as a reference the magnitudes represented in figure 4.9, we see that this $\mathbf{H}_{\mathrm{D}}$ contribution will dominate over the cross-talk contribution from Rx1-Tx5 by Rx1-Tx3. In general, there is always a term contributing to $\mathbf{H}_{\mathrm{D}}$ larger than terms part of $\boldsymbol{\eta}$ involving the same receiver and one of the source positions.
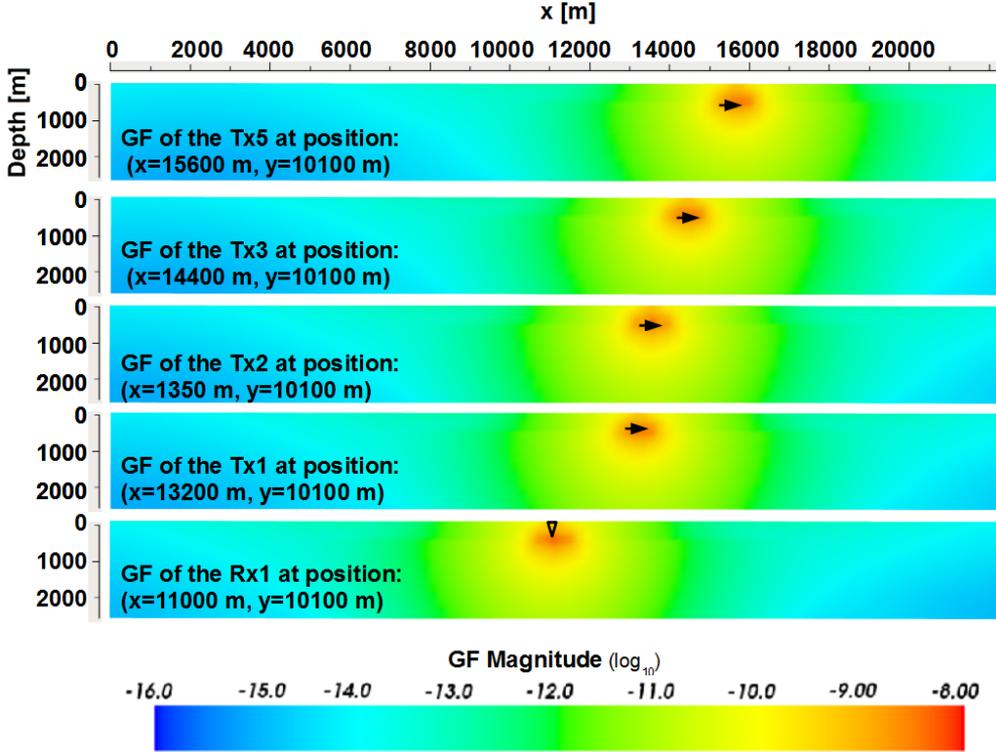
Figure 4.8: Green function amplitude at different source (Tx) and receiver (Rx) positions, for the electric field E_x at frequency $f = 0.25$ Hz, when using the survey layout represented at figure 4.2 for a larger model with the same background and water conductivity but without the 50 Ωm target.

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 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 $\tilde{\mathbf{\Delta}\boldsymbol{\sigma}}$. 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.
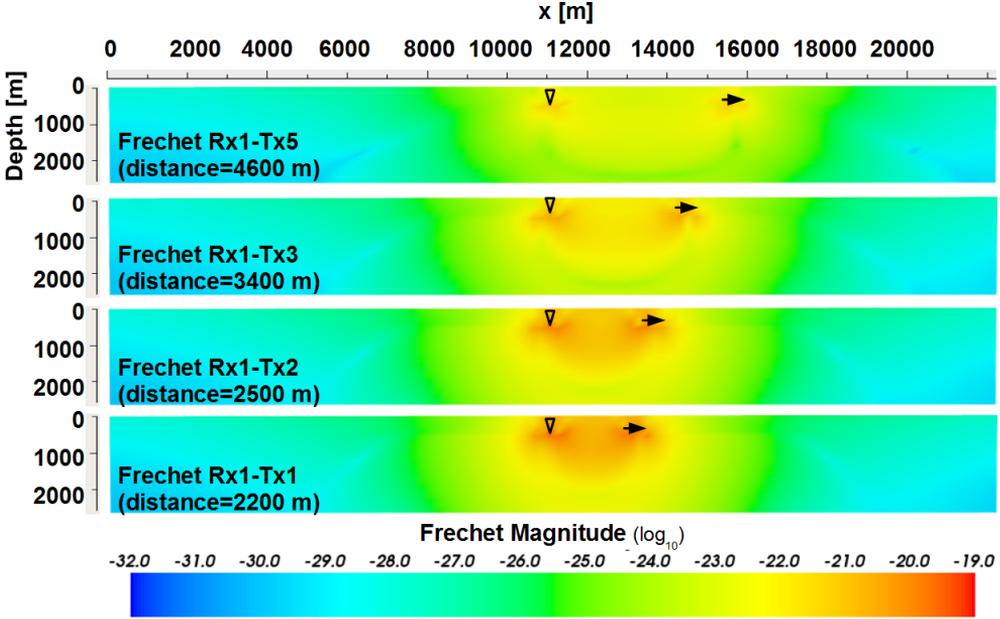
Figure 4.9: Frechet derivatives computed from the Green functions in figure 4.8.

4.2.2 Data Hessian matrix free recursive direct solver

As we mention in the introduction, 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 eq. 4.15 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 case, a parameter compression strategy that can reduce the order of magnitude of the parameter number by 10-100 is required in order to solve numerical linear algebra problem. For the low-rank approximation, we can however make use of the implicit matrix representation in eq. 4.21 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 eq. 4.15, 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,

$$\tilde{\mathbf{H}}^{-1} = \left(\mathbf{H_R} + \sum_{m=1}^{\tilde{M}} \left(\tilde{\mathbf{J}}_m \tilde{\mathbf{J}}_m^\dagger + \text{c.c.} \right) \right)^{-1}, \tag{4.32}$$

where $\tilde{M}$ is the number of approximated Jacobian columns as in eq. 4.22, and $\tilde{\mathbf{J}}_m$ represents the m-th column. The superscript † symbol represents the conjugate transpose. Introducing the short-hand notation

$$\tilde{\mathbf{J}}_{m_{\mathrm{Re}}} = \mathrm{Re}(\sqrt{2}\,\tilde{\mathbf{J}}_m), \quad \tilde{\mathbf{J}}_{m_{\mathrm{Im}}} = \mathrm{Im}(\sqrt{2}\,\tilde{\mathbf{J}}_m), \tag{4.33}$$

we can rewrite the eq. 4.32 in the following form,

$$\tilde{\mathbf{H}}^{-1} = \left(\mathbf{H}_{\mathrm{R}} + \sum_{j=1}^{2\tilde{M}} \left(\tilde{\mathbf{J}}_j \tilde{\mathbf{J}}_j^T \right) \right)^{-1}, \tag{4.34}$$

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

$$\left(\mathbf{A} + \mathbf{u}\mathbf{u}^T \right)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{u}^T\mathbf{A}^{-1}}{1 + \mathbf{u}^T\mathbf{A}^{-1}\mathbf{u}}. \tag{4.35}$$

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 in order to avoid storing the whole $\tilde{\mathbf{J}}$ matrix. At iteration j it is only necessary to keep in memory the sparse $\mathbf{H}_{\mathrm{R}}$, the so-called iteration vectors $\mathbf{w}_j$ and the scalars c_j. The algorithm is described below in three steps.

Initialization,

$$\tilde{\mathbf{H}}_{(0)}^{-1} = \mathbf{H}_{\mathrm{R}}^{-1}. \tag{4.36}$$

First iteration ($j = 1$),

$$\begin{aligned} \tilde{\mathbf{H}}_{(1)}^{-1} &= \left(\tilde{\mathbf{H}}_{(0)} + \tilde{\mathbf{J}}_1 \tilde{\mathbf{J}}_1^T \right)^{-1} \\ &= \tilde{\mathbf{H}}_{\mathrm{R}}^{-1} - \frac{\tilde{\mathbf{H}}_{\mathrm{R}}^{-1} \tilde{\mathbf{J}}_1 \tilde{\mathbf{J}}_1^T \tilde{\mathbf{H}}_{\mathrm{R}}^{-1}}{1 + \tilde{\mathbf{J}}_1^T \tilde{\mathbf{H}}_{\mathrm{R}}^{-1} \tilde{\mathbf{J}}_1} \\ &= \tilde{\mathbf{H}}_{\mathrm{R}}^{-1} - c_1 \mathbf{w}_1 \mathbf{w}_1^T, \end{aligned} \tag{4.37}$$

$$\mathbf{w}_1 = \tilde{\mathbf{H}}_{\mathrm{R}}^{-1} \tilde{\mathbf{J}}_1. \tag{4.38}$$

$$c_1 = \left(1 + \tilde{\mathbf{J}}_1^T \mathbf{w}_1 \right)^{-1}. \tag{4.39}$$

The Sherman-Morrison formula, eq. 4.35, is applied in the second step of eq. 4.37. Second iteration ($j = 2$),

$$
\begin{aligned}
\tilde{\mathbf{H}}_{(2)}^{-1} &= \left(\tilde{\mathbf{H}}_{(1)} + \tilde{\mathbf{J}}_2 \tilde{\mathbf{J}}_2^T \right)^{-1} \\
&= \tilde{\mathbf{H}}_{(1)}^{-1} - \frac{\tilde{\mathbf{H}}_{(1)}^{-1} \tilde{\mathbf{J}}_2 \tilde{\mathbf{J}}_2^T \tilde{\mathbf{H}}_{(1)}^{-1}}{1 + \tilde{\mathbf{J}}_2^T \tilde{\mathbf{H}}_{(1)}^{-1} \tilde{\mathbf{J}}_2} \\
&= \tilde{\mathbf{H}}_{(1)}^{-1} - c_2 \mathbf{w}_2 \mathbf{w}_2^T \\
&= \tilde{\mathbf{H}}_{\mathrm{R}}^{-1} - c_1 \mathbf{w}_1 \mathbf{w}_1^T - c_2 \mathbf{w}_2 \mathbf{w}_2^T ,
\end{aligned}
\tag{4.40}
$$

$$
\begin{aligned}
\mathbf{w}_2 &= \tilde{\mathbf{H}}_{(1)}^{-1} \tilde{\mathbf{J}}_2 \\
&= \tilde{\mathbf{H}}_{\mathrm{R}}^{-1} \tilde{\mathbf{J}}_2 - c_1 \mathbf{w}_1 \mathbf{w}_1^T \tilde{\mathbf{J}}_2 ,
\end{aligned}
\tag{4.41}
$$

$$
c_2 = \left(1 + \tilde{\mathbf{J}}_2^T \mathbf{w}_2 \right)^{-1} .
\tag{4.42}
$$

When generalizing for iteration j,

$$
\begin{aligned}
\tilde{\mathbf{H}}_{(j)}^{-1} &= \tilde{\mathbf{H}}_{(j-1)}^{-1} - c_j \mathbf{w}_j \mathbf{w}_j^T \\
&= \tilde{\mathbf{H}}_{(0)}^{-1} - \sum_{q=1}^{j} c_q \mathbf{w}_q \mathbf{w}_q^T ,
\end{aligned}
\tag{4.43}
$$

$$
\begin{aligned}
\mathbf{w}_j &= \tilde{\mathbf{H}}_{(j-1)}^{-1} \tilde{\mathbf{J}}_j \\
&= \tilde{\mathbf{H}}_{(0)}^{-1} \tilde{\mathbf{J}}_j - \sum_{q=1}^{j-1} c_q \mathbf{w}_q \mathbf{w}_q^T \tilde{\mathbf{J}}_q ,
\end{aligned}
\tag{4.44}
$$

$$
c_j = \left(1 + \tilde{\mathbf{J}}_j^T \mathbf{w}_j \right)^{-1} .
\tag{4.45}
$$

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

$$
\tilde{\mathbf{H}}^{-1} = \tilde{\mathbf{H}}_{(2\tilde{M})}^{-1} ,
\tag{4.46}
$$

and then the approximated model update is obtained as

$$
\mathbf{\Delta}\tilde{\sigma} = - \left(\tilde{\mathbf{H}}_{(0)}^{-1} \mathbf{g} - \sum_{j=1}^{2\tilde{M}} c_j \mathbf{w}_j \mathbf{w}_j^T \mathbf{g} \right) .
\tag{4.47}
$$

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 eq. 4.44 and eq. 4.47 can be obtained without the explicit inverse. Since the regularization will be a sparse matrix 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 eq. 4.15 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 to approaches where the Hessian is constructed explicitly, it becomes feasible to solve linear systems in eq. 4.15 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,

$$\tilde{\mathbf{H}}_R = \text{diag}(\mathbf{H}_R), \tag{4.48}$$

and use the Hessian

$$\hat{\mathbf{H}} = \tilde{\mathbf{H}}_D + \tilde{\mathbf{H}}_R. \tag{4.49}$$

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 eq. 4.36.

The a priori model regularization matrix $\mathbf{H}_{\text{ap mod}}$ is not modified when it is used in the approximated regularization Hessian matrix (eq. 4.48) because $\mathbf{H}_{\text{ap mod}}$ is diagonal by definition. For the gradient smoothness $\mathbf{H}_{\text{grad}}$ the approximation will affect the resulting update, but since we are carrying out iterations of the non-linear optimization, the lack of information can be compensated at later iterations.

4.3   Inversion results

In this section we present and compare the inversion results from a recent 3D CSEM field dataset. The geological complexity and the scale of the acquisition of the field data makes this a challenging and interesting survey to use to test the performance of inverse schemes. We consider four different schemes: (A) quasi-Newton L-BFGS-B (Zhu et al., 1997), (B) 3D Gauss-Newton using model parameter compression and a conjugate gradient (CG) solver for eq. 4.4, (C) the proposed low-rank approach using model parameter compression and a CG solver for eq. 4.15, and (D) the low-rank approach using the presented recursive direct solver with the approximation in eq. 4.48. The approaches (A) and (B) are in the class of optimization methods most commonly applied for inversion of marine CSEM data today. For all the tests we use L2-norm model smoothness regularization. In Appendix (section 4.7) we include inversion results from the synthetic survey in figure 4.2.

The survey data considered here were acquired by EMGS ASA in 2014. Due to showright restrictions, we can not 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 is shown in figure 4.10. The water depth in this area is about 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 $0.55 \cdot 10^6$ in these examples.
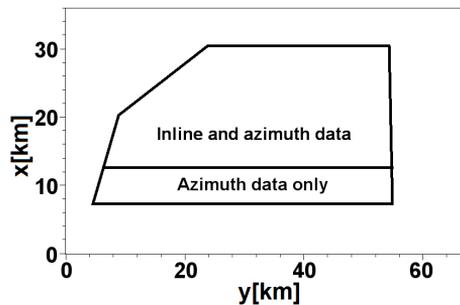
Figure 4.10: Receiver coverage for the subset of the survey data considered in the inversion examples. Receivers were positioned in a 2 km×2 km grid inside the area shown by the polygons. The source towing was focused on the top region, where receivers recorded both inline and azimuth data. The receivers in the lower region only recorded azimuth data.

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 typ-

ically some structural information like *e.g.* salt outlines from seismic is utilized to obtain the best inversion results. However, in this case we instead use a simple initial model in order to highlight the performance of different inversion schemes. We consider anisotropic inversion with transverse isotropy defined by a vertical axis, *i.e.* a VTI model where 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 in order to have a fair comparison between methods. We define the RMS value as $\sqrt{\varepsilon/N}$ where ε is defined in eq. 4.2 and N is the number of observations. Although results with misfit this large are not suited for geological 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 (B) and (C) defined above 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 in the order of a factor 200 compared to 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 (A) are shown in figure 4.11. To achieve this result we carried out 60 iterations of the inversion, where each iteration required 1004 forward solutions (60240 solutions in total). The memory usage to compute the update was about 10 GB for 10^7 free parameters.
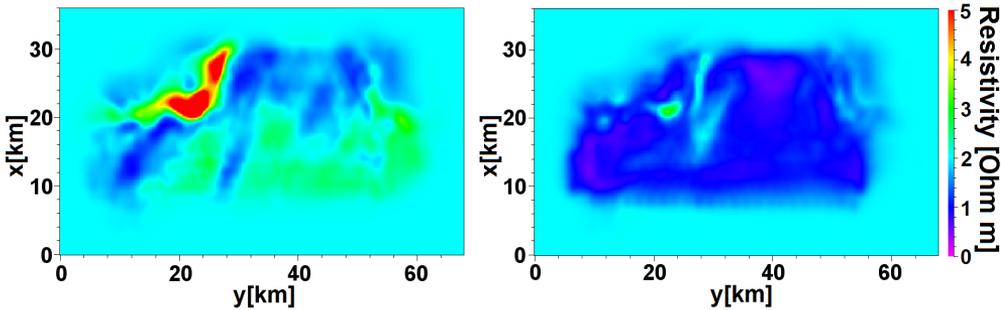
Figure 4.11: Survey scheme (A) inversion result: L-BFGS-B optimizer result; depth-slices at 4200 m for the vertical resistivity model (left) and the horizontal resistivity model (right); the total number of free parameters was 10^7.

In figure 4.12 we show results from the Gauss-Newton scheme (B), where we utilized the model parameter compression and used a CG method to solve the eq. 4.4. This result was obtained after 13 iterations, using 2806 forward solutions per iteration (36478 simulations in total) and a memory usage of 12 GB for $5 \cdot 10^4$ parameters.
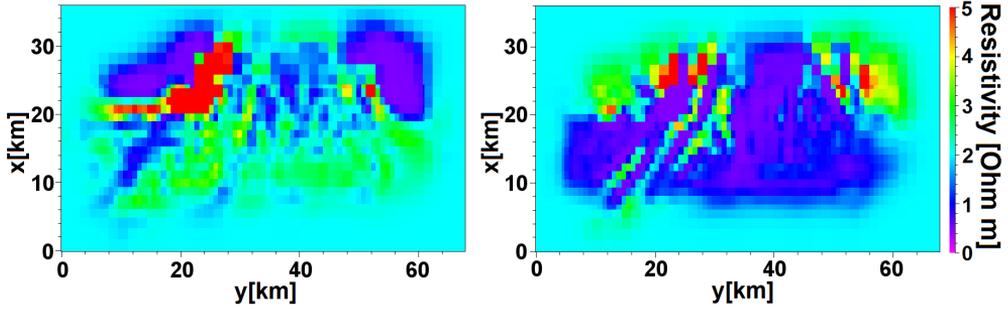
Figure 4.12: Survey scheme (B) inversion result: 3D Gauss-Newton optimizer result ($\mathbf{H} = \mathbf{H}_D + \mathbf{H}_R$) using CG solver and model parameter compression; depth-slices at 4200 m for the vertical resistivity model (left) and the horizontal resistivity model (right); the total number of free parameters was $5 \cdot 10^4$.

Figure 4.13 shows results obtained using the low-rank approximation (C) to the data part of the Hessian matrix in eq. 4.15. For this case we used the same model compression as in case (B). The results were obtained after 15 iterations, and the number of forward solutions per iteration was 1506 (22590 forward solutions in total) using $N_s = 1$ source groups. The maximum memory usage was similar to case (B).
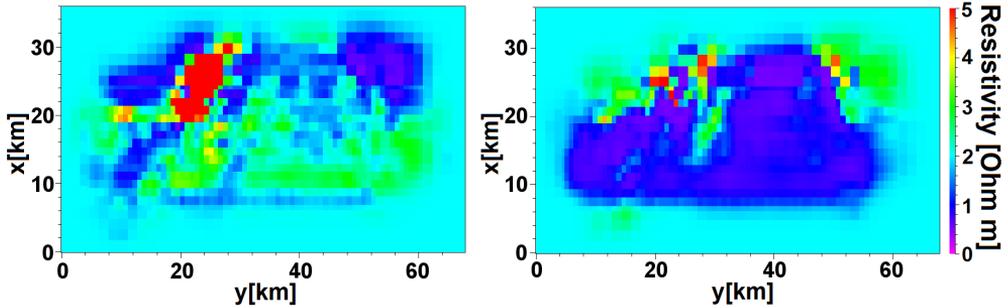
Figure 4.13: Survey scheme (C) inversion result: low-rank optimizer ($\tilde{\mathbf{H}} = \tilde{\mathbf{H}}_D + \mathbf{H}_R$) using CG solver and model parameter compression; depth-slices at 4200 m for the vertical resistivity model (left) and the horizontal resistivity model (right); the total number of free parameters was $5 \cdot 10^4$ and the number of source groups was $N_s = 1$.

In figure 4.14 we compare the eigenvalue distribution of the Hessian matrices for the first iteration in cases (B) and (C), with the eigenvalue distribution of their data Hessian matrices (without 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 shown in figure 4.5.
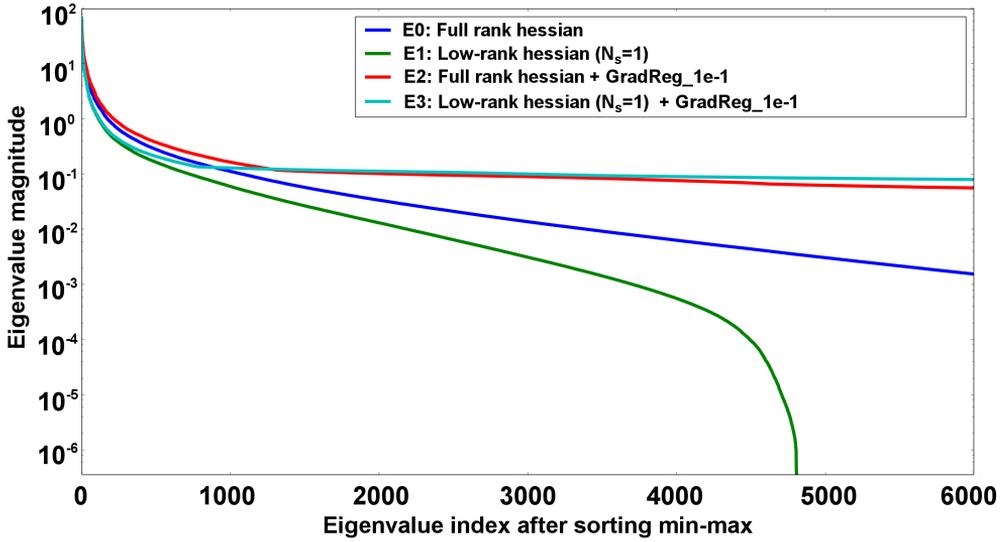
Figure 4.14: Eigenvalue decomposition for the first-iteration Hessian matrices (with and without regularization) of the inversions (B) and (C). E0: data Hessian matrix $\mathbf{H}_D$ of inversion (B). E1: data Hessian matrix $\tilde{\mathbf{H}}_D$ of inversion (C). E2: Hessian matrix $\mathbf{H} = \mathbf{H}_D + \mathbf{H}_{grad}$, with $W_{grad}(\sigma) = 1e^{-1}$, of inversion (B). E3: Hessian matrix $\tilde{\mathbf{H}} = \tilde{\mathbf{H}}_D + \mathbf{H}_{grad}$, with $W_{grad}(\sigma) = 1e^{-1}$, of inversion (C).

Finally, we show inversion results for the case (D) 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 the eq. 4.47.
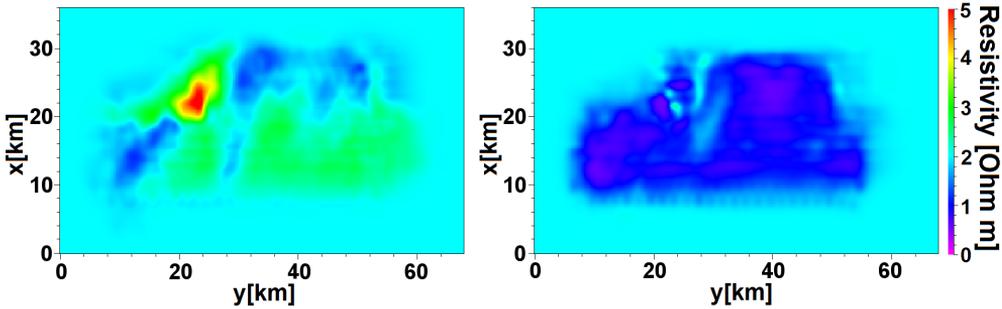
Figure 4.15: Survey scheme (D) inversion result: low-rank optimizer using the recursive solver with the diagonal of the regularization Hessian matrix ($\hat{\mathbf{H}} = \tilde{\mathbf{H}}_D + \tilde{\mathbf{H}}_R$); depth-slices at 4200 m for the vertical resistivity model (left) and the horizontal resistivity model (right); the total number of free parameters was 10^7 and the number of source groups was $N_s = 1$.

For this case we executed 21 iterations of the inversion, and the number of forward solutions per iteration was 1506 (31584 simulations in total). We used the same number of free parameters, 10^7, as in case (A) with no compression of model parameters. The memory usage to construct the update was 180 GB. As explained above, for this par-

ticular case we neglected the off-diagonal parts of the regularization Hessian matrix. This approximation may have impaired the inversion convergence.

The large-scale structures in the recovered resistivity models are similar for the four inversion approaches. There is a difference in scale between A,D and B,C 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,

$$\chi = \frac{\left|F^{\mathrm{Obs}} - F^{\mathrm{Synth}}\right|}{\delta F^{\mathrm{obs}}}. \tag{4.50}$$

This quantity, that we call "significant misfit", gives the discrepancy between observed and synthetic data in units of the estimated observed data measurement uncertainty, denoted by δF^{obs}. Figure 4.16 shows the common mid-point sorted data misfit χ for offsets 4 km and 8 km, for the case of the low-rank approximation with the recursive solver (D). 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 NW and NE corners of the survey area, where salt bodies are present. The data coverage for these structures is poor due to lack of receivers over these regions. Only the extension of source towlines out of the receiver grid give data that sample this part of the model. The resulting lack of information makes it difficult for the inversion to fit these data.
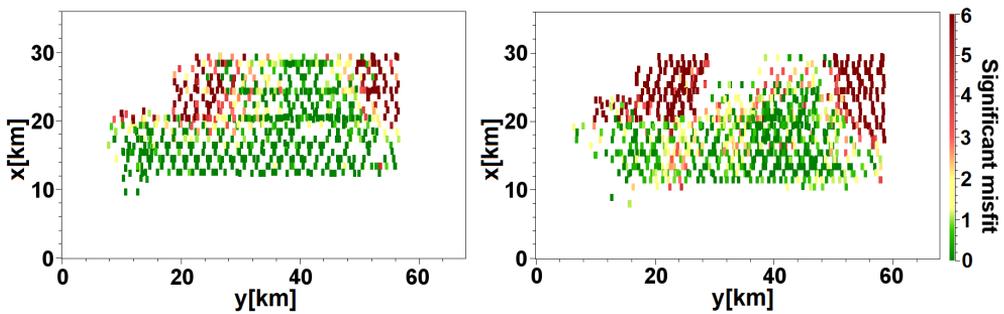
Figure 4.16: Significant misfit (eq. 4.50) for survey scheme (D) inversion results: low-rank optimizer using the recursive solver with the diagonal of the regularization Hessian matrix ($\hat{\mathbf{H}} = \tilde{\mathbf{H}}_{\mathrm{D}} + \tilde{\mathbf{H}}_{\mathrm{R}}$). For offsets of 4 km (left) and for offsets of 8 km (right).

4.4   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 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. 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 4.1 we show a comparison between the computational cost of the different schemes for which we showed inversion results above.

	(A) L-BFGS-B $N = 10^7$	(B) 3DGN + CG $N_s = 1$, $N = 5 \cdot 10^4$	(C) LR + CG $N_s = 1$, $N = 5 \cdot 10^4$	(D) LR + DS diag $N_s = 1, N = 10^7$
Number of simulations per iteration	$\mathcal{O}(N_{rx})$	$\mathcal{O}(N_{tx})$	$\mathcal{O}(N_{rx}N_s)$	$\mathcal{O}(N_{rx}N_s)$
Jacobian columns (rank) (M or $\tilde{M}$)	10 updates	$\mathcal{O}(N_{tx})$	$\mathcal{O}(N_s)$	$\mathcal{O}(N_s)$
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	$\mathcal{O}(N)$	$\mathcal{O}(N^2)$	$\mathcal{O}(N^2)$	$\mathcal{O}(N\tilde{M})$

Table 4.1: Computational complexity for the different schemes: (A) L-BFGS-B; (B) 3D Gauss-Newton using Conjugate Gradient (3DGN + CG); (C) Low-rank approach using Conjugate gradient (LR + CG); (D) Low-rank approach using the proposed recursive direct solver but just using the diagonal of the regularization matrix (LR + DS diag). k is the condition number of the Hessian matrix, that in general depends on the size of the matrix.

Table 4.1 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 number of sources N_{tx} like 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 (5 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 (C) and (D) described above. In table 4.2 we summarize the computational complexity of each of the four tests.

	(A) L-BFGS-B	(B) 3DGN + CG	(C) LR + CG	(D) LR + DS diag
Number of free parameters (N)	10^7	$5 \cdot 10^4$	$5 \cdot 10^4$	10^7
Number of simulations per iteration	1004	2806	1506	1506
Jacobian columns (rank) (M or $\tilde{M}$)	10 updates	550000	2510	2510
Solver time complexity (FLOPs)	10^9	$2.5 \cdot 10^{12}$ $[10^{17}]$	$2.5 \cdot 10^{12}$ $[10^{17}]$	$6.3 \cdot 10^{12}$
Solver memory complexity (GB)	0.149	12 $[93132]$	12 $[93132]$	180
Total number of iterations	60	13	15	21
Total number of simulations	60240	36478	22590	31584

Table 4.2: Computational cost to obtain inversion results:(A) L-BFGS-B ; (B) 3D Gauss-Newton using Conjugate Gradient (3DGN + CG); (C) Low-rank approach using Conjugate gradient (LR + CG) ; (D) 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 (B) and (C) show the cost in the case where the model parameter compression scheme is not applied, i.e. in the case of using 10^7 free parameters.

The 3D Gauss-Newton scheme (B) and low-rank schemes (C) and (D) require fewer iterations than the quasi-Newton case (A) 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 (C) and (D)), the total number

of simulations demonstrates that fewer forward solutions are required than in case of the 3D Gauss-Newton scheme (B) 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 (C) and (D) 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. Since computational complexity is proportional to the number of Jacobian columns, this is an important factor to take into consideration. In implementations where 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 table 4.1 and table 4.2, and in our example leads to reduction by a factor $M/\tilde{M} = 550000/2510 \approx 200$.

4.5   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 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 receiver 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 full rank 3D Gauss-Newton, but a number of simulations per iteration closer to a quasi-newton scheme. The recursive direct solver allows to use a larger number of inversion parameters than with the 3D Gauss-Newton scheme due to the smaller use of memory.

4.6   Acknowledgements

4.7 Appendix

In this appendix we will present inversion results for the synthetic survey shown in figure 4.2. This survey example was used to illustrate certain theoretical aspects of the low-rank approximation in the section 4.2. The numerical complexity involved in the inversion of this synthetic dataset is not very large, but the results are useful to validate the different schemes considered.

We consider four different inversion methodologies: (A_s) quasi-Newton L-BFGS-B (Zhu et al., 1997), (B_s) 3D Gauss-Newton with a conjugate gradient (CG) solver for eq. 4.4, (C_s) the proposed low-rank approach with a CG solver for eq. 4.15, and (D_s) the low-rank approach using the presented recursive direct solver with the approximation in eq. 4.48. Due to the small size of the synthetic survey considered, and in contrast to the field data example studied in the section 4.3, we do not need to use a model parameter compression. The inversion parametrization is a regular grid with cell size 200 m×200 m×100 m. We consider a VTI model, resulting in $5 \cdot 10^4$ 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 Ωm, and a water resistivity of 0.25 Ωm. We terminated the iterations when the convergence rate became too slow (misfit change in new iteration smaller than 1%).

In figure 4.17 we show the inversion results for the scheme (A_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 about 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 favour a model with resistivity contrasts rather than smooth variations. Note that the target is reconstructed in the vertical resistivity component since 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.
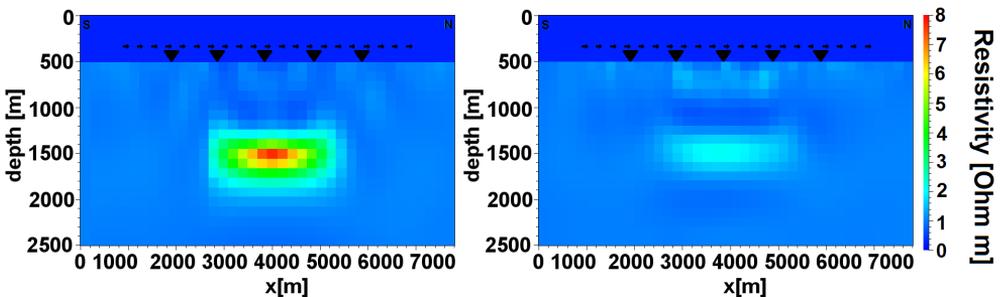
Figure 4.17: Synthetic-survey scheme (A_s) inversion result: L-BFGS-B optimizer result; vertical slice at y=2900 m for the vertical resistivity model (left) and the horizontal resistivity model (right); iteration 74 with misfit=0.382 RMS.

The inversion result for the scheme (B_s) is shown in figure 4.18. This result was obtained after 53 iterations, using 645 forward solutions per iteration (34185 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.
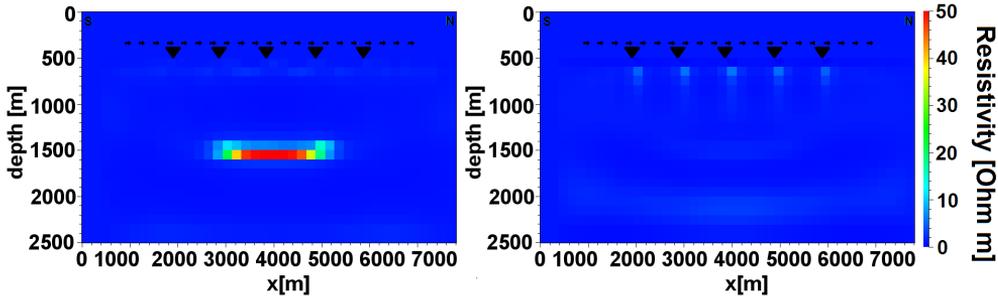
Figure 4.18: Synthetic-survey scheme (B_s) inversion result: 3D Gauss-Newton optimizer result $(\mathbf{H} = \mathbf{H_D} + \mathbf{H_R})$ using CG solver; vertical slice at y=2900 m for the vertical resistivity model (left) and the horizontal resistivity model (right); iteration 53 with misfit=0.125 RMS.

Figure 4.19 shows the inversion results for the scheme (C_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 (B_s). However, the target is situated 100 m above its position in the true model. Considering the low frequencies involved, we belive that this difference (on the scale of the discretization) is below the resolution of the data.
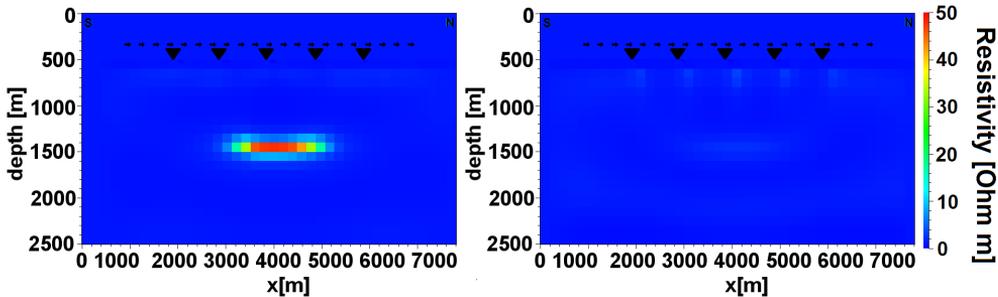
Figure 4.19: Synthetic-survey (C_s) inversion result: low-rank optimizer $(\tilde{\mathbf{H}} = \tilde{\mathbf{H}}_\mathbf{D} + \mathbf{H_R})$; vertical slice at y=2900 m for the vertical resistivity model (left) and the horizontal resistivity model (right); iteration 76 with misfit=0.137 RMS.

For the synthetic survey that we considered in this appendix, the reduced memory usage from the recursive direct solver is not relevant since the low number of model parameters makes the CG solver feasible even without model parameter compression.

However in order to validate the different approaches, we show in figure 4.20 the inversion results for scheme (D_s). This result was achived after 119 iterations, using 150 forward solutions per iteration (17850 simulations in total). The memory usage to compute the update was about 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 utilized when constructing the update in the data Hessian matrix free solver.
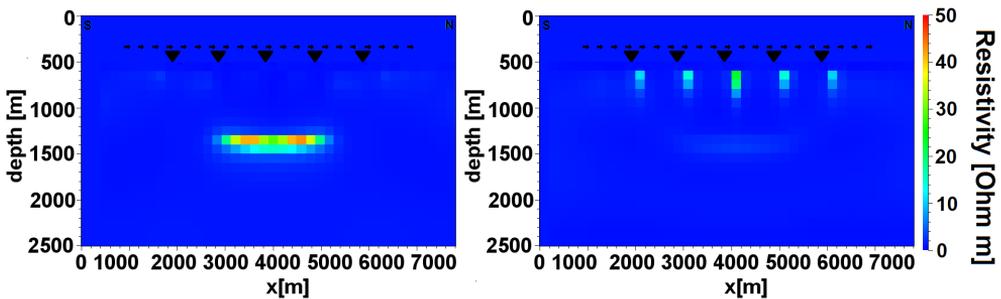
Figure 4.20: Synthetic-survey (D_s) inversion result: low-rank optimizer using the recursive solver with the diagonal of the regularization Hessian matrix ($\hat{\mathbf{H}} = \tilde{\mathbf{H}}_D + \tilde{\mathbf{H}}_R$); vertical slice at y=2900 m for the vertical resistivity model (left) and the horizontal resistivity model (right); iteration 119 with misfit=0.139 RMS.

4.8   References

Abubakar, A., Habashy, T., Druskin, V., Alumbaugh, D., Zerelli, A., and Knizhnerman, L.   Two-and-half-dimensional forward and inverse modeling for marine CSEM. *SEG Expanded Abstracts*, 150:750–754, 2006. doi: 10.1190/1.2370366. URL http://dx.doi.org/10.1190/1.2370366.

Abubakar, A., Habashy, T. M., Li, M., and Liu, J. Inversion algorithms for large-scale geophysical electromagnetic measurements. *IOP Science Inverse Problems*, 25: 123012, 2009. URL http://stacks.iop.org/0266-5611/25/i=12/a=123012.

Bansal, R., Krebs, J., Routh, P., Lee, S., Anderson, J., Baumstein, A., Mullur, A., Lazaratos, S., Chikichev, I., and McAdow, D. Simultaneous-source full-wavefield inversion. *The Leading Edge*, 32(9):1100–1108, 2013. doi: 10.1190/tle32091100. 1. URL http://dx.doi.org/10.1190/tle32091100.1.

Ben-Hadj-Ali, H., Operto, S., and Virieux, J.   An efficient frequency-domain full waveform inversion method using simultaneous encoded sources. *Geophysics*, 76

(4):R109–R124, 2011. doi: 10.1190/1.3581357. URL http://dx.doi.org/10.1190/1.3581357.

Boonyasiriwat, C. and Schuster, G. T. 3d multisource full-waveform inversion using dynamic random phase encoding. In *SEG Technical Program Expanded Abstracts*, pages 1044–1049, 2010. doi: 10.1190/1.3513025. URL http://dx.doi.org/10.1190/1.3513025.

deGroot Hedlin, C. and Constable, S. Occam's inversion to generate smooth, two-dimensional models from magnetotelluric data. *Geophysics*, 55(55):1613–1624, 1990. doi: 10.1190/1.1442813. URL http://dx.doi.org/10.1190/1.1442813.

Grayver, A. V., Streich, R., and Ritter, O. Three-dimensional parallel distributed inversion of CSEM data using a direct forward solver. *Geophysical Journal International*, 193(3):1432–1446, 2013. doi: 10.1093/gji/ggt055. URL http://dx.doi.org/10.1093/gji/ggt055.

Hesthammer, J., Stefatos, A., Boulaenko, M., Fanavoll, S., and Danielsen, J. CSEM performance in light of well results. *The Leading Edge*, 29(1):34–41, 2010. doi: 10.1190/1.3284051. URL http://dx.doi.org/10.1190/1.3284051.

Hoversten, G., Myer, D., Key, K., Hermann, O., Hobbet, R., and Alumbaugh, D. Csem & mmt base basalt imaging. In *75th EAGE Conference & Exhibition incorporating SPE EUROPEC*, 2013. doi: 10.3997/2214-4609.20130137. URL http://dx.doi.org/10.3997/2214-4609.20130137.

Jing, X., Finn, C. J., Dickens, T. A., and Willen, D. E. Encoding multiple shot gathers in prestack migration. In *SEG Technical Program Expanded Abstracts*, pages 786–789, 2000. doi: 10.1190/1.18161887. URL http://dx.doi.org/10.1190/1.18161887.

Krebs, J. R., Anderson, J. E., Hinkley, D., Neelamani, R., Lee, S., Baumstein, A., and Lacasse, M.-D. Fast full-wavefield seismic inversion using encoded sources. *Geophysics*, 74(6):WCC177–WCC188, 2009. doi: 10.1190/1.3230502. URL http://dx.doi.org/10.1190/1.3230502.

Li, M., Abubakar, A., Liu, J., Pan, G., and Habashy, T. M. A compressed implicit jacobian scheme for 3d electromagnetic data inversion. *Geophysics*, 76(3):F173–F183, 2011. doi: 10.1190/1.3569482. URL http://dx.doi.org/10.1190/1.3569482.

Li, Y. and Key, K. 2D marine controlled-source electromagnetic modeling: Part 1 an adaptive finite element algorithm. *Geophysics*, 72(2):WA51–WA62, 2007. doi: 10.1190/1.2432262. URL http://dx.doi.org/10.1190/1.2432262.

Lin, Y., Li, M., Abubakar, A., and Habashy, T. M. A wavelet-based model compression method for three-dimensional electromagnetic data inversion. In *SEG Technical Program Expanded Abstracts*, pages 707–712, 2013. doi: 10.1190/segam2013-0395.1. URL http://dx.doi.org/10.1190/segam2013-0395.1.

Loke, M. and Dahlin, T. A comparison of the gauss-newton and quasi-newton methods in resistivity imaging inversion. *Journal of Applied Geophysics*, 49(3):149–162, 2002. doi: 10.1016/S0926-9851(01)00106-9. URL http://doi.org/10.1016/S0926-9851(01)00106-9.

Mackie, R., Watts, M., and Rodi, W. Joint 3D inversion of marine CSEM and MT data. In *SEG Expanded Abstracts*, volume 116, pages 574–578, 2007. doi: 10.1190/1.2792486. URL http://dx.doi.org/10.1190/1.2792486.

Mittet, R., Maulana, H., Brauti, K., and Wicklund, T. A. CMP inversion of marine CSEM data. In *EGM 2007 International Workshop*, April 2007. URL http://www.earthdoc.org/publication/publicationdetails/?publication=41234.

Morten, J. P., Roth, F., Timko, D., Pacurar, C., Nguyen, A. K., and Olsen, P. A. 3D reservoir characterization of a North Sea oil field using quantitative seismic & CSEM interpretation. In *SEG Technical Program Expanded Abstracts*, pages 1903–1907, 2011. doi: 10.1190/1.3627578. URL http://dx.doi.org/10.1190/1.3627578.

Morten, J., Twarz, C., Ricoy-Paramo, V., and Sun, S. Improved resolution salt imaging from 3D CSEM anisotropic inversion. *75th EAGE Conference & Exhibition*, 2013. doi: 10.3997/2214-4609.20130139. URL http://dx.doi.org/10.3997/2214-4609.20130139.

Morton, S. A. and Ober, C. C. Fastershot-record depth migrations using phase encoding. In *SEG Technical Program Expanded Abstracts*, pages 1131–1134, 1998. doi: 10.1190/1.1820088. URL http://dx.doi.org/10.1190/1.1820088.

Press, W. H., Teukolosky, S. A., Vetterling, W. T., and Flannery, B. P. *Numerical recipies in C++*, chapter 2, page 76. Cambridge University Press, second edition edition, 2002. ISBN 978-0521431088. URL http://apps.nrbook.com/c/index.html.

Romero, L. A., Ghiglia, D. C., Ober, C. C., and Morton, S. A. Phase encoding of shot records in prestack migration. *Geophysics*, 65(2):426–436, 2000. doi: 10.1190/1.1444737. URL http://dx.doi.org/10.1190/1.1444737.

Schiemenz, A. and Igel, H. 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(3):1970–1988, 2013. doi: 10.1093/gji/ggt362. URL http://dx.doi.org/10.1093/gji/ggt362.

Støren, T., Zach, J., and Maaø, F. Gradient calculations for 3D inversion of CSEM data using a fast finite-difference time-domain modelling code. In *70th EAGE Conference & Exhibition, P194*, 2008. doi: 10.3997/2214-4609.20147963. URL http://dx.doi.org/10.3997/2214-4609.20147963.

Zhdanov, M. S. *Geophysical Electromagnetic Theory and Methods*. Elsevier Science, 1st edition, 2009. ISBN 9780444529633. URL http://store.elsevier.com/Geophysical-Electromagnetic-Theory-and-Methods/Michael-Zhdanov/isbn-9780080931760/.

Zhdanov, M. S., Gribenko, A., and Cuma, M. Regularized focusing inversion of marine csem data using minimum verticalsupport stabilizer. In *SEG Technical Program Expanded Abstracts*, pages 579–583, 2007. doi: 10.1190/1.2792487. URL http://dx.doi.org/10.1190/1.2792487.

Zhu, C., Byrd, H., and Nocedal, J. L-BFGS-B: Algorithm 778: L-BFGS-B, FORTRAN routines for large scale bound constrained optimization. *ACM Transactions on Mathematical Software*, 23(4):550–560, 1997. doi: 10.1145/279232.279236. URL http://dx.doi.org/10.1145/279232.279236.

Chapter 5

3D CSEM data inversion using Newton and Halley class methods.

Beyond the state-of-the-art frontiers, the use of higher-order methods than Gauss-Newton, like Newton and Halley class methods, is not generally considered due to the computational cost involved. In this paper this problem is efficiently solved allowing the inversion of moderate size 3D CSEM survey data.

Authors: M. Amaya, K. R. Hansen and J.P. Morten

Abstract: We introduce the use of the Newton and the Halley class optimization methods for the inversion of 3D controlled source electromagnetic data. The inversion is formulated as a constrained non-linear least-squares problem which is solved by iterative optimization. These methods require the derivatives up to second order of the residuals with respect to model parameters. We show how Green functions determine the high order derivatives, and develop a diagrammatical representation of the residual derivatives. The Green functions are efficiently calculated on-the-fly, making use of a finite-differences frequency-domain forward modelling code based on a multi-frontal sparse direct solver. This allow us to build the second-order derivatives of the residuals keeping the memory cost in the same order as in a Gauss-Newton scheme. Model updates are computed with a trust-region based conjugate-gradient solver which does not require the computation of a stabilizer. We present inversion results for a synthetic survey and compare the Gauss-Newton, Newton, and super-Halley optimization schemes, and consider two different approaches to set the initial trust-region radius. All the methods reach convergence by the same number of iterations. Our analysis shows that each approach drives the convergence through a different path.

5.1 Introduction

Marine controlled-source electromagnetic (CSEM) surveys are used to map the conductivity of the subsurface. In offshore hydrocarbon exploration applications, the CSEM data are typically interpreted together with other geophysical data, such as seismic data, to determine the presence and properties of hydrocarbon reservoirs (Eidesmo et al., 2002). The CSEM data can also be used to support exploration and seismic processing by structural imaging applications (Hoversten et al., 2013; Morten et al., 2011). In a CSEM survey (Constable and Srnka, 2007), a powerful electric dipole is towed close to the seabed transmitting a low-frequency signal that penetrates the subsurface. The low frequency ensures a significant penetration depth. The presence of thin resistive layers, such as resulting from a hydrocarbon accumulation, produces a wave-guide effect. This effect can give a large response in the data, which makes detection of hydrocarbon charged reservoirs possible. A grid of receivers deployed on the seabed measures the EM field. Imaging of the processed data is achieved by iterative modelling and inversion workflows that produce a resistivity model of the subsurface. State-of-the-art techniques for acquisition and imaging generate 3D resistivity models of the survey area, for a recent example see Fanavoll et al. (2014). The presence, volume, and properties of imaged resistors can be used for risk assessment in exploration.

In this paper we focus on the development and implementation of high-order inversion methods that so far are not applied in 3D CSEM inversion. These inverse problems are typically formulated as a local optimization problem that finds a minimum of the total residual between the observed data and the synthetic responses, the cost function, from a model within the neighbourhood of the initial guess. The global minimum of such an optimization problem could in principle be found by a stochastic approach. However, the number of free parameters and modelling complexity involved in a 3D CSEM inverse problem make stochastic methods unaffordable due to the very large computational cost.

The main interest for developing high-order inversion methods is to achieve faster convergence without the need of expending too much manual effort building detailed initial models. An example of this interest is the transition from quasi-Newton methods (Plessix and Mulder, 2008) to Gauss-Newton methods (Abubakar et al., 2009). Higher order methods than Gauss-Newton are not generally considered efficient due to the amount of storage and computation needed. There are some problems where properties like system matrix (tensor) sparsity can make it affordable (Gundersen and Steihaug, 2011). However, as we show in this paper, when developing high-order inversions methods in 3D CSEM, large and dense matrices (tensors) need to be processed.

To date, inversion based on Gauss-Newton optimization is the highest order method typically in use for imaging of 3D CSEM data. This method considers derivatives of the optimization cost function with respect to model parameters to the second order by an approximate Hessian matrix. The Gauss-Newton approximation to the Hessian

only includes the first-order derivative of the residuals. A typical 3D CSEM survey involves $N \sim 10^6$ inversion parameters and $N_\kappa \sim 10^7$ data samples. The size of the Gauss-Newton Hessian matrix scales as N^2, making its storage requirement a limiting factor when processing large surveys. In Lin et al. (2013) a model compression method is introduced that allows to reduce the number of inversion parameters. Another challenge is the number of forward simulations required to compute the Hessian matrix for a 3D CSEM Gauss-Newton inversion with a large number of data samples. In order to reduce this computational cost, a low-rank approximation to the Hessian matrix was presented in Amaya et al. (2014). There are other approaches to cope with the computational complexity of the 3D CSEM Gauss-Newton scheme: in Li et al. (2010) a compressed implicit Jacobian method is presented; Zaslavsky et al. (2013) shows the use of a fast 3D forward solver using an optimal subspace selection to reduce the size of the Jacobian matrix.

Beyond the Gauss-Newton method, it is necessary to compute higher order derivatives of the residuals with respect to model parameters. The computation of these high-order derivatives requires the computation of tensor Green functions for the entire parameter domain, *i.e.* the complete electromagnetic field resulting from a unit point source at any spatial position in the model. In 3D CSEM time complexity from obtaining millions of these Green functions could become the main bottle neck of high-order inversion schemes. The performance achieved by current direct-solvers (Grayber and Streich, 2012) makes those tensor Green functions simulations affordable for the purposes of this paper.

With this paper we would like to introduce inversion using high-order methods beyond the Gauss-Newton approach for imaging of 3D CSEM data. We show that if Green functions can be computed efficiently, high-order methods can be implemented without a significant increase in storage requirements. The main increase in the computational cost would be based on the construction and storage of the derivatives of the misfit function.

The storage of tensor Green functions and multi-dimensional matrices, that appear in higher-order methods than 3D CSEM Gauss-Newton, significantly increases the memory complexity. We have solved the storage difficulties for developing the two inversion methods that we present in this paper, 3D CSEM Newton and 3D CSEM Halley class methods (Ezquerro and Hernández, 2003; Gutierrez and Hernandez, 2001; Han, 2001). We compute the Green functions on the fly using a finite-difference frequency-domain (FDFD) sparse direct solver, avoiding the construction of the rank-3 tensor in a way that allows us to keep the memory cost in the same order as in the case of a Gauss-Newton approximation to the Hessian matrix.

Higher-order inversion methods than 3D CSEM Gauss-Newton have the difficulty of solving linear equation systems with indefinite matrices. Therefore we suggest the implementation of a trust-region solver that can handle the non-positive-definite matrices.

We present the implementation of a 3D CSEM Newton and a 3D CSEM super-Halley's method, both based on the computation of second-order derivatives of the

residuals, and we show synthetic results.

The organization of this paper is the following. First we show how to obtain the third-order approximation of the data-misfit model cost function. After a diagrammatical explanation of such expansions, we derive the stationary point equations. Next we explain how we obtain the forward modelling solutions and the way we store the matrices. Then we discuss the implementation of the trust-region solution algorithm used in this paper. Finally, we apply our scheme to a synthetic model, and compare results from the Gauss-Newton, Newton, and super-Halley inversions.

5.2 Theory

The inversion of CSEM data is formulated as a constrained non-linear least-squares optimization problem,

$$\boldsymbol{\sigma}^{\star} = \arg \min_{\boldsymbol{\sigma} \in \mathcal{M}} \varepsilon(\boldsymbol{\sigma}), \tag{5.1}$$

where $\boldsymbol{\sigma}$ is a discrete parameterization of the 3D conductivity model, $\mathcal{M}$ represents the set of possible models based on a priori information, and

$$\varepsilon(\boldsymbol{\sigma}) = \varepsilon_{\mathrm{D}}(\boldsymbol{\sigma}) + \lambda \, \varepsilon_{\mathrm{R}}(\boldsymbol{\sigma}), \tag{5.2}$$

is the cost function. This cost function includes the regularization misfit term, ε_{R}, with λ an adjustable scalar parameter. The data misfit term, ε_{D}, depends on the observations through a L_2-norm,

$$\varepsilon_{\mathrm{D}}(\boldsymbol{\sigma}) = d_{\kappa} d_{\kappa}^{*}, \tag{5.3}$$

$$d_{\kappa} = W_{\kappa} \Delta F_{\kappa}. \tag{5.4}$$

In order to simplify the notation, we use the Einstein summation convention so a sum over the repeated index κ is implied in the above equations. The short hand notation $\kappa = (F, i, f, \mathbf{r_{rx}}, \mathbf{r_{tx}})$ uniquely labels a measurement, F represents a field component ($F = E$ for electric and $F = H$ for magnetic), i are the spatial components (x, y) of the field recordings, f are the frequencies, $\mathbf{r_{rx}}$ is a receiver position, and $\mathbf{r_{tx}}$ is a source position. Here $\boldsymbol{d} = (d_1, ..., d_{N_{\kappa}})^T$ is a vector that contains N_{κ} residuals, and $\boldsymbol{d}^*$ is its complex conjugate. W_{κ} is a datum weight (typically inverse standard deviation) and $\Delta F_{\kappa} = F_{\kappa}^{\mathtt{Synth}}(\boldsymbol{\sigma}) - F_{\kappa}^{\mathtt{Obs}}$ represents the difference between synthetic and observed fields.

The regularization term $\varepsilon_{\mathrm{R}}(\boldsymbol{\sigma})$ introduces a priori information about the solution model $\boldsymbol{\sigma}^{\star}$ (Zhdanov, 2009; Portniaguine and Zhdanov, 1999). The regularization term can be designed to achieve a semipositive-definite Hessian matrix by adjusting the "trade-off factor" λ in eq. 5.2. In a Gauss-Newton approximation, the sum of the data Hessian $\mathcal{H}^{\mathtt{GN}} \approx \boldsymbol{\nabla}^2 \varepsilon_{\mathrm{D}}$ and the regularization Hessian matrices will be a positive-definite matrix. However, in the higher order methods that we apply in this paper, the

data Hessian matrix $\mathcal{H} = \nabla^2 \varepsilon_{\mathrm{D}}$ and the tensor $\mathcal{T} \approx \nabla^3 \varepsilon_{\mathrm{D}}$ are indefinite. Therefore, the regularization term that we use in this paper just imposes model smoothness, and is not used to ensure positive definite characteristic of the Hessian matrix $\mathcal{H}$ and the tensor $\mathcal{T}$. In the following, we focus on the data misfit term. Details about the regularization can be found in our references.

5.2.1 Data-misfit model cost function

The problem (eq. 5.1) is iteratively solved starting from an initial guess $\boldsymbol{\sigma}^{(1)}$, by updating the model with a new step $\Delta\boldsymbol{\sigma}$ at each iteration, so that iterate $\boldsymbol{\sigma}^{(k+1)} = \boldsymbol{\sigma}^{(k)} + \Delta\boldsymbol{\sigma}^{(k)}$. To obtain the step $\Delta\boldsymbol{\sigma}^{(k)}$ we define the model cost function $m^{(k)}(\Delta\boldsymbol{\sigma})$ that is an approximation to the cost function $\varepsilon(\boldsymbol{\sigma})$ in the vicinity of the current iterate $\boldsymbol{\sigma}^{(k)}$. At each iteration we constrain the updated model $\boldsymbol{\sigma}^{(k+1)}$ by applying a non-linear transformation of the parameters, forcing them to lie within the physical bounds (Habashy and Abubakar, 2004).

We define the data-misfit model cost function $m_{\mathrm{D}}^{(k)}(\Delta\boldsymbol{\sigma})$ that is used at each iteration by expanding the data-misfit $\varepsilon_{\mathrm{D}}(\boldsymbol{\sigma} + \Delta\boldsymbol{\sigma})$ in eq. 5.3 to the third-order in $\Delta\boldsymbol{\sigma}$, but approximating it by removing the third-order residual derivative contribution from $\nabla^3 \varepsilon_{\mathrm{D}}$. This expansion becomes,

$$
\begin{aligned}
m_{\mathrm{D}}^{(k)}(\Delta\boldsymbol{\sigma}) = {}& d_\kappa^{(k)} d_\kappa^{*\,(k)} + g_n^{(k)} \Delta\sigma_n^{(k)} + \frac{1}{2}\mathcal{H}_{n\,n'}^{(k)}\Delta\sigma_n^{(k)}\Delta\sigma_{n'}^{(k)} \\
& + \frac{1}{6}\mathcal{T}_{n\,n'\,n''}^{(k)}\Delta\sigma_n^{(k)}\Delta\sigma_{n'}^{(k)}\Delta\sigma_{n''}^{(k)},
\end{aligned}
\tag{5.5}
$$

$$
\nabla\varepsilon_{\mathrm{D}\,n}^{(k)} = g_n^{(k)} = d_\kappa^{*\,(k)}\mathcal{J}_{\kappa\,n}^{(k)} + \mathsf{c.c.},
\tag{5.6}
$$

$$
\nabla^2\varepsilon_{\mathrm{D}\,n\,n'}^{(k)} = \mathcal{H}_{n\,n'}^{(k)} = \mathcal{J}_{\kappa\,n}^{(k)}\mathcal{J}_{\kappa\,n'}^{*\,(k)} + d_\kappa^{*\,(k)}\mathcal{S}_{\kappa\,n\,n'}^{(k)} + \mathsf{c.c.},
\tag{5.7}
$$

$$
\nabla^3\varepsilon_{\mathrm{D}\,n\,n'\,n''}^{(k)} \approx \mathcal{T}_{n\,n'\,n''} = \mathcal{J}_{\kappa\,n}^{*\,(k)}\mathcal{S}_{\kappa\,n'\,n''}^{(k)} + \mathcal{J}_{\kappa\,n'}^{*\,(k)}\mathcal{S}_{\kappa\,n\,n''}^{(k)} + \mathcal{J}_{\kappa\,n''}^{*\,(k)}\mathcal{S}_{\kappa\,n\,n'}^{(k)} + \mathsf{c.c.},
\tag{5.8}
$$

where c.c. represents the complex conjugate of the preceding terms. Defining N as the number of model parameters with $n = 1, \cdots, N$, the gradient $\boldsymbol{g}$ is a vector of N real-elements and the Hessian $\mathcal{H}$ is a real $N \times N$ matrix. The Jacobian matrix $\mathcal{J}$ is a $N_\kappa \times N$ complex matrix containing the first derivatives of the residuals d_κ with respect to model parameters. The quantity $\mathcal{S}$ is a $N_\kappa \times N \times N$ tensor formed by the second derivatives of the residuals with respect to model parameters.

The Jacobian matrix $\mathcal{J}$ can be constructed from the first-order derivatives of the synthetic fields, see in the appendix (section 5.7) eq. 5.51 and eq. 5.52,

$$
\mathcal{J}_{\kappa\,n} = W_\kappa \frac{\partial F_\kappa^{\mathtt{Synth}}(\boldsymbol{\sigma})}{\partial\sigma_n},
\tag{5.9}
$$

$$\frac{\partial F_\kappa^{\texttt{Synth}}(\boldsymbol{\sigma})}{\partial \sigma_n} = \Delta V\left(\mathbf{r}_n\right) \sum_{p\in A_n,\,q} G_{i\,p}^{F\,J}\left(\mathbf{r}_{\mathbf{rx}}|\mathbf{r}_n,f,\boldsymbol{\sigma}\right) G_{p\,q}^{E\,J}\left(\mathbf{r}_n|\mathbf{r}_{\mathbf{tx}},f,\boldsymbol{\sigma}\right) J_q\left(\mathbf{r}_{\mathbf{tx}},f\right).$$

(5.10)

In this equation, $\Delta V(\mathbf{r}_n)$ is a volume factor that applies for the spatial discretization, and $\mathbf{r}_n$ is a suitably chosen interpolation point for the region associated with parameter σ_n. Note that an additional grid decoupling technique can be applied in eq. 5.10 to introduce more flexibility in the choice of inversion parameters. For example, extended regions consisting of several discretization cells for the evaluation of Green functions can be combined in the parameterization. However, in this paper we consider that the discretization of the computational domains for the Green functions and the inversion parameterization is the same. The index range denoted A_n represents the model anisotropy (Zhdanov, 2009), so that the range of the index p depends on the type of model parameter σ_n. In this work we consider a vertical transverse isotropic (VTI) model, *i.e.* we consider independent horizontal and vertical components for the conductivity of each parameter region. This means that the conductivity tensor has the structure $\sigma_{i\,j}(\mathbf{r}_n) = \delta_{i\,j}\left[\sigma_{\texttt{H}}(\mathbf{r}_n)(\delta_{i\,x}+\delta_{j\,y})+\sigma_{\texttt{V}}(\mathbf{r}_n)\delta_{i\,z}\right]$, where $\delta_{i\,j}$ is the Kronecker delta. Thus, for the horizontal conductivity parameters $p = x, y$, and for the vertical conductivity parameters $p = z$ in eq. 5.10. The Green function $G_{i\,p}^{F\,J}(\mathbf{r}_{\mathbf{rx}}|\mathbf{r},f,\boldsymbol{\sigma})$ denotes the field F for component i ($i = x, y$) at position $\mathbf{r}_{\mathbf{rx}}$, given a unit current source in direction p operating at position $\mathbf{r}$ with frequency f. All Green function expressions in this paper use an equivalent notation. The factor $J_q(\mathbf{r}_{\mathbf{tx}})$ approximates the vector current source contribution in direction q at position $\mathbf{r}_{\mathbf{tx}}$ as a point dipole.

The second-order derivatives of the residuals that appear in the tensor $\mathcal{S}$ can also be derived from Maxwell's equation,

$$\mathcal{S}_{\kappa\,n\,n'} = W_\kappa \frac{\partial^2 F_\kappa^{\texttt{Synth}}(\boldsymbol{\sigma})}{\partial \sigma_n \partial \sigma_{n'}},$$

(5.11)

$$\frac{\partial^2 F_\kappa^{\texttt{Synth}}(\boldsymbol{\sigma})}{\partial \sigma_n \partial \sigma_{n'}} = \Delta V\left(\mathbf{r}_n\right) \Delta V\left(\mathbf{r}_{n'}\right) \sum_{m\in A_n,\,p\in A_{n'},\,q}$$
$$\left[G_{i\,m}^{F\,J}\left(\mathbf{r}_{\mathbf{rx}}|\mathbf{r}_n,f,\boldsymbol{\sigma}\right) G_{m\,p}^{E\,J}\left(\mathbf{r}_n|\mathbf{r}_{n'},f,\boldsymbol{\sigma}\right) G_{p\,q}^{E\,J}\left(\mathbf{r}_{n'}|\mathbf{r}_{\mathbf{tx}},f,\boldsymbol{\sigma}\right) J_q\left(\mathbf{r}_{\mathbf{tx}},f\right) \right.$$
$$\left. + G_{i\,p}^{F\,J}\left(\mathbf{r}_{\mathbf{rx}}|\mathbf{r}_{n'},f,\boldsymbol{\sigma}\right) G_{p\,m}^{E\,J}\left(\mathbf{r}_{n'}|\mathbf{r}_n,f,\boldsymbol{\sigma}\right) G_{m\,q}^{E\,J}\left(\mathbf{r}_n|\mathbf{r}_{\mathbf{tx}},f,\boldsymbol{\sigma}\right) J_q\left(\mathbf{r}_{\mathbf{tx}},f\right) \right]. \quad (5.12)$$

The details of the derivation of this result can be found in our appendix (section 5.7) eq. 5.54, where we also explain that we apply reciprocity to reduce the number of simulations when computing the Green functions related to receiver position ($G_{i\,m}^{F\,J}\left(\mathbf{r}_{\mathbf{rx}}|\mathbf{r},f,\boldsymbol{\sigma}\right)$ and $G_{i\,p}^{F\,J}\left(\mathbf{r}_{\mathbf{rx}}|\mathbf{r},f,\boldsymbol{\sigma}\right)$).

The following symmetries follow from the definitions,

$$\mathcal{S}_{\kappa\,n\,n'} = \mathcal{S}_{\kappa\,n'\,n}, \tag{5.13}$$

$$\mathcal{H}_{n\,n'} = \mathcal{H}_{n'\,n}, \tag{5.14}$$

$$\mathcal{T}_{n\,n'\,n''} = \mathcal{T}_{n'\,n''\,n} = \mathcal{T}_{n''\,n\,n'}. \tag{5.15}$$

These symmetries facilitate a compressed storage scheme.

5.2.2   Diagrammatic representation of the derivatives

We develop a diagrammatic representation of the derivatives of the data-misfit ε_{D} from eq. 5.6, eq. 5.7, eq. 5.8, and the contributions to the data-misfit model cost function m_{D} (eq. 5.5). The diagrammatical representation appeals to the interpretation of the Green function $\boldsymbol{G}(\mathbf{r}_b|\mathbf{r}_a, \boldsymbol{\sigma})$ as a propagator. Here we suppress the frequency argument for the Green function. A propagator describes how a "disturbance" or source for the electric field at $\mathbf{r}_a$ leads to a response at position $\mathbf{r}_b$. In the diagrammatic representation this process is indicated as an arrow originating at $\mathbf{r}_a$ and terminating at $\mathbf{r}_b$. Using this identification, *e.g.* the first and second-order derivatives of the fields that appear in eq. 5.10 and eq. 5.12 can be represented as in figure 5.1.
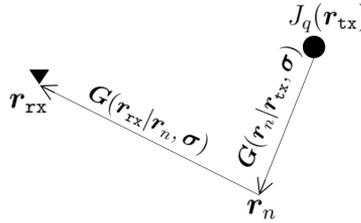
(a) Diagrammatic representation of the first derivative of the synthetic field with respect to the conductivity parameter, $\partial F_{\kappa}^{\mathrm{Synth}}(\boldsymbol{\sigma})/\partial\sigma_n$. In eq. 5.10 it appears multiplied by the corresponding weight.
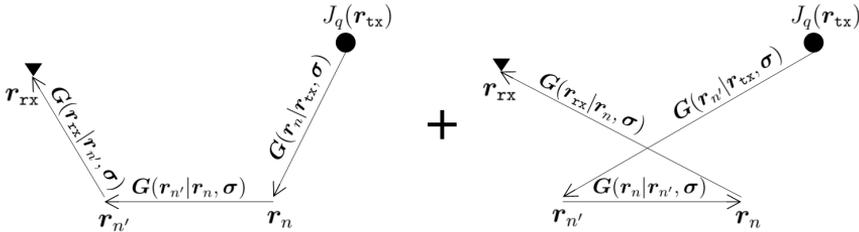
(b) Diagrammatic representation of the second derivative of the synthetic field with respect to conductivity parameters σ_n and $\sigma_{n'}$, $\partial^2 F_{\kappa}^{\mathrm{Synth}}(\boldsymbol{\sigma})/\partial\sigma_n\partial\sigma_{n'}$. In eq. 5.12 it appears multiplied by the corresponding weights.

Figure 5.1: Diagrams of the first-order and second-order derivatives of the fields with the source and receiver in the spatial representation $\sigma(\mathbf{r})$. The source distribution $J_q(\mathbf{r}_{\mathrm{tx}})$ is shown with a filled circle, and the receiver by a filled triangle.

Note that we do not make use of the reciprocity principle when constructing diagrammatic representations (see Appendix in section 5.7).

Such diagrammatic representations can be very useful to simplify the development of higher order derivatives, and to appreciate the physical significance of the various terms in the data-misfit model function. Moreover, a set of simple rules for the one-to-one identification of a diagram and a mathematical representation, allows for diagrams to be used systematically to construct the expressions for the contributions to a high-order expansion. In order to construct an n-th order derivative, we construct all possible connected diagrams with n internal nodes. The mathematical expressions are given by the product of the source factors and propagators along the path. See Appendix in section 5.7 for more details.

In figure 5.2 we present diagrammatical representations of the terms that make up the derivatives of the data-misfit term $\boldsymbol{\nabla}\varepsilon_{\mathrm{D}}$, $\boldsymbol{\nabla}^2\varepsilon_{\mathrm{D}}$, and $\boldsymbol{\nabla}^3\varepsilon_{\mathrm{D}}$. The gradient $\boldsymbol{\nabla}\varepsilon_{\mathrm{D}}$ (eq. 5.6) is given by the diagram in figure 5.2(a). This contribution to m_{D} is thus illustrated as the result of the propagation from the source position to the perturbed point in the model, and subsequent propagation to the receiver. This visual illustration corresponds with an interpretation of the perturbation as a scatterer. The factor $\left[W_\kappa^2 \Delta F_\kappa\right]^* J_q(\mathbf{r_{tx}})$, applied at the source coordinate $\mathbf{r_{tx}}$, denotes the source amplitude for the propagation. The total gradient will be a sum over all the data samples in the dataset, *i.e.* all field components, frequencies, and source-receiver combinations $(\mathbf{r_{rx}}, \mathbf{r_{tx}})$, as in eq. 5.3. In the diagrammatic representation we implicitly include the complex conjugate pair of each diagram. For the gradient, this would involve the corresponding diagram with $W_\kappa^2 \Delta F_\kappa J_q^*(\mathbf{r_{tx}})$ as the source, and the complex conjugation of the Green function propagators. We draw diagrams with complex conjugated Green functions as "reverse-time" propagation, in the sense that any series of propagators emanating at a receiver coordinate $\mathbf{r_{rx}}$ and ending at a source coordinate $\mathbf{r_{tx}}$ has an implied complex conjugation on the Green function. Explicitly, the mathematical expression for the gradient (eq. 5.6) is

$$
\begin{aligned}
g_n = &\Delta V\left(\mathbf{r}_n\right) \sum_{F,i,f,\mathbf{r_{rx}},\mathbf{r_{tx}}} \sum_{p\in A_n,q} \left[G_{ip}^{FJ}\left(\mathbf{r_{rx}}|\mathbf{r}_n, f, \boldsymbol{\sigma}\right) G_{pq}^{EJ}\left(\mathbf{r}|\mathbf{r_{tx}}, f, \boldsymbol{\sigma}\right) \right.\\
&\left. \times \left[W_i^F\left(\mathbf{r_{rx}}|\mathbf{r_{tx}}, f\right)\right]^2 \Delta F_i^*(\mathbf{r_{rx}}|\mathbf{r_{tx}}, f, \boldsymbol{\sigma}) J_q\left(\mathbf{r_{tx}}, f\right) \right] + \text{c.c.}.
\end{aligned} \tag{5.16}
$$

This expression can be seen to be one-to-one with the representation in figure 5.2(a) given the above set of rules for identification of elements in these diagrams and physical quantities.

Let us now consider the diagrammatic representation of the second-order terms involving the Hessian $\boldsymbol{\nabla}^2\varepsilon_{\mathrm{D}}$ (eq. 5.7). The diagrams are shown in figure 5.2(b). The first diagram shown involves a two-step propagation from the source to the receiver, as well as a similar propagation from the receiver to the source. The latter two Green functions, from the receiver to the source, are associated with complex conjugation

given the direction of the arrows from the receiver to the source, and the prescription described above. The complex conjugate pair diagram is implicitly included. Note that the diagram obtained when $\mathbf{r}' \to \mathbf{r}$ corresponds to the magnitude of the model parameter sensitivity.
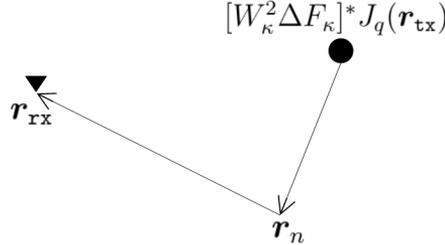
$$[W_\kappa^2 \Delta F_\kappa]^* J_q(\boldsymbol{r}_{\text{tx}})$$

$\boldsymbol{r}_{\text{rx}}$

$\boldsymbol{r}_n$

(a) Diagrammatic representation of contributions to $\boldsymbol{\nabla}\varepsilon_{\text{D}} = \boldsymbol{g}$.

$$W_\kappa^2 J_q(\boldsymbol{r}_{\text{tx}}) \qquad\qquad [W_\kappa^2 \Delta F_\kappa]^* J_q(\boldsymbol{r}_{\text{tx}})$$

$\boldsymbol{r}_{\text{rx}}$ $\boldsymbol{r}_{\text{rx}}$

$+$

$\boldsymbol{r}_{n'}$ $\boldsymbol{r}_n$ $\boldsymbol{r}_{n'}$ $\boldsymbol{r}_n$
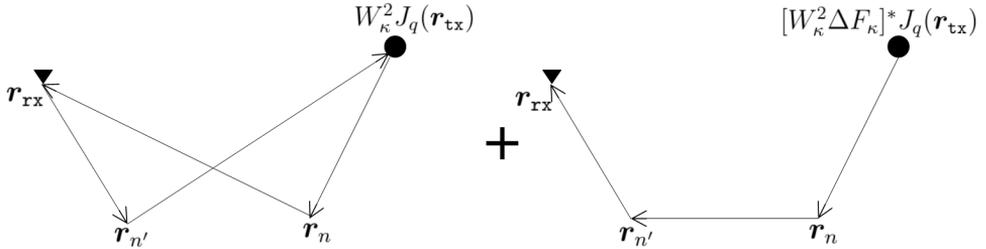
(b) Diagrammatic representation of contributions to $\boldsymbol{\nabla}^2\varepsilon_{\text{D}} = \boldsymbol{\mathcal{H}}$. The Gauss-Newton Hessian matrix $\boldsymbol{\mathcal{H}}^{\text{GN}}$ is an approximation only formed by the diagram on the left.

$$W_\kappa^2 J_q(\boldsymbol{r}_{\text{tx}}) \qquad\qquad [W_\kappa^2 \Delta F_\kappa]^* J_q(\boldsymbol{r}_{\text{tx}})$$

$\boldsymbol{r}_{\text{rx}}$ $\boldsymbol{r}_{\text{rx}}$

$\boldsymbol{r}_{n''}$

$+$

$\boldsymbol{r}_{n'}$

$\boldsymbol{r}_{n'}$ $\boldsymbol{r}_n$ $\boldsymbol{r}_{n''}$ $\boldsymbol{r}_n$
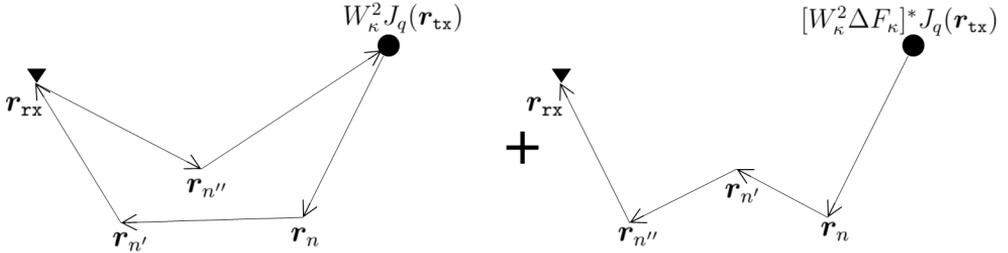
(c) Diagrammatic representation of contributions to $\boldsymbol{\nabla}^3\varepsilon_{\text{D}}$. In this paper, the quantity $\boldsymbol{\mathcal{T}}$ is an approximation only formed by the terms represented by the diagram on the left.

Figure 5.2: Diagrams of the data-misfit derivatives. Note that diagrams corresponding to complex conjugate pairs and permutations of the internal spatial nodes $\mathbf{r}_n, \mathbf{r}_{n'}, \mathbf{r}_{n''}$ are not shown, but implicitly assumed added.

The second diagram in figure 5.2(b) includes an "internal propagation" between the two positions $\mathbf{r}$ and $\mathbf{r}'$. Such contributions to the Hessian are omitted in a Gauss-Newton approach. In addition to the complex conjugate pair that can be constructed, the diagrammatic representation also implicitly includes all permutations of such internal nodes. This is similar to the situation shown in figure 5.1(b). In total there will be four contributions represented by the second diagram in figure 5.2(b). Note that

these latter four contributions are significantly more complicated to compute because of the internal propagator, which is discussed below.

Higher order contributions can be developed systematically by constructing all possible diagrams with the number of internal nodes corresponding to the order of the expansion. The diagrams for the third-order terms involving $\nabla^3 \varepsilon_D$ are shown in figure 5.2(c). The first type of diagram is a combination of a first and a second derivative. Complex conjugate pairs and internal node permutation introduces in total 12 of these terms. These are the type of contributions involved in the tensor $\mathcal{T}$ (eq. 5.8). The second type of diagram in figure 5.2(c) involves two internal propagators. There are also here 12 implied diagrams from complex conjugation and internal node permutation. Note that the second type of diagram in figure 5.2(c) is omitted in the approximation eq. 5.8 due to the numerical complexity involved in computing the double internal propagation.

5.2.3 Stationary point equations

Once the data-misfit model function m_D is defined, it is possible to get the stationary point equations that determines the steps $\Delta\boldsymbol{\sigma}^{(k)}$. A necessary condition of the solution eq. 5.1 is that the gradient of the objective function vanishes at $\boldsymbol{\sigma}^\star$. Applying this condition to eq. 5.5,

$$\frac{\partial m_D^{(k)}(\Delta\boldsymbol{\sigma})}{\partial \Delta\sigma_n} = 0, \tag{5.17}$$

we get the stationary point equation

$$g_n^{(k)} + \mathcal{H}_{n\,n'}^{(k)}\Delta\sigma_{n'}^{(k)} + \frac{1}{2}\mathcal{T}_{n\,n'\,n''}^{(k)}\Delta\sigma_{n'}^{(k)}\Delta\sigma_{n''}^{(k)} = 0. \tag{5.18}$$

In order to simplify the notation, from this point we suppress the iteration index k.

When applying a Gauss-Newton (GN) optimization scheme, eq. 5.18 is approximated to first-order in the residual derivatives, *i.e.* all terms determined by $\mathcal{S}$ are omitted. At each iteration, the step $\Delta\boldsymbol{\sigma}^{GN}$ is obtained from

$$\mathcal{H}_{n\,n'}^{GN}\,\Delta\sigma_{n'}^{GN} = -g_n, \tag{5.19}$$

$$\mathcal{H}_{n\,n'}^{GN} = \mathcal{J}_{\kappa\,n}^*\mathcal{J}_{\kappa\,n'} + \text{c.c..} \tag{5.20}$$

In this paper we apply two higher-order methods in 3D CSEM optimization, Newton's method and a Halley class method. As mentioned in the introduction, our intention is to explore higher-order methods by solving the numerical challenges that these methods present.

In order to apply Newton's method, the eq. 5.18 is approximated to first-order. At each iteration the step $\Delta\boldsymbol{\sigma}^N$ is obtained from,

$$\mathcal{H}_{n\,n'}\,\Delta\sigma_{n'}^N = -g_n, \tag{5.21}$$

where we use the complete expression for the Hessian matrix by including the second-order derivative of the residuals $\mathcal{S}$ (compare eq. 5.7 with eq. 5.20).

Another higher-order proposal is to solve eq. 5.18 by applying a Halley class method for multidimensional functions in the way that it is formulated in Gundersen and Steihaug (2011). This method can be formulated as applying two Newton steps at each iteration to obtain the step $\Delta\boldsymbol{\sigma}^{\mathtt{H}}$,

$$\Delta\sigma_n^{\mathtt{H}} = \Delta\sigma_n^{\mathtt{A}} + \Delta\sigma_n^{\mathtt{B}}, \tag{5.22}$$

$$\mathcal{H}_{n\,n'}\Delta\sigma_{n'}^{\mathtt{A}} = -g_n, \tag{5.23}$$

$$\left(\mathcal{H}_{n\,n'} + \alpha\mathcal{T}_{n\,n'\,n''}\Delta\sigma_{n''}^{\mathtt{A}}\right)\Delta\sigma_{n'}^{\mathtt{B}} = -\frac{1}{2}\mathcal{T}_{n\,n'\,n''}\Delta\sigma_{n'}^{\mathtt{A}}\Delta\sigma_{n''}^{\mathtt{A}}. \tag{5.24}$$

For the inversion results in this paper, we consider the super-Halley method where the parameter $\alpha = 1$. Other alternatives are the Chebyshev's method ($\alpha = 0$), and the Halley's method ($\alpha = 1/2$).

5.2.4 Forward solutions and system of equation construction

A 3D CSEM optimization method based on first-order derivatives of the residuals (eq. 5.10), like Gauss-Newton (eq. 5.19), requires the computation of the Green function for each source $\boldsymbol{G}(\mathbf{r}_n|\mathbf{r}_{\mathtt{tx}}, \boldsymbol{\sigma})$ and receiver position $\boldsymbol{G}(\mathbf{r}_{n'}|\mathbf{r}_{\mathtt{rx}}, \boldsymbol{\sigma})$ (here reciprocity has been applied), for each field-component and frequency. For higher order derivatives methods, in addition it is necessary to evaluate, for each field-component and frequency, a Green function $\boldsymbol{G}(\mathbf{r}_{n'}|\mathbf{r}_n, \boldsymbol{\sigma})$ for all pairs of cell positions in the model (internal propagation in the parameterized domain). In these cases, for a regular size 3D model, the number of internal Green function simulations is several orders of magnitude larger than the number of source and receiver Green function simulations.

Direct solvers have started to become practical for modelling of moderate-size 3D CSEM surveys and using state-of-the-art computer clusters (Da Silva et al., 2012; Yang and Oldenburg, 2012; Grayver et al., 2013; Schwarzbach and Haber, 2013). In some cases, direct solvers can be faster than iterative solutions when a problem requires many right hand sides (RHS) (Grayber and Streich, 2012). Our two 3D CSEM optimization proposals (Newton and Halley class methods) require the computation of the second-derivatives of the residuals (eq. 5.12), which are the lowest-order derivatives that require the simulation of internal Green functions $\boldsymbol{G}(\mathbf{r}_n|\mathbf{r}_{n'}, \boldsymbol{\sigma})$. Therefore, due to the large amount of required RHS, we suggest the use of a FDFD direct-solver to obtain the forward solutions.

We use a FDFD modelling tool that uses the MUMPS sparse direct solver to obtain a numerically efficient solution for forward modelling. MUMPS factorizes the system matrix, making use of a multifrontal approach and a MPI standard parallelization (Amestoy et al., 2001, 2006). Once the matrix is factorized, each RHS is obtained in

a few seconds. A similar modelling approach for 3D CSEM Gauss-Newton inversion was considered by Streich (2009).

The storage required for the complete set of Green functions becomes a challenge when using a model of size typical in current commercial applications. We keep in memory only the Green functions $G(\mathbf{r}_n|\mathbf{r}_{\mathtt{tx}}, \boldsymbol{\sigma})$ associated with the source positions and the Green functions $G(\mathbf{r}_{n'}|\mathbf{r}_{\mathtt{rx}}, \boldsymbol{\sigma})$ associated with the receiver positions to reduce the memory requirement. The internal Green functions $G(\mathbf{r}_n|\mathbf{r}_{n'}, \boldsymbol{\sigma})$ (for its three field components) are computed on-the-fly, and considered one at a time. Next we combine this internal Green function with all the Green functions associated with sources and receiver positions, to obtain the contribution in the calculation of the high-order derivative of the residuals. In this way we only buffer an internal Green function (three field components) simultaneously, avoiding the use of the large amount of memory needed in case of storing the whole tensor $G(\mathbf{r}_n|\mathbf{r}_{n'}, \boldsymbol{\sigma})$.

In the methods that we present in this paper, the dimensions of the tensors $\boldsymbol{S}$ and $\boldsymbol{T}$ make their storage impractical for a moderate-size model. As we comment on the introduction, a realistic survey involves $N \sim 10^6$ inversion parameters (Hessian storage $\sim$ TB) and $N_\kappa \sim 10^7$ data samples. The memory requirement scales as,

$$\dim(\boldsymbol{\mathcal{H}}) = N \times N, \tag{5.25}$$

$$\dim(\boldsymbol{S}) = N_\kappa \times N \times N, \tag{5.26}$$

$$\dim(\boldsymbol{T}) = N \times N \times N. \tag{5.27}$$

To cope with the storage challenges following from these scaling behaviours, we do not explicitly compute $\boldsymbol{S}$ and $\boldsymbol{T}$. Rather, we compute the product $d_\kappa^* \mathcal{S}_{\kappa\,n\,n'}$ when obtaining the Hessian matrix $\boldsymbol{\mathcal{H}}$, and the product $\mathcal{T}_{n\,n'\,n''} \Delta\sigma_{n''}^{\mathtt{A}}$ when obtaining a super-Halley update, from eq. 5.6 to eq. 5.8 and from eq. 5.22 to eq. 5.24. In this way we keep the memory complexity in the order of the Hessian matrix.

$$\dim(d_\kappa^* \mathcal{S}_{\kappa\,n\,n'}) = N \times N, \tag{5.28}$$

$$\dim(\mathcal{T}_{n\,n'\,n''} \Delta\sigma_{n''}^{\mathtt{A}}) = N \times N. \tag{5.29}$$

Following these strategies, we can obtain the model update for 3D CSEM inversion for the Newton (eq. 5.21) and Halley class methods (eq. 5.23 and eq. 5.24), with the same memory complexity as in the case of a 3D CSEM Gauss-Newton inversion.

5.2.5 Trust-region solver

In order to solve the linear equation systems eq. 5.21, eq. 5.23 and eq. 5.24, there are two main strategies that can be applied to deal with the fact that matrices $\mathcal{H}$ and $\mathcal{T}$ are

generally indefinite. One option is to apply a Hessian matrix modification, *i.e.* finding at each iteration the smallest matrix $\mathcal{B}^{(k)}$ such that,

$$\mathcal{H}^{(k)} + \mathcal{B}^{\text{A}\,(k)} \qquad\qquad \text{is positive definite,} \qquad (5.30)$$

$$\text{and}$$

$$\mathcal{H}^{(k)} + \mathcal{T}^{(k)}\Delta\boldsymbol{\sigma}^{\text{B}\,(k)} + \mathcal{B}^{\text{B}\,(k)} \qquad\qquad \text{is positive definite.} \qquad (5.31)$$

For a typical 3D CSEM inversion this process can involve a large computational cost to determine the optimal matrix $\mathcal{B}$.

Another alternative to solve eq. 5.21, eq. 5.23, eq. 5.24 is to use one of the inexact Newton methods described in Nocedal and Wright (2006). They are iterative methods with modifications to handle negative curvature in the Hessian matrix. We have implemented the trust-region Newton conjugate gradient (CG) Steihaug algorithm (Steihaug, 1983) because the alternative, based on line-search Newton direction, can require many line-search evaluations or/and giving a small reduction in the misfit, and a trust-region Newton-Lanczos method is computationally too expensive.

At each iteration of an optimization scheme, for solving a linear equation systems (*e.g.* eq. 5.19, eq. 5.21, eq. 5.23 and eq. 5.23) with a trust-region method like CG-Steihaug, a mechanism is necessary to establish the trust-region radius $R^{(k)}$. We consider two alternatives to set the initial trust-region radius $R^{(1)}$. The first one is to set

$$R^{(1)} = \|\Delta\boldsymbol{\sigma}_{\text{GN}}\|_2, \qquad (5.32)$$

i.e. we calculate an initial Gauss-Newton step $\Delta\boldsymbol{\sigma}_{\text{GN}}$ with a regular conjugate gradient solver and then we assign the Euclidean norm of this step to the initial radius of the trust-region algorithm. Another way is to establish the initial trust-region radius based on an initial steepest-descent step $\Delta\boldsymbol{\sigma}_{\text{SD}}$, *i.e.*

$$R^{(1)} = \|\Delta\boldsymbol{\sigma}_{\text{SD}}\|_2. \qquad (5.33)$$

Once the initial trust-region radius $R^{(1)}$ is set, we calculate the first-iteration step $\Delta\boldsymbol{\sigma}^{(1)}$. For the following iterations, the trust-region radius is adjusted according to the following strategy. We define the parameter

$$\psi^{(k)} = \left(\varepsilon(\boldsymbol{\sigma}^{(k)}) - \varepsilon(\boldsymbol{\sigma}^{(k)} + \Delta\boldsymbol{\sigma}^{(k)})\right)\left(m^{(k)}(0) - (m^{(k)}(\Delta\boldsymbol{\sigma}^{(k)})\right)^{-1}, \qquad (5.34)$$

i.e. the ratio between the actual cost function reduction and the predicted cost function reduction. As we show in algorithm 1, the ratio ψ is used to determine when to adjust the trust-region radius, and when to accept the step $\Delta\boldsymbol{\sigma}$. In the case of a Halley class method, at each iteration we use the same trust-region radius to obtain the two Newton steps $\Delta\boldsymbol{\sigma}^{\text{A}}$ (eq. 5.23) and $\Delta\boldsymbol{\sigma}^{\text{B}}$ (eq. 5.24) independently, but adjusting this radius for the following iteration based on the L_2-norm of the resultant step $\Delta\boldsymbol{\sigma}^{\text{H}}$ (eq. 5.22).

Algorithm 1 Trust-region algorithm used in this paper

1: Set the initial radius $R^{(1)}$ based on a initial step: eq. 5.32 or eq. 5.33
2: Given $R_{max} \geq R^{(1)}$ and $\eta \in [0, 1/4]$
3: **for** k=1,2,... **do**
4:     Obtain the iteration step $\boldsymbol{\Delta\sigma}^{(k)}$ by applying the CG-Steihaug solver (Steihaug, 1983) in the corresponding linear equation system: in eq. 5.19 for $\boldsymbol{\Delta\sigma}^{\text{GN}\,(k)}$; or in eq. 5.21 for $\boldsymbol{\Delta\sigma}^{\text{N}\,(k)}$; or in eq. 5.23 and eq. 5.24 for $\boldsymbol{\Delta\sigma}^{\text{H}\,(k)}$
5:     Evaluate $\psi^{(k)}$ from eq. 5.34
6:     **if** $\psi^{(k)} < 1/4$ or (actual reduction < 0 and predicted reduction < 0) **then**
7:         $R^{(k+1)} \leftarrow 1/4\ R^{(k)}$
8:     **else**
9:         **if** $\psi^{(k)} > 3/4$ and $||\boldsymbol{\Delta\sigma}^{(k)}|| = R^{(k)}$ **then**
10:             $R^{(k+1)} \leftarrow min(2R^{(k)}, R_{max})$
11:         **else**
12:             $R^{(k+1)} \leftarrow R^{(k)}$
13:         **end if**
14:     **end if**
15:     **if** $\psi^{(k)} > \eta$ **then**
16:         $\boldsymbol{\sigma}^{(k+1)} \leftarrow \boldsymbol{\sigma}^{(k)} + \boldsymbol{\Delta\sigma}^{(k)}$
17:     **else**
18:         $\boldsymbol{\sigma}^{(k+1)} \leftarrow \boldsymbol{\sigma}^{(k)}$
19:   **end if**
20: **end for**

The trust-region algorithm 1 that we implement is based on the one described in Nocedal and Wright (2006), but with one modification. A cubic approximation like a Halley class method (eq. 5.22) based on a trust-region implementation, could result in an iteration step $\boldsymbol{\Delta\sigma}^{\text{H}}$ from eq. 5.22 that is not a descent direction. This situation can arise while computing the step $\boldsymbol{\Delta\sigma}^{\text{B}}$ from eq. 5.24 if R is too large. Therefore we have extended the algorithm to reduce R when the actual reduction and predicted reduction are simultaneously negative.

5.3 Inversion results

We present 3D CSEM inversion results based on a synthetic survey dataset. These results allow us to evaluate our implementation of the methods that we present in this paper. We compare three different optimization schemes: (A) Gauss-Newton, (B) Newton, and (C) super-Halley. We use a trust-region conjugate gradient method (see subsection 5.2.5) to solve the linear systems for the model updates. For each scheme, we consider the two alternatives eq. 5.32 and eq. 5.33 to set the initial trust-region radius.
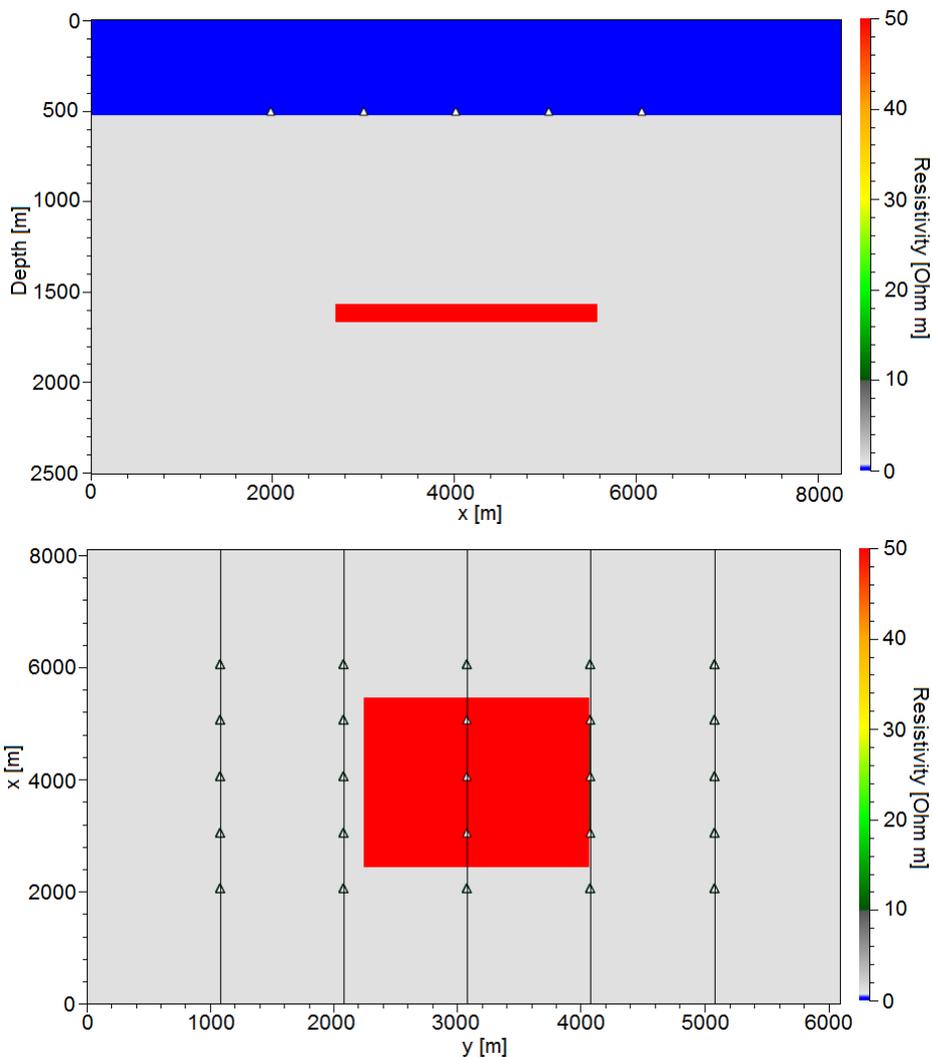
Figure 5.3: True model including the survey layout used in the tests. Data for E_x and E_y at 0.25 Hz and 1.0 Hz is used in the inversion tests. Top: cross-section at $y = 3100$ m. Bottom: depth-slice at $z = 1500$ m.
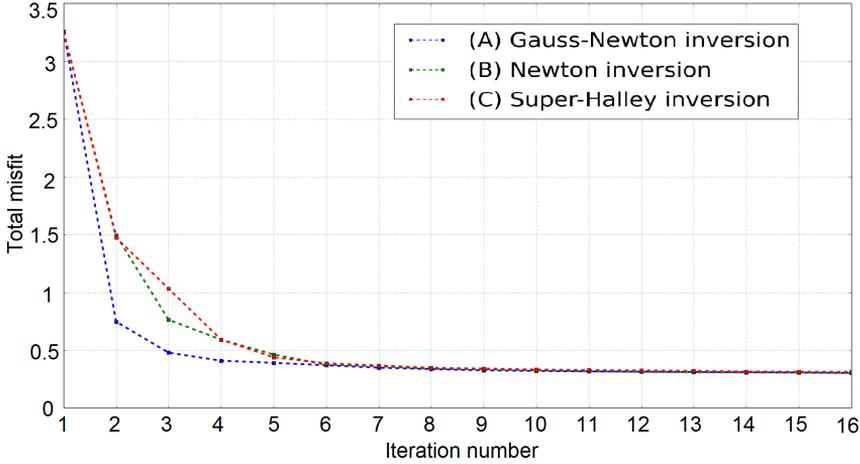
(a) Misfit per iteration for the different schemes when the initial trust-region radius $R^{(1)} = \|\Delta\boldsymbol{\sigma}_{\text{GN}}\|_2$ (L_2-norm of a initial Gauss-Newton step).
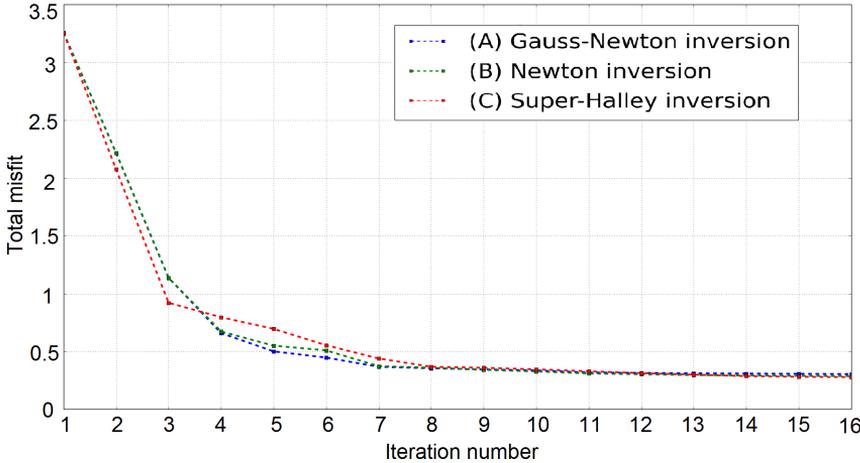
(b) Misfit per iteration for the different schemes when the initial trust-region radius $R^{(1)} = \|\Delta\boldsymbol{\sigma}_{\text{SD}}\|_2$ (L_2-norm of a initial steepest-descent step).

Figure 5.4: Total misfit for the three schemes while using different initial trust-region radius (eq. 5.32 and eq. 5.33).

In figure 5.3 we show the synthetic model used in these tests, and the source-receiver layout. The synthetic survey layout consists of 25 receivers and 5 towlines. The receiver and towline spacing is 1 km. The true model has an homogeneous background resistivity of 1 Ωm, and contains a thin resistor of 50 Ωmat 1000 m below the seabed. The water depth is 500 m, and the water has a resistivity of 0.25 Ωm. The recorded data components are the E_x and E_y fields for the frequencies 0.25 Hz and 1.0 Hz. For these tests we did not contaminate the dataset with noise. The total number of data samples is $N_\kappa = 13368$. For the VTI model we invert for the horizontal

and vertical resistivity components. The model is discretized as a rectilinear grid with cells of size 200 m×200 m×100 m. This gives a total of $N = 45240$ free parameters. In all the inversion results that we show, the initial guess is a half-space model with a background resistivity of 1 Ωmand water resistivity of 0.25 Ωm, *i.e.* corresponding to the background in the true model. We do not show inversion results for the horizontal resistivity since this component does not show significant differences comparing to the initial model. This is due to the well-known fact that the electric field is vertically polarized for the guided-wave produced in a thin resistive layer (see an explanation of this effect in Mittet and Morten (2013)).

In the figure 5.4 we show the total misfit per iteration for the six different tests. The initial total misfit is $\varepsilon^{(1)} = 3.26$. We stopped all the tests at iteration $k = 16$ because, as we show in detail in the table 5.1, the convergence was very slow at that stage.

The same number of iterations is required to reach similar final misfit for all the six tests, see figure 5.4. However, the two different strategies for setting the initial trust-region radius results in a different evolution of the convergence. At the first iterations, the Gauss-Newton scheme in figure 5.4(a) shows a faster convergence than the Newton and the super-Halley schemes. In figure 5.4(b), at early iterations the three schemes show a similar convergence speed, but the super-Halley case is slightly faster in the first two iterations. At iteration 8, all the tests reach a similar low total misfit, but the inversion result models still show important differences when comparing to the true model. At iteration 16 all the inversion result models are close to the true model, but we can observe some minor differences.

In table 5.1 we gather detailed information about the first and last iteration of the different tests. This allows us to analyse the differences observed in figure 5.4. In the table we also show the computed values for the angle between the steepest descent direction and the model update following from each inversion approach,

$$\phi^{(k)} = \texttt{arccos}\left(\|\boldsymbol{\Delta\sigma}^{(k)} \cdot (-\boldsymbol{g}^k)\| \,/\, \left(\|\boldsymbol{\Delta\sigma}^{(k)}\| \cdot \|(-\boldsymbol{g}^k)\|\right)\right) \qquad (5.35)$$

The rest of parameters that appear in each column correspond with the definition previously used in this paper: k = iteration number; $\varepsilon^{(k+1)}$ = total misfit (eq. 5.2) of the updated model $\boldsymbol{\sigma}^{(k+1)} = \boldsymbol{\sigma}^{(k)} + \boldsymbol{\Delta\sigma}^{(k)}$; $\|\boldsymbol{\Delta\sigma}\|_2 = L_2$-norm of the model update.

For the iteration $k = 1$, the data in table 5.0(a) shows that the (A) Gauss-Newton scheme produces a long step $\boldsymbol{\Delta\sigma}^{(1)}$ that reduces the total misfit much more than in the case of the (B) Newton and (C) super-Halley schemes. In the same table and at the same iteration, we observe that although the result of the Newton and super-Halley steps produce a similar total misfit, the angles $\phi^{(k)}$ for both cases are very different. The angle $\phi^{(1)}$ of the scheme (B) step in table 5.0(a) is almost zero since the trust-region algorithm reduces the radius one time before finding a descent direction. We can observe all these differences comparing from figure 5.5 to figure 5.7. The ratio $\psi^{(1)} \approx 1$ indicates that the predicted misfit reduction is quite accurate for the three schemes, at iteration $k = 1$ in table 5.0(a) and table 5.0(b).

(a) Information about the iteration results when using an initial trust-region radius $R^{(1)} = \|\Delta\boldsymbol{\sigma}_{\text{SD}}\|_2$.

scheme	k	$\varepsilon^{(k+1)}$	$\psi^{(k)}$	$\|\Delta\boldsymbol{\sigma}^{(k)}\|_2$	$\phi^{(k)}$
(A)	1	0.76	1.01	18.63	57.4 °
(B)	1	1.49	1.03	4.65	0.07 °
(C)	1	1.47	1.04	6.53	45.5 °
(A)	15	0.31	0.51	0.01	52.7 °
(B)	15	0.31	0.63	0.02	78.24 °
(C)	15	0.31	1.38	0.01	38.7 °

(b) Information about the iteration results when using an initial trust-region radius $R^{(1)} = \|\Delta\boldsymbol{\sigma}_{\text{GN}}\|_2$.

| scheme	k	$\varepsilon^{(k+1)}$	$\psi^{(k)}$	$\|\Delta\boldsymbol{\sigma}^{(k)}\|_2$	$\phi^{(k)}$
(A)	1	2.21	1.04	2.33	0.0 °
(B)	1	2.21	1.04	2.33	0.0 °
(C)	1	2.07	1.07	3.50	41.23 °
(A)	15	0.31	0.87	0.03	67.98 °
(B)	15	0.29	0.32	0.01	0.29 °
(C)	15	0.28	1.18	0.02	59.83 °

Table 5.1: Information about iteration $k = 1$ and iteration $k = 15$ of the three schemes when using different initial trust-region radius (eq. 5.32 and eq. 5.33).
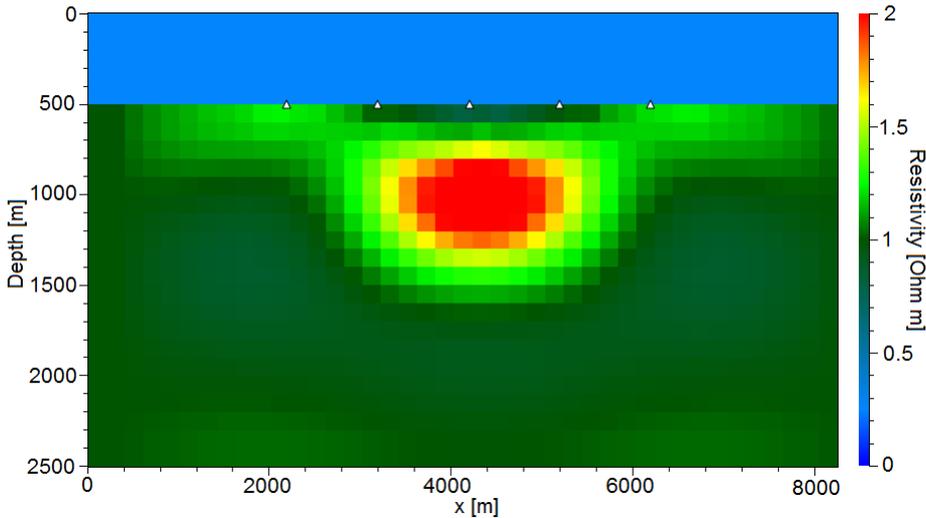
Figure 5.5: Inversion result $\boldsymbol{\sigma}^{(2)}$ of the iteration $k = 1$ and the scheme (A) when using $R^{(1)} = \|\Delta\boldsymbol{\sigma}_{\text{GN}}\|_2$. Section at $y = 3100$ m.
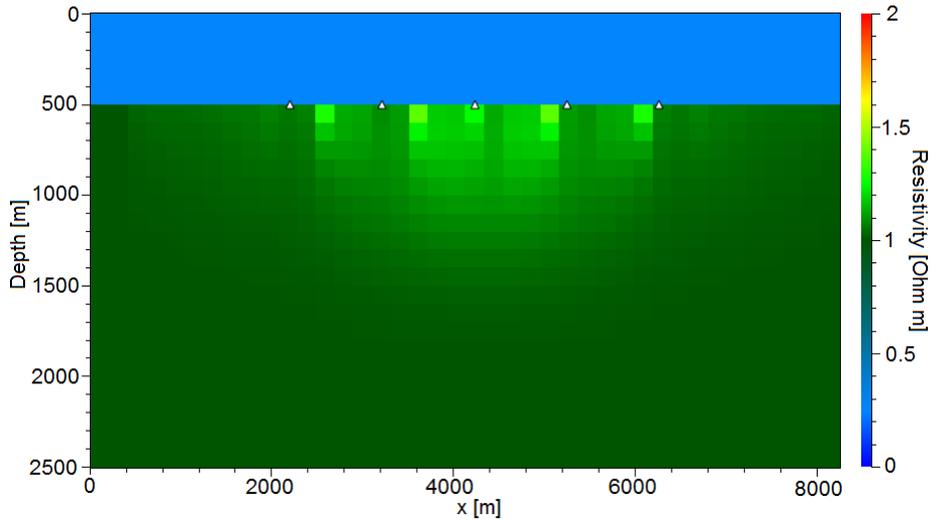
Figure 5.6: Inversion result $\sigma^{(2)}$ of the iteration $k = 1$ and the scheme (B) when using $R^{(1)} = \|\mathbf{\Delta\sigma}_{\text{GN}}\|_2$. Section at $y = 3100$ m.
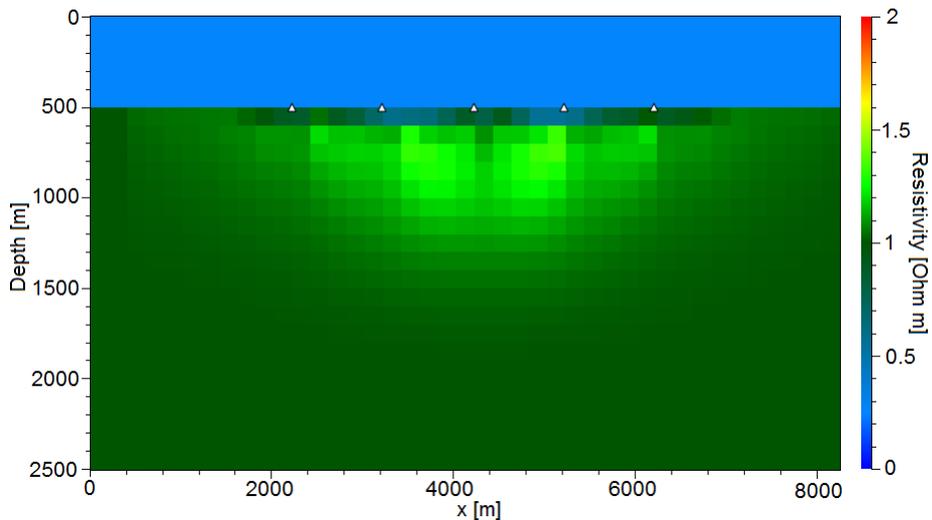

Figure 5.7: Inversion result $\sigma^{(2)}$ of the iteration $k = 1$ and the scheme (C) when using $R^{(1)} = \|\mathbf{\Delta\sigma}_{\text{GN}}\|_2$. Section at $y = 3100$ m.

For the tests based on the use of $R^{(1)} = \|\mathbf{\Delta\sigma}_{\text{SD}}\|_2$ (table 5.0(b)), the initial steps are shorter than in table 5.0(a). In table 5.0(b), scheme (C) at iteration $k = 1$, the step is a bit larger than in scheme (A) and (B), as a result of the two independent super-Halley steps (eq. 5.22), however it does not show a significant difference in the reduction of the total misfit. Only slight differences are observed when comparing the inversion results figure 5.8 and figure 5.9, with figure 5.10.
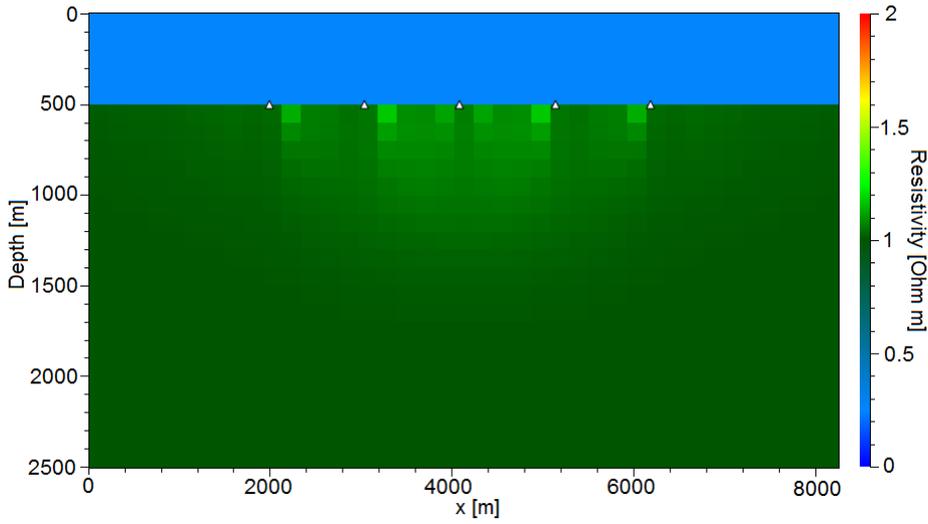
Figure 5.8: Inversion result $\boldsymbol{\sigma}^{(2)}$ of the iteration $k = 1$ and the scheme (A) when using $R^{(1)} = \|\boldsymbol{\Delta\sigma}_{\text{SD}}\|_2$. Section at y = 3100 m.
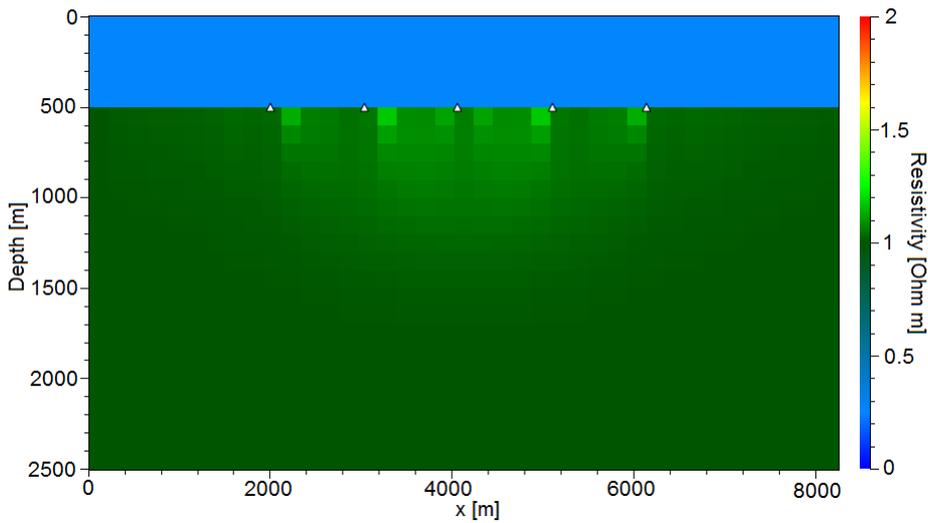
Figure 5.9: Inversion result $\boldsymbol{\sigma}^{(2)}$ of the iteration $k = 1$ and the scheme (B) when using $R^{(1)} = \|\boldsymbol{\Delta\sigma}_{\text{SD}}\|_2$. Section at y = 3100 m.
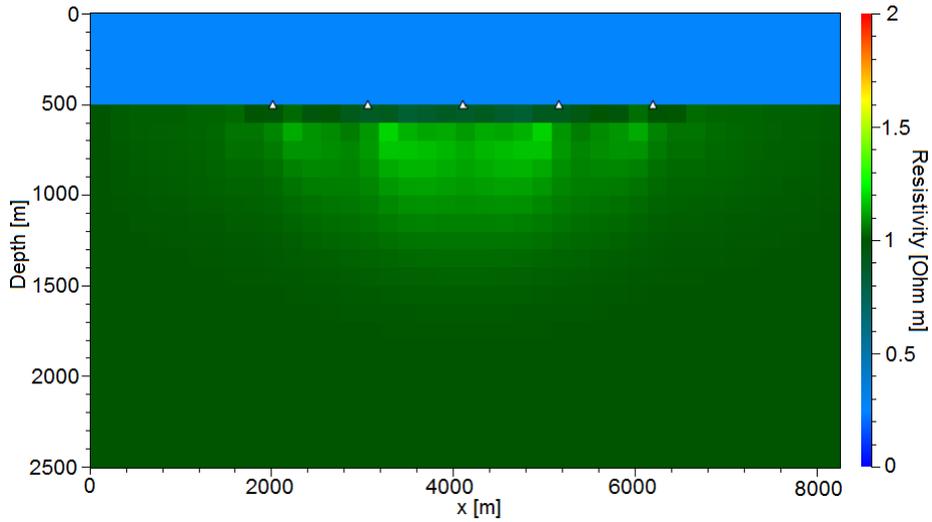
Figure 5.10: Inversion result $\boldsymbol{\sigma}^{(2)}$ of the iteration $k = 1$ and the scheme (C) when using $R^{(1)} = \|\boldsymbol{\Delta\sigma}_{\mathrm{SD}}\|_2$. Section at y = 3100 m.

At iteration $k = 15$, the six tests in table 5.1 produce very small steps $\boldsymbol{\Delta\sigma}^{(15)}$ with different angles $\phi^{(15)}$. Although all of them reach the total misfit $\varepsilon^{(16)} \approx 0.3$, the inversion results (from figure 5.11 to figure 5.16) show some differences between them.
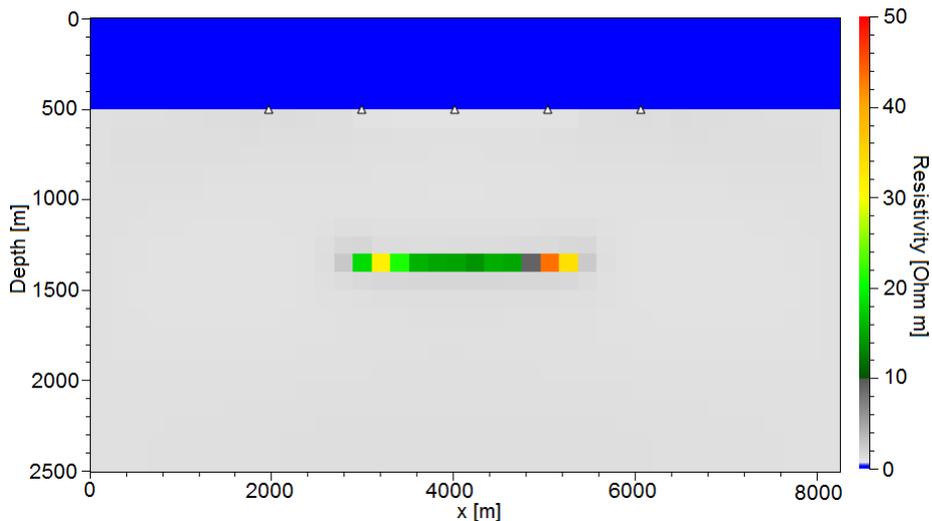

Figure 5.11: Inversion result $\boldsymbol{\sigma}^{(16)}$ of the iteration $k = 15$ and the scheme (A) when using $R^{(1)} = \|\boldsymbol{\Delta\sigma}_{\mathrm{GN}}\|_2$. Section at y = 3100 m.
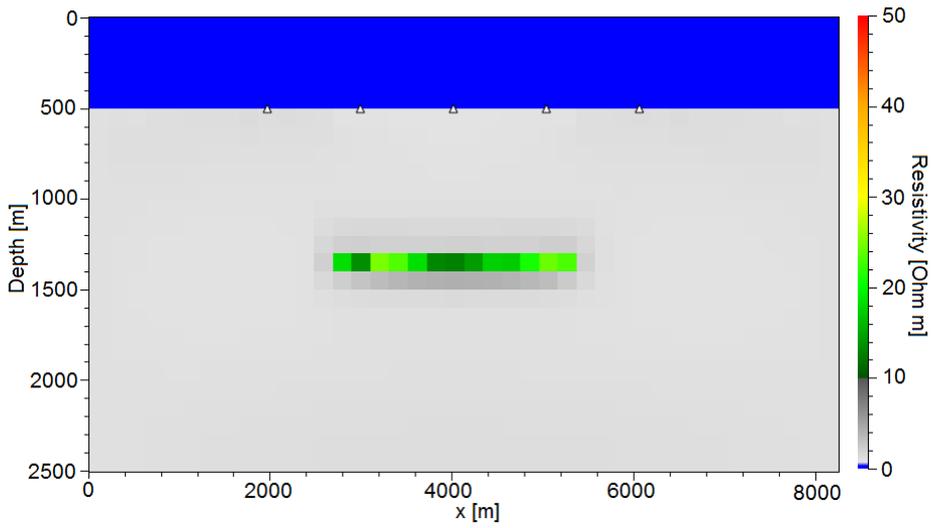
Figure 5.12: Inversion result $\boldsymbol{\sigma}^{(16)}$ of the iteration $k = 15$ and the scheme (B) when using $R^{(1)} = \|\boldsymbol{\Delta\sigma}_{\text{GN}}\|_2$. Section at y = 3100 m.
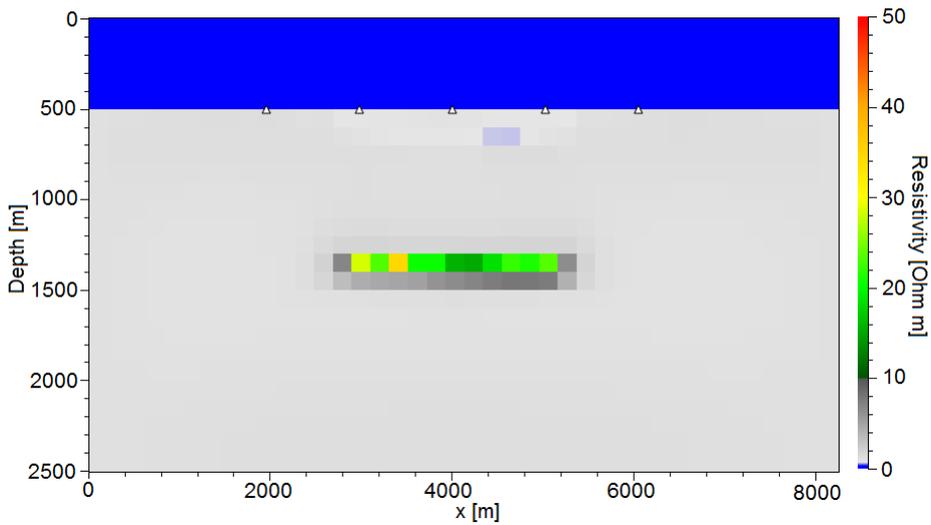
Figure 5.13: Inversion result $\boldsymbol{\sigma}^{(16)}$ of the iteration $k = 15$ and the scheme (C) when using $R^{(1)} = \|\boldsymbol{\Delta\sigma}_{\text{GN}}\|_2$. Section at y = 3100 m.
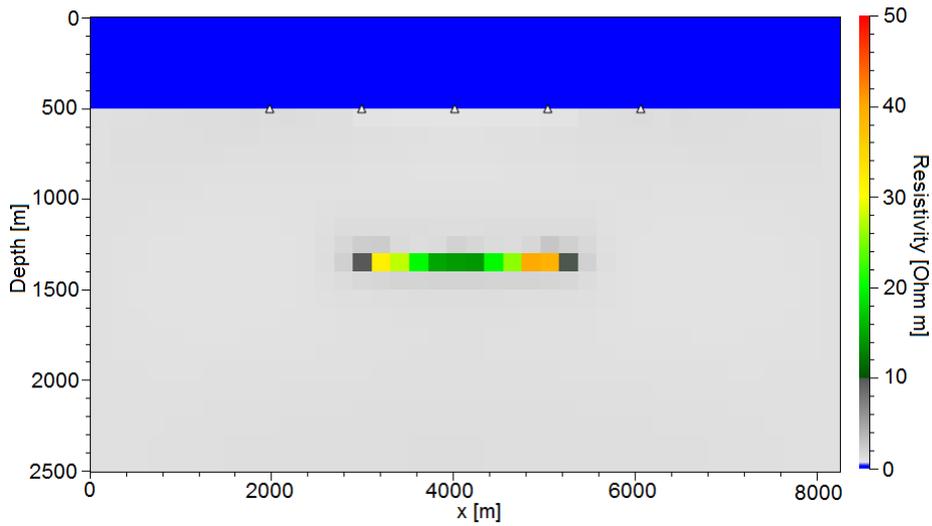
Figure 5.14: Inversion result $\boldsymbol{\sigma}^{(16)}$ of the iteration $k = 15$ and the scheme (A) when using $R^{(1)} = \|\boldsymbol{\Delta\sigma}_{\text{SD}}\|_2$. Slice at y = 3100 m.

The inversion results from figure 5.11 to figure 5.14 place the resistor 200 m above its position in the true model. In these results, the maximum resistivity is 40 Ωm. Since at each iteration the regularization (smoothness) contribution is reduced, at iteration $k = 15$ the weak regularization produces irregularities in the target resistivity. In figure 5.15 and figure 5.16 the resistor is situated 100 m above its position in the true model and the maximum resistivity reaches 50 Ωm.
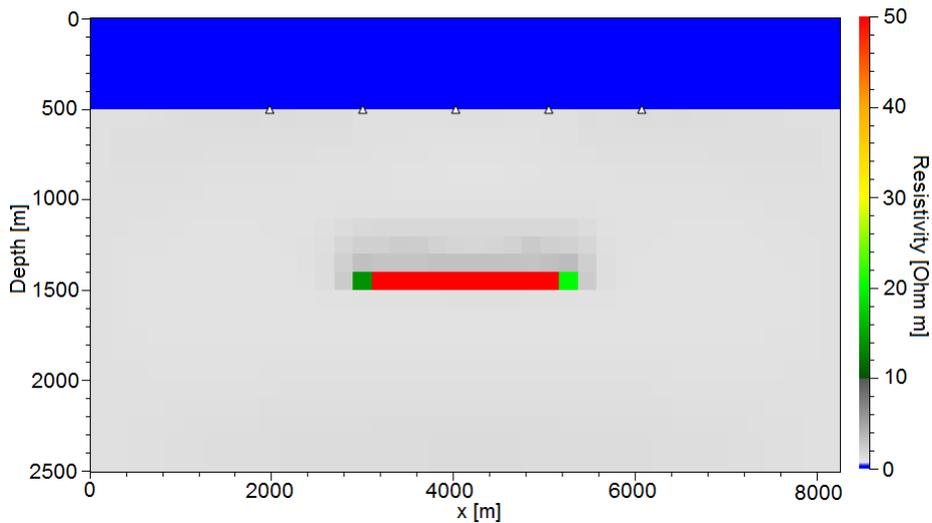
Figure 5.15: Inversion result $\boldsymbol{\sigma}^{(16)}$ of the iteration $k = 15$ and the scheme (B) when using $R^{(1)} = \|\boldsymbol{\Delta\sigma}_{\text{SD}}\|_2$. Slice at y = 3100 m.
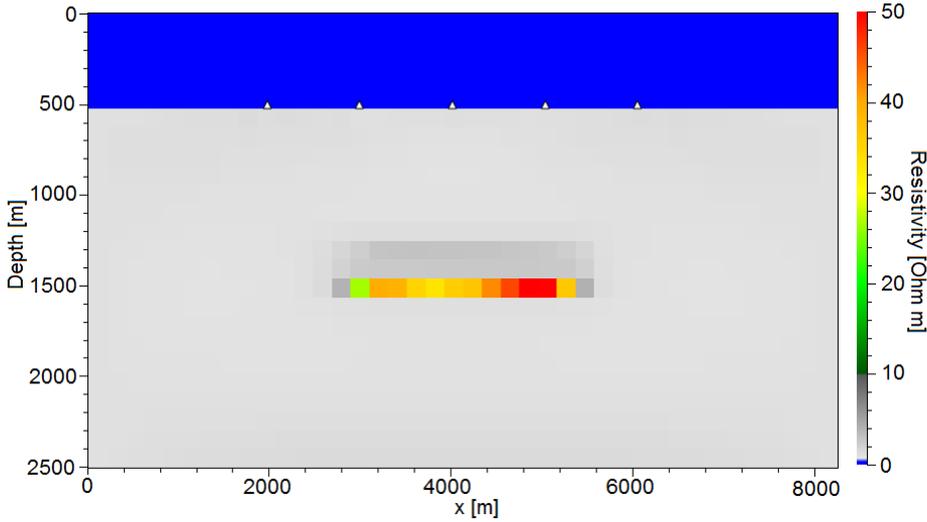
Figure 5.16: Inversion result $\sigma^{(16)}$ of the iteration $k = 15$ and the scheme (C) when using $R^{(1)} = \|\Delta\sigma_{\mathrm{SD}}\|_2$. Slice at y = 3100 m.

In order to compute these results, each test used 8 MPI-processes per frequency in a single isolated-cluster node, with a maximum total (sum of all the processes) RAM memory of 60 GB per frequency. The FDFD forward modelling tool has the largest memory usage. The average computing time needed by each test was: test (A) used 2.5 hours per iteration; test (B) used 13 hours per iteration; test(C) used 23 hours per iteration.

5.4 Discussion

In this section we compare the results of the 3D CSEM Newton and super-Halley schemes, to the results of 3D CSEM Gauss-Newton scheme.

In the six tests, the different angles $\phi^{(1)}$, the different L_2-norms of the initial steps $\|\Delta\sigma^{(1)}\|_2$, and the evolution of their total misfits, suggest that each method drives the inversion through different paths. Indeed, although the last inversion results $\sigma^{(16)}$ have a similar total misfit, we can observe differences when comparing them.

The implementation of the trust-region approach does not guarantee the optimal set of steps in the constrained minimization problem. For the six tests presented here, the differences in the convergence speed at the initial iterations are not sufficient to conclusively evaluate the initial trust-region radius schemes described above. Although the two approaches to set the initial trust-region radius apparently change the behaviour of the inversion, we think that the initial misfit reduction is too low for the tests that use $R^{(1)} = \|\Delta\sigma_{\mathrm{SD}}\|_2$.

The main difference between the results at iteration $k = 15$ and the true model is the depth placement of the resistor. We note that for the low frequencies considered,

the smallest propagation length-scale is the skin-depth which is approximately 500 m for the highest frequency. Variations in depth placement that are significantly smaller than this length are considered below the resolution of the data. However, it is interesting to see that there are two inversion results that place the resistor closer to its true position.

When evaluating integrals over the internal Green functions for which the unit source is contained within the integration volume, we approximate the integral by using the solution for $\mathbf{G}(\mathbf{r}_n|\mathbf{r}_{n'},\sigma)$ obtained by the FDFD method. Doing this we are approximating an integral (over a singular function) by an average value of the integrand over the cell volume. In higher order approaches like Newton and Halley class methods, these approximations affect all the values of the diagonal (and its vicinity) of the Hessian matrix $\mathcal{H}$ and the tensor $\mathcal{T}$. In the presented tests we do not observe incorrect resistors or non-converging inversions that might be produced by such approximations. It may be possible to obtain a better approximation to $\mathbf{G}(\mathbf{r}_n|\mathbf{r}_{n'},\sigma)$, but that is beyond the scope of our paper.

5.5   Conclusions

In this paper we introduce the use of the high-order optimization schemes Newton and Halley-class methods for 3D CSEM inversion. These schemes represent the next step in complexity over Gauss-Newton optimization, which is currently becoming an established methodology for 3D CSEM inversion. Our results contribute to the understanding of the numerical complexity and performance of high-order methods compared to current inverse schemes.

The use of a FDFD direct solver, for performing the tens of thousands of forward simulations per iteration, and on-the-fly computations, allow to implement 3D CSEM higher order optimization methods than Gauss-Newton due to the resultant limited computational cost. We show that it is possible to perform these computations keeping the memory complexity low. Once the system matrix is factorized using the direct solver, it is possible to explore high-order methods without further time-consuming modeling operations. Our diagrammatic representations show how the higher order methods can all be developed from internal propagations, *i.e.* Green functions with both spatial indices inside the parameterized domain.

A trust-region solver that can handle the negative curvature of the system matrices, avoids the computational cost associated with finding an optimal stabilizer. Our results suggest the use of a more sophisticated trust-region algorithm that could modify the trust-region radius following a more efficient pattern.

The results that we present show that the three different schemes (Gauss-Newton, Newton and super-Halley) drive the inversion through different paths, making it difficult to compare the convergence speed of these methods.

In future work we would consider more complex conductivity models, that involve a more complex cost function with local minima corresponding to qualitatively

different models.

5.6   Acknowledgments

5.7   Appendix

In this appendix, we consider the derivatives of the cost function and the synthetic data, with respect to model parameters $\sigma(\mathbf{r})$. These derivatives follow from the variation of the fields when we vary the parameters,

$$
\begin{aligned}
F_\kappa^{\texttt{Synth}}(\sigma + \delta\sigma) = {} & F_\kappa^{\texttt{Synth}}(\sigma) \\
& + \sum_n \frac{1}{n!} \int \prod_{m=1}^n \mathrm{d}\mathbf{r}_m \, \frac{\partial^n F_\kappa^{\texttt{Synth}}(\sigma)}{\partial\sigma(\mathbf{r}_1)\cdots\partial\sigma(\mathbf{r}_n)} \delta\sigma(\mathbf{r}_1)\cdots\delta\sigma(\mathbf{r}_n).
\end{aligned} \quad (5.36)
$$

In this expression, we defined the derivatives $\partial^n F_\kappa^{\texttt{Synth}}/\partial\sigma(\mathbf{r}_1)\cdots\partial\sigma(\mathbf{r}_n)$ of the expansion of the fields in the parameters, and the variation in the parameters $\delta\sigma(\mathbf{r})$. We will give a simplified derivation of these derivatives, focusing first on the development, and subsequently introducing more details. Other derivations of the parameter derivatives pertaining to geophysics applications in particular can be found in *e.g.* McGillivray et al. (1994) and Plessix (2006).

The physical equations that $F_\kappa^{\texttt{Synth}}$ satisfies, are the Maxwell equations. The set of first-order partial differential equations for the electric and magnetic fields can be combined as a second-order equation for the electric field, *i.e.* we consider that $F^{\texttt{Synth}} = E$ in this section. In the frequency domain, and utilizing the quasi-static approximation, this equation becomes,

$$
\frac{i}{\omega\mu} \nabla \times \nabla \times \mathbf{E}(\mathbf{r}, \omega) - \sigma(\mathbf{r})\mathbf{E}(\mathbf{r}, \omega) = \mathbf{J}(\mathbf{r}, \omega). \quad (5.37)
$$

Here $\omega = 2\pi f$ is the angular frequency, μ is the magnetic permeability assumed constant, $\mathbf{E}$ is the electric field, σ is the conductivity tensor, and $\mathbf{J}$ are the electric current sources. A similar equation can be derived for the magnetic field, but since we only consider electric field responses in this paper we have omitted that derivation. Note however, that the subsequent steps follow analogously.

We represent the linear operator associated with the inhomogeneous partial differential eq. 5.37 depending on the conductivity as $\mathbf{L_r}(\sigma, \omega)$. The subscript $\mathbf{r}$ will

denote the spatial argument of the operator. Using this definition, and suppressing the angular frequency argument, the inhomogeneous equation to solve can be written

$$\mathbf{L_r}(\sigma)\mathbf{E}(\mathbf{r}|\sigma) = \mathbf{J}(\mathbf{r}), \tag{5.38}$$

where we have added an explicit label for the conductivity model on the electric field.

Let us consider the effect of a perturbation of the conductivity model, $\delta\sigma$, at position $\mathbf{p}$. The operator can be expanded as

$$\mathbf{L_r}\left(\sigma + \delta\sigma(\mathbf{p})\right) = \mathbf{L_r}(\sigma) - \delta\sigma\,\delta(\mathbf{r} - \mathbf{p}). \tag{5.39}$$

We also expand the electric field in the perturbed model,

$$\mathbf{E}(\mathbf{r}|\sigma + \delta\sigma(\mathbf{p})) \approx \mathbf{E}(\mathbf{r}|\sigma) + \delta\sigma\,\frac{\partial\mathbf{E}(\mathbf{r}|\sigma)}{\partial\sigma(\mathbf{p})}. \tag{5.40}$$

We now introduce the Green function tensor for the electric field at position $\mathbf{r}$ in direction i, due to a unit electrical current source at position $\mathbf{r}'$ with spatial direction j, $G_{ij}(\mathbf{r}|\mathbf{r}', \sigma)$. This quantity is associated with the operator $\mathbf{L_r}(\sigma)$, and defined as the solution of the equation

$$\mathbf{L_r}(\sigma)\mathbf{G}(\mathbf{r}|\mathbf{r}', \sigma) = \mathbf{I}\,\delta(\mathbf{r} - \mathbf{r}'), \tag{5.41}$$

where $\mathbf{I}$ is a 3×3 unit matrix. In the above equation, we have reverted to the vector notation for the tensor components i, j. The Green function allows us to express the solution to eq. 5.38 as

$$\mathbf{E}(\mathbf{r}|\sigma) = \int d\mathbf{r}'\mathbf{G}(\mathbf{r}|\mathbf{r}', \sigma)\,\mathbf{J}(\mathbf{r}'). \tag{5.42}$$

The eq. 5.38 for the electric field in the perturbed model can be written,

$$[\mathbf{L_r}(\sigma) - \delta\sigma\,\delta\,(\mathbf{r} - \mathbf{p})]\left(\mathbf{E}(\mathbf{r}|\sigma) + \delta\sigma\,\frac{\partial\mathbf{E}(\mathbf{r}|\sigma)}{\partial\sigma(\mathbf{p})}\right) = \mathbf{J}(\mathbf{r}). \tag{5.43}$$

To first order in $\delta\sigma$, and using the identity in eq. 5.38, we get

$$\mathbf{L_r}(\sigma)\frac{\partial\mathbf{E}(\mathbf{r}|\sigma)}{\partial\sigma(\mathbf{p})} = \delta\,(\mathbf{r} - \mathbf{p})\,\mathbf{E}(\mathbf{r}|\sigma). \tag{5.44}$$

This equation is solved using the Green function, and we also substitute the result in eq. 5.42 to obtain,

$$\frac{\partial\mathbf{E}(\mathbf{r}|\sigma)}{\partial\sigma(\mathbf{p})} = \int d\mathbf{r}'\,\mathbf{G}(\mathbf{r}|\mathbf{p}, \sigma)\,\mathbf{G}(\mathbf{p}|\mathbf{r}', \sigma)\,\mathbf{J}(\mathbf{r}'). \tag{5.45}$$

This is the desired result for the derivative. If we can approximate the source as a unit dipole point-source at $\mathbf{r_{tx}}$, the integration of the coordinate $\mathbf{r}'$ simplifies as

$\int d\mathbf{r}' f(\mathbf{r}') J_l(\mathbf{r}') \approx f(\mathbf{r}_{tx}) \hat{J}_l(\mathbf{r}_{tx})$ with $\hat{J}_l$ designating the source components with dipole moment $\left|\hat{\mathbf{J}}\right| = 1$ Am. This approximation is generally good for CSEM data when the separation between the source and the receiver is at least a wavelength, and the source length is small compared to the wavelength. Note that the numerical results in this paper were computed with a line-source approximation rather than the approximation described above.

We can go on to derive an expression for the second derivative by considering a perturbation at position $\mathbf{q}$ and the expansion,

$$\frac{\partial \mathbf{E}(\mathbf{r}|\sigma + \delta\sigma(\mathbf{q}))}{\partial\sigma(\mathbf{p})} \approx \frac{\partial \mathbf{E}(\mathbf{r}|\sigma)}{\partial\sigma(\mathbf{p})} + \delta\sigma \frac{\partial^2 \mathbf{E}(\mathbf{r}|\sigma)}{\partial\sigma(\mathbf{p})\partial\sigma(\mathbf{q})}. \tag{5.46}$$

Following the same procedure as above, we construct an equation for this quantity in the perturbed model based on eq. 5.44,

$$\mathbf{L_r}\left(\sigma + \delta\sigma\left(\mathbf{q}\right)\right) \frac{\partial \mathbf{E}(\mathbf{r}|\sigma + \delta\sigma(\mathbf{q}))}{\partial\sigma(\mathbf{p})} = \delta\left(\mathbf{r} - \mathbf{p}\right) \mathbf{E}\left(\mathbf{r}|\sigma + \delta\sigma\left(\mathbf{q}\right)\right). \tag{5.47}$$

Inserting the expansions for the quantities in the perturbed model, we find to the same accuracy used above that

$$\mathbf{L_r}(\sigma) \frac{\partial^2 \mathbf{E}(\mathbf{r}|\sigma)}{\partial\sigma(\mathbf{p})\partial\sigma(\mathbf{q})} = \delta(\mathbf{r} - \mathbf{q}) \frac{\partial \mathbf{E}(\mathbf{r}|\sigma)}{\partial\sigma(\mathbf{p})} + \delta(\mathbf{r} - \mathbf{p}) \frac{\partial \mathbf{E}(\mathbf{r}|\sigma)}{\partial\sigma(\mathbf{q})}. \tag{5.48}$$

We can finally use the Green function and the expression in eq. 5.45 to express the second derivative as,

$$\frac{\partial^2 \mathbf{E}(\mathbf{r}|\sigma)}{\partial\sigma(\mathbf{p})\partial\sigma(\mathbf{q})} = \int d\mathbf{r}' \, \mathbf{G}(\mathbf{r}|\mathbf{q}, \sigma) \, \mathbf{G}(\mathbf{q}|\mathbf{p}, \sigma) \, \mathbf{G}(\mathbf{p}|\mathbf{r}', \sigma) \, \mathbf{J}(\mathbf{r}')$$
$$+ \int d\mathbf{r}' \, \mathbf{G}(\mathbf{r}|\mathbf{p}, \sigma) \, \mathbf{G}(\mathbf{p}|\mathbf{q}, \sigma) \, \mathbf{G}(\mathbf{q}|\mathbf{r}', \sigma) \, \mathbf{J}(\mathbf{r}'). \tag{5.49}$$

This is the expression for the second derivative.

Expressions for higher order derivatives are obtained in the same way. For example, the third derivative becomes,

$$\frac{\partial^3 \mathbf{E}(\mathbf{r}|\sigma)}{\partial\sigma(\mathbf{p}_1)\partial\sigma(\mathbf{p}_2)\partial\sigma(\mathbf{p}_3)} = \sum_{(1\,2\,3)} \int d\mathbf{r}' \mathbf{G}(\mathbf{r}|\mathbf{p}_3, \sigma) \, \mathbf{G}(\mathbf{p}_3|\mathbf{p}_2, \sigma)$$
$$\times \mathbf{G}(\mathbf{p}_2|\mathbf{p}_1, \sigma) \, \mathbf{G}(\mathbf{p}_1|\mathbf{r}', \sigma) \, \mathbf{J}(\mathbf{r}'), \tag{5.50}$$

where the notation $(1\,2\,3)$ denotes that we should include all permutations of the indices in the sum. The generalization to arbitrary order derivatives follow the same pattern, with the number of "internal nodes", above denoted $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ for third order, corresponding to the order of the derivative. Such generalization accommodates

the diagrammatic technique described in subsection 5.2.2 to construct arbitrary order derivatives.

In order to make contact with the notation used in the main text, we will now write down the expression above for the first derivative in components. When we consider the Green function for general field components $F = E, H$ due to an electric source, $\mathbf{J}$, or a magnetic source, $\mathbf{K}$, we use the notation $G_{i,j}^{F,U}$ for $U = J, K$. Moreover, we consider a discrete approximation, where parameter variations are defined for finite regions of volume $\Delta V(\mathbf{r}_k)$. The coordinate $\mathbf{r}_k$ is considered at the region center point. The integration over the variations in the definition in eq. 5.36 is then approximated $\int \mathrm{d}\mathbf{r} f(\mathbf{r}) \delta\sigma(\mathbf{r}) \approx \sum_k \Delta V(\mathbf{r}_k) f(\mathbf{r}_k) \Delta\sigma(\mathbf{r}_k)$ where the sum is over the discretization. We also generalize the result in eq. 5.45 to consider tensor structure on the parameter, $\sigma \to \sigma_{j,k}$. If we consider a VTI medium, the perturbation becomes $\delta\sigma \to \delta\sigma_{j,k} = \delta_{j,k} [\delta\sigma_\mathtt{H} (\delta_{j,x} + \delta_{j,y}) + \delta\sigma_\mathtt{V}\delta_{j,z}]$, with independent horizontal, $\delta\sigma_\mathtt{H}$, and vertical, $\delta\sigma_\mathtt{V}$, components. Introducing this tensor structure in eq. 5.43 leads to a component projection of the Green functions. The expression for the first derivative then becomes,

$$\frac{\partial F_i(\mathbf{r}|\mathbf{r}_\mathtt{tx}, f, \sigma)}{\partial\sigma_\mathtt{H}(\mathbf{p})} = \Delta V(\mathbf{p}) \sum_{j=x,y} G_{ij}^{FJ}(\mathbf{r}|\mathbf{p}, f, \sigma) \, G_{jl}^{EJ}(\mathbf{p}|\mathbf{r}_\mathtt{tx}, f, \sigma) \, \hat{J}_l(\mathbf{r}_\mathtt{tx}, f),$$
(5.51)

$$\frac{\partial F_i(\mathbf{r}|\mathbf{r}_\mathtt{tx}, f, \sigma)}{\partial\sigma_\mathtt{V}(\mathbf{p})} = \Delta V(\mathbf{p}) G_{iz}^{FJ}(\mathbf{r}|\mathbf{p}, f, \sigma) \, G_{zl}^{EJ}(\mathbf{p}|\mathbf{r}_\mathtt{tx}, f, \sigma) \, \hat{J}_l(\mathbf{r}_\mathtt{tx}, f).$$
(5.52)

Above we reinstated the frequency argument f. To simplify the computational load in numerical evaluations of the derivatives, the reciprocity principle can be applied to compute the Green function connecting $\mathbf{r}$ and $\mathbf{p}$ where $\mathbf{r}$ corresponds to fixed receiver positions,

$$G_{ij}^{FJ}(\mathbf{r}|\mathbf{p}, f, \sigma) = \mathtt{sign}(F) \, G_{ji}^{EU}(\mathbf{p}|\mathbf{r}, f, \sigma).$$
(5.53)

Here the index $U = J$ when $F = E$, and $U = K$ when $F = H$.

The component representation for the higher order derivatives is obtained following the same generalizations. For example, the mixed second derivative for $\sigma_\mathtt{H}$ and $\sigma_\mathtt{V}$ from eq. 5.49 becomes,

$$\frac{\partial^2 F_i(\mathbf{r}|\mathbf{r}_\mathtt{tx}, f, \sigma)}{\partial\sigma_\mathtt{H}(\mathbf{p})\partial\sigma_\mathtt{V}(\mathbf{q})} = \Delta V(\mathbf{p})\Delta V(\mathbf{q}) \sum_{j=x,y}$$

$$\times \Big[G_{iz}^{FJ}(\mathbf{r}|\mathbf{q}, f, \sigma) \, G_{zj}^{EJ}(\mathbf{q}|\mathbf{p}, f, \sigma) \, G_{jl}^{EJ}(\mathbf{p}|\mathbf{r}_\mathtt{tx}, f, \sigma)\hat{J}_l(\mathbf{r}_\mathtt{tx}, f)$$

$$+ G_{ij}^{FJ}(\mathbf{r}|\mathbf{p}, f, \sigma) \, G_{jz}^{EJ}(\mathbf{p}|\mathbf{q}, f, \sigma) \, G_{zl}^{EJ}(\mathbf{q}|\mathbf{r}_\mathtt{tx}, f, \sigma)\hat{J}_l(\mathbf{r}_\mathtt{tx}, f) \Big].$$
(5.54)

Again, the reciprocity principle can be used to reduce computational load in numerical evaluation of this derivative when the number of receiver positions $\mathbf{r}$ is small.

5.8 References

Abubakar, A., Habashy, T. M., Li, M., and Liu, J. Inversion algorithms for large-scale geophysical electromagnetic measurements. *IOP Science Inverse Problems*, 25(12):123012, 2009. doi: 10.1088/0266-5611/25/12/123012. URL http://dx.doi.org/10.1088/0266-5611/25/12/123012.

Amaya, M., Morten, J. P., and Boman, L. Efficient computation of approximate low-rank hessian for 3D CSEM inversion. In *SEG Technical Program Expanded Abstracts*, pages 722–726, 2014. doi: 10.1190/segam2014-0493.1. URL http://doi.org/10.1190/segam2014-0493.1.

Amestoy, P. R., Duff, I., L'Excellent, J.-Y., and Koster, J. A fully asynchronous multifrontal solver using distributed dynamic scheduling. *SIAM Journal on Matrix Analysis and Applications*, 23(1):15–41, 2001. doi: 10.1137/S0895479899358194. URL http://dx.doi.org/10.1137/S0895479899358194.

Amestoy, P. R., Guermouche, A., L'Excellent, J.-Y., and Pralet, S. Hybrid scheduling for the parallel solution of linear systems. *Parallel Computing*, 32(2):136–156, 2006. doi: 10.1016/j.parco.2005.07.004. URL http://dx.doi.org/10.1016/j.parco.2005.07.004.

Constable, S. and Srnka, L. An introduction to marine controlled-source electromagnetic methods for hydrocarbon exploration. *Geophysics*, 72(2):WA3–WA12, 2007. doi: 10.1190/1.2432483. URL http://dx.doi.org/10.1190/1.2432483.

Da Silva, N. V., Morgan, J. V., Macgregor, L., and Warner, M. A finite element multifrontal method for 3D CSEM modeling in the frequency domain. *Geophysics*, 77(2):E101–E115, 2012. doi: 10.1190/geo2010-0398.1. URL http://dx.doi.org/10.1190/geo2010-0398.1.

Eidesmo, T., Ellingsrud, T., MacGregor, L., Constable, S., Sinha, M., Johansen, S. E., Kong, F., and Westerdahl, H. Sea bed logging (sbl), a new method for remote and direct identification of hydrocarbon filled layers in deepwater areas. *First Break*, 20(3):144–152, 2002. URL http://fb.eage.org/publication/content?id=25008.

Ezquerro, J. A. and Hernández, M. A. A new class of third-order methods in banach spaces. *Bulletin of the Institute of Mathematics. Academia Sinica*, 31(1):33–58, 2003. ISSN 0304-9825. URL http://folk.uib.no/ssu029/Pdf_file/Ezquerro03.pdf.

Fanavoll, S., Gabrielsen, P., and Ellingsrud, S. The impact of CSEM on exploration decisions and seismic: two case studies from the barents sea. *First Break*, 32(11): 105–110, 2014. doi: 10.1190/INT-2013-0171.1. URL http://dx.doi.org/10.1190/INT-2013-0171.1.

Grayber, A. and Streich, R. Comparison of iterative and direct solvers for 3D CSEM modeling. In *SEG Technical Program Expanded Abstracts*, pages 1–6, 2012. doi: 10.1190/segam2012-0727.1. URL http://dx.doi.org/10.1190/segam2012-0727.1.

Grayver, A. V., Streich, R., and Ritter, O. Three-dimensional parallel distributed inversion of CSEM using a direct forward solver. *Geophysics Journal International*, 193(3):1432–1446, 2013. doi: 10.1093/gji/ggt055. URL http://dx.doi.org/10.1093/gji/ggt055.

Gundersen, G. and Steihaug, T. On large-scale scale unconstrained optimization problems and higher order methods. *Optimization Methods and Software*, 35(3): 337–358, 2011. doi: 10.1080/10556780903239071. URL http://doi.org/10.1080/10556780903239071.

Gutierrez, J. M. and Hernandez, M. A. An acceleration of Newton's method: Super-halley method. *Applied Mathematics and Computation*, 117(2-3):223–239, 2001. doi: 10.1016/S0096-3003(99)00175-7. URL http://dx.doi.org/10.1016/S0096-3003(99)00175-7.

Habashy, T. and Abubakar, A. A general framework for constraint minimization for the inversion of electromagnetic measurements. *Progress in Electromagnetics Research*, 46:265–312, 2004. doi: 10.2528/PIER03100702. URL http://dx.doi.org/10.2528/PIER03100702.

Han, D. The convergence on a family of iterations with cubic order. *Journal of Computational Mathematics*, 19(5):467–474, 2001.

Hoversten, G., Myer, D., Key, K., Hermann, O., Hobbet, R., and Alumbaugh, D. CSEM & MMT base basalt imaging. In *75th EAGE Conference & Exhibition incorporating SPEEUROPEC 2013*, 2013. doi: 10.3997/2214-4609.20130137. URL http://doi.org/10.3997/2214-4609.20130137.

Li, M., Abubakar, A., Liu, J., Pan, G., and Habashy, T. M. Three-dimensional regularized gauss-newton inversion algorithm using a compressed implicit jacobian calculation for electromagnetic applications. In *SEG Technical Program Expanded Abstracts*, pages 593–598, 2010. doi: 10.1190/1.3513853. URL http://dx.doi.org/10.1190/1.3513853.

Lin, Y., Li, M., Abubakar, A., and Habashy, T. A wavelet-based model compression method for three-dimensional electromagnetic data inversion. In *SEG Technical Program Expanded Abstracts*, pages 707–712, 2013. doi: 10.1190/segam2013-0395.1. URL http://dx.doi.org/10.1190/segam2013-0395.1.

McGillivray, P. R., Oldenburg, D. W., Ellis, R. G., and Habashy, T. M. Calculation of sensitivities for the frequency-domain electromagnetic problem. *Geophysical Journal International*, 1:1–4, 1994. doi: 10.1111/j.1365-246X.1994.tb02121.x. URL http://dx.doi.org/10.1111/j.1365-246X.1994.tb02121.x.

Mittet, R. and Morten, J. P. The marine controlled-source electromagnetic method in shallow water. *Geophysics*, 78(2):E67–E77, 2013. doi: 10.1190/GEO2012-0112.1. URL http://dx.doi.org/10.1190/GEO2012-0112.1.

Morten, J. P., Roth, F., Timko, D., Pacurar, C., Nguyen, A. K., and Olsen, P. A. 3D reservoir characterization of a North Sea oil field using quantitative seismic & CSEM interpretation. In *SEG Technical Program Expanded Abstracts 2011*, pages 1903–1907, 2011. doi: 10.1190/1.3627578. URL http://library.seg.org/doi/abs/10.1190/1.3627578.

Nocedal, J. and Wright, S. J. *Numerical optimization*. Springer, 2nd edition, 2006. ISBN 978-0387-30303-1. URL http://www.springer.com/us/book/9780387303031.

Plessix, R.-. and Mulder, W. A. Resistivity imaging with controlled-source electromagnetic data: depth and data weighting. *IOP Science Inverse Problems*, 24(3):034012, 2008. doi: 10.1088/0266-5611/24/3/034012. URL http://dx.doi.org/10.1088/0266-5611/24/3/034012.

Plessix, R. E. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167:495–503, 2006. doi: 10.1111/j.1365-246X.2006.02978.x. URL http://dx.doi.org/10.1111/j.1365-246X.2006.02978.x.

Portniaguine, O. and Zhdanov, S. Focusing geophysical inversion images. *Geophysics*, 64(3):874–887, 1999. doi: 10.1190/1.1444596. URL http://dx.doi.org/10.1190/1.1444596.

Schwarzbach, C. and Haber, E. Finite element based inversion for time-harmonic electromagnetic problems. *Geophysics Journal International*, 193(2):615–634, 2013. doi: 10.1093/gji/ggt006. URL http://dx.doi.org/10.1093/gji/ggt006.

Steihaug, T. The conjugate gradient method and trust regions in large scale optimization. *SIAM Journal of Numerical Analysis*, 20:626–637, 1983. doi: 10.1137/0720042. URL http://dx.doi.org/10.1137/0720042.

Streich, R. 3D finite-difference frequency-domain modeling of controlled-source electromagnetic data: Direct solution and optimization for high accuracy. *Geophysics*, 74(5):F95–F105, 2009. doi: 10.1190/1.3196241. URL http://dx.doi.org/10.1190/1.3196241.

Yang, D. and Oldenburg, D. W. Three-dimensional inversion of airborne time-domain electromagnetic data with applications to a porphyry deposit. *Geophysics*, 77(2):B23–B34, 2012. doi: 10.1190/geo2011-0194.1. URL http://dx.doi.org/10.1190/geo2011-0194.1.

Zaslavsky, M., Druskin, V., Abubakar, A., Habashy, T., and Simoncini, V. Large-scale gauss-newton inversion of transient controlled-source electromagnetic measurement data using the model reduction framework. *Geophysics*, 78(4):E161–E171, 2013. doi: 10.1190/geo2012-0257.1. URL http://dx.doi.org/10.1190/geo2012-0257.1.

Zhdanov, M. S. *Geophysical Electromagnetic Theory and Methods*. Elsevier Science, 1st edition, 2009. ISBN 9780444529633. URL http://store.elsevier.com/Geophysical-Electromagnetic-Theory-and-Methods/Michael-Zhdanov/isbn-9780080931760/.

This page is intentionally left blank.

Chapter 6

Future work

This PhD thesis is focused on relevant topics related to current methods for large-scale 3D CSEM data inversion. There are complementary issues that can be addressed. This chapter describes several directions to extend this research.

The research on higher-order methods than Gauss-Newton for 3D CSEM problems was not a primary objective of this thesis. However, about this topic a complementary subject is the validation of the approximation that it is done in chapter 5 when using the finite-difference frequency-domain solution output for evaluating the Green functions at the same position of the unit dipole moment source. Although nonsignificant effects were observed in the presented inversion results, it is interesting to compare the numerical results with the analytical results of these functions. The time necessary to compute this validation was the main reason for leaving this issue for future papers.

Another topic outside the scope of this PhD research is the efficient distributed computation of the 3D CSEM Gauss-Newton inversion (not applying the presented lowrank approximation). When working with large-scale surveys, the number of data is such that it is necessary a large number of computers. Currently, an implicit representation of the Jacobian matrix or the Gauss-Newton Hessian matrix is not affordable in a common computer node when a large number of inversion parameters are involved. It is necessary to study different ways to divide this computation, looking for the best balance between the lowest communication cost and the maximum computation performance.

An extension of this PhD thesis is to test the proposed optimization methods with synthetic resistivity models with more complicated topology. It would be useful to evaluate the results of the different inversion paths in these scenarios.

It would also be interesting to apply the methods included in this thesis in joint-inversion schemes (that include seismic and/or magnetotelluric data), to evaluate their effect in this type of inversions.

The development of preconditioners for solving the linear equation systems that

appear in 3D CSEM problems would help to speed up the inversion process. This is not necessary when applying the presented matrix free recursive solver in the lowrank approach, but it would be useful when applying the straight forward Gauss-Newton method.