ELECTROMAGNETIC SUBSURFACE IMAGING AT VLF WITH DISTRIBUTED OPTIMIZATION

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

David Strauss

June, 2013

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Umran Inan, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Sigrid Close

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Ivan Linscott

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Abstract

The purpose of radio remote sensing is to learn more about the environment by observing how the environment interacts with radio waves. Radio remote sensing can be seen as three generic steps: measurement, processing, and interpretation. In this thesis, we discuss methods to address the latter components of radio remote sensing as they pertain to electromagnetic subsurface imaging. We describe a signal processing technique, known as sparse separation, that allows us to decompose the observed data into components that are of scientific value and those that are considered interference. We formulate electromagnetic subsurface imaging as an optimization problem constrained by a set of partial differential equations (PDEs), specifically Maxwell's equations, which govern electromagnetic wave propagation. Algorithms for approximating the solution to these optimization problems are presented and tested numerically.

We image conductivities underground with natural sources of electromagnetic radio waves. In the VLF (Very Low Frequency, 3-30kHz) band there are many types of naturally occurring electromagnetic waves. We are primarily interested in a class of electromagnetic waves known as radio atmospherics, or sferics for short. Sferics are short-duration, broadband radio bursts produced by lightning discharges worldwide. Sferics provide three basic advantages for electromagnetic subsurface imaging. First sferics are plentiful; lightning occurs at a rate of 45 discharges per second. Second, because of the global distribution of lightning discharges, we are able to observe sferics incident from multiple different directions of arrival. Each independent incident direction provides more information about the subsurface conductivities. Third, sferics are broadband in nature. Each sferic contains information at multiple frequencies, further augmenting the information we have collected about the subsurface area of

interest. The techniques developed in this thesis are designed to take advantage of these three basic properties.

To better understand the data, we introduce and develop the technique of sparse separation in an overcomplete dictionary. By solving for a sparse, higher dimensional representation of our data we are able to partition the data into components that are scientifically valuable and components that are considered interference. We develop methods based on soft thresholding, the proximal operator for the sparsity promoting ℓ_1 bound. These first order methods scale well, thus permitting the processing of large datasets.

Electromagnetic subsurface imaging is formulated as a nonlinear, non-convex, PDE-constrained optimization problem. Using the finite difference frequency domain (FDFD) method for modeling Maxwell's equations on a discrete grid, we can predict the electromagnetic fields everywhere within a computational domain for any wave illumination and any set of scattering conductivities underground. The goal of electromagnetic subsurface imaging is to find an optimal set of conductivities within our model of Maxwell's equations that predict electromagnetic fields that closely match a set of electromagnetic field observations. Many techniques exist for approximating the solution to non-convex optimization problems including sequential linear approximation, alternating projections, and semidefinite relaxations and embeddings. Adaptations of the alternating directions method of multipliers (ADMM) to this non-convex problem have shown the best performance for obtaining the best accuracy and incorporating the multiple frequency and multiple direction of arrival information from sferics in a distributed and scalable manner.

Development of these algorithms leads to computational experiments to discover the real-world performance bounds. Adding more information to the problem increases estimation accuracy of subsurface conductivities. We investigate performance as a function of background conductivity and sensor noise. Varying these parameters allows us to predict the influence of basic physics principles and test the algorithms in a variety of possible real-world scenarios. Furthermore, we explore the relationship between the number of sensor measurements necessary and the accuracy we can obtain for determining the subsurface conductivity parameters. Numerical testing

leads to a subsampling principle which gives a relationship between accuracy and the number of radio observations.

Acknowledgements

It is a pleasure to thank the many people who have made the work of this thesis possible. I would especially like to thank my advisors, Dr. Ivan Linscott, Dr. Umran Inan, and Dr. Sigrid Close, for encouraging me to explore and learn. I would also like to thank the members of the VLF and SESS research Groups for helping to conduct experiments and to keep servers running.

I also want to thank my friends, my sisters, Sarah and Anne, and my parents, Jim and Carla, for their kind words and endless support.

David Strauss
Stanford, California
June, 2013

This research was supported by the Bosack-Kruger Foundation.

Contents

List of Tables

List of Figures

Part I

Introduction

Chapter 1

Introduction

Electromagnetic subsurface imaging is the process of estimating and mapping material properties under the surface of the Earth using radio observations. Electromagnetic subsurface imaging is a specific application of radio remote sensing. Radio remote sensing techniques have been used for many years to measure and determine geophysical and astronomical properties including terrestrial ionospheric studies [32], measurement of distant atmospheres [139], and snow accumulation [114], [124], [60].

In this thesis, we are concerned with the use of radio remote sensing as a way to image valuable resources underground. Specifically, we investigate the use of natural sources of radio waves at VLF (3-30 kHz) to image conductivity structures underground. Remote sensing is a diverse field incorporating aspects of physics, applied mathematics, and computer science. The work of this thesis incorporates all three aspects, but focuses mostly on the applied mathematics issues of radio remote sensing. We are faced with challenges for collecting data, storing data, processing data, and building models to explain and understand the data. Electromagnetic subsurface imaging combines all of these aspects. In this thesis, we focus on the latter two aspects, *i.e.*, processing the data to determine and extract interesting content and building models to explain and understand what we have observed.

The work of this thesis began with a series of field campaigns to directly measure the electromagnetic perturbations near active gold mines including the Rawhide Mine near Fallon, NV, and the Edgar Mine near Idaho Springs, CO. Armed with the

background expertise and hardware development for studying VLF radio wave propagation a project to investigate local environmental imaging, specifically subsurface mineral imaging using a network of the low noise VLF receivers developed at the Stanford VLF Group, we set out to map subsurface conductivity parameters.

Ultimately, more work was necessary in order to understand and interpret the data that were collected during these field campaigns. Without any analytic work backing our subsurface imaging idea and without taking much more time to design a really robust data acquisition system, we were unable to formulate experiments and collect data in the field that would explore questions related to electromagnetic subsurface imaging.

The key insights for the development of the work presented in this thesis came after a concerted effort was made to disconnect the logistics of collecting data and the difficulties of a noisy data set from the theoretical and mathematical ideals necessary for electromagnetic subsurface imaging. A framework based on linear algebra, optimization, and computational electromagnetics was formulated and explored based on the supposition of an ideal data set. This numerical and computationally intensive approach was developed so that we could explicitly develop a technique for producing images of the conductivity distribution of subsurface materials.

Along the way, we developed methods for filtering the data and separating the interesting features of the data from interference. In conjunction with data processing tasks necessary for a side project on hypervelocity dust impacts [93], we developed the method of sparse separation presented in Part II. The research related to these tools helped develop fundamental and necessary insights into numerical optimization and ill-posed problems that were influential in formulating and developing the algorithms and numerical results for electromagnetic subsurface imaging.

As the details of the numerical problem became more apparent, it was clear that the data sets previously collected would be insufficient to provide the quality of data necessary for electromagnetic subsurface imaging. Furthermore, the details of solving the electromagnetic subsurface imaging problem made us realize that a full three dimensional model of the subsurface electromagnetics necessary for working with real

data was computationally infeasible. Therefore, the focus of the work on electromagnetic subsurface imaging in this thesis is on the numerical machinery necessary for transforming a set of data, in our case artificial data produced by computational models, into the best estimate of subsurface conductivity. The realization of algorithms that can incorporate the unique aspects of the natural sources of VLF radio waves that we can collect and assessment of their performance is the main contribution of this thesis.

The methods developed here rely on optimality principles. We wish to find the best estimate of the separation of our data signals and the best estimate of the subsurface conductivity parameters. These "best" parameters are mathematically expressed as the solution to a numerical optimization problem in which we want to find the optimal parameters in a model of VLF propagation that closely match field observations. We design an algorithm that can choose the best conductivity image from the set of possible conductivity maps that best explains our observations. The structure of these subsurface conductivity maps is arbitrary; we have very little information about the structure and magnitude of subsurface conductivities. Due to this fact, we must resort to numerical techniques in order to compute a possible subsurface conductivity map from the infinite space of possible conductivity maps. The details of these methods are explained in Part III.

In the rest of this chapter we set the stage for the later developments of the thesis. In Section 1.1 we review many of the basic concepts of VLF radio propagation, a field of study and expertise of the Stanford VLF research group, representing the accumulated knowledge of decades of research. Both sparse separation and electromagnetic subsurface imaging are ill-posed optimization problems; we briefly introduce the basics of numerical optimization necessary to understand the development of both sparse separation and electromagnetic subsurface imaging in Section 1.2.

In this thesis, we pursue numerical optimization strategies due to the data-driven nature of the work. For many data driven problems, a formula or a set of computations can be described that will ultimately lead to the correct or optimal answer; it is then up to a computer to execute these computations in order to complete the learning process.

The main contributions of this thesis require the combination of these two academic disciplines; we apply optimization routines to the well developed observation and modeling of VLF radio propagation. By investigating optimal principles of VLF radio wave propagation, we can directly use the observed data to learn and image specific aspects of the local environment. However simple explanation of the VLF subsurface imaging problem is not enough; we develop and implement tools for distributed, scalable algorithms of recent interest in computing and numerical optimization. With recent trends toward introducing computing power as a simple utility available to everyone, we are motivated to investigate these resources and develop code that is adaptable to these new computing paradigms.

1.1 VLF Radio Propagation

The propagation of Very Low Frequency (VLF, 3-30 kHz) radio waves has been an active area of research for many years and has led to the discovery of many geophysical phenomenon, [28], [8], [77], including the development of a global lightning detection networks [123], observation of particle precipitation [89], and observation of trans-ionospheric propagation [95], [133]. There are numerous sources of electromagnetic radiation in the VLF frequency band (VLF - Very Low Frequency, 3-30 kHz) including radio atmospherics, man-made transmitters, and geophysical phenomenon such as chorus, whistlers, and hiss [67].

Radio signals at VLF frequencies are easily observed with a magnetic field loop antenna. The Atmospheric Weather Electromagnetic System for Observation Modeling and Education (AWESOME) developed by the Stanford University VLF Group has enabled global observation of VLF phenomenon [42]. Adaptations of this receiver have enabled three-axis measurements of the magnetic field using orthogonal loop antennas. Further adaptations have enabled field observation in very remote settings by constructing weatherproofed versions of these receivers that operate on battery power. The AWESOME receiver directly digitizes signals produced by oscillating magnetic fields. By digitizing at a rate of 100 kHz, we are able to resolve frequency content up to 50 kHz thus encompassing the entire VLF band. Deployment of these

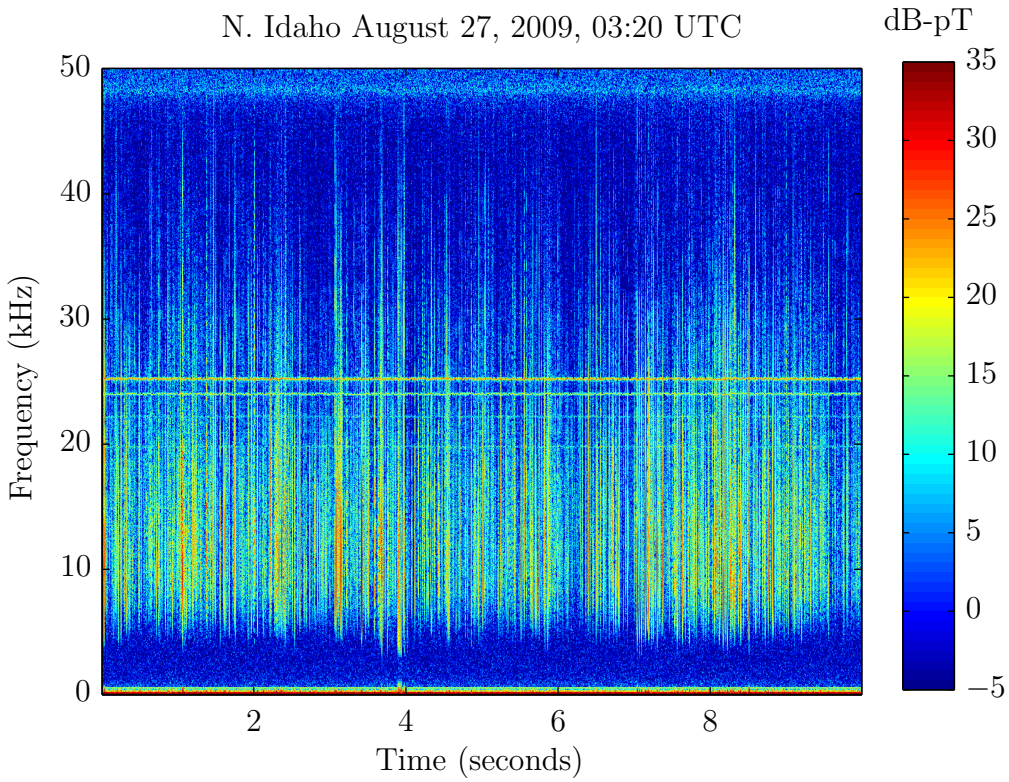

Figure 1.1: Typical VLF spectrogram recorded in Northern Idaho on August 27, 2009. Color maps the intensity of the observed magnetic field at a particular point in time-frequency space. Two distinct classes of signal can be seen, transmitters (horizontal lines) and sferics (vertical lines). Signal from the north-south facing antenna is shown.

receivers has led to the recording of hundreds of terabytes of data over the last few decades.

A typical example of VLF observation data is shown in Figure 1.1. This spectrogram representation of the data shows the evolution of the signals in time-frequency space. It is formed by plotting the amplitude of the Fourier coefficients of overlapping short-time segments of the data. Color corresponds to the intensity of the observed signal at a particular point on the time-frequency plane.

Two distinct classes of signals can be seen in this particular spectrogram. The first class of signals are constant in time and narrow in frequency. These signals are

typically man-made sources including VLF transmitters such as NAA (24 kHz) and NWC (19.8 kHz), and interference from power lines (60 Hz and harmonics). These steady tone signals are valuable for certain remote-sensing applications [44], [117]. The power line harmonics are often considered undesirable interference. The second distinct class of signal that we can identify in the data are signals that are short in time, but broad in frequency. These signals correspond to radio atmospherics, or sferics for short. Sferics are the VLF component of radiation produced by distant lightning discharges. Each discharge radiates across a broad spectrum, but the bulk of the energy is radiated in the VLF band between 5 kHz and 10 kHz [119].

Transmitter signals and sferics propagate for thousands of kilometers in the Earth-ionosphere waveguide [8]. An illustration of Earth-ionosphere waveguide propagation is shown in Figure 1.2. Propagation in the Earth-ionosphere waveguide can be approximated to first order as a parallel plate waveguide where the lower boundary is formed by the conductive Earth and the upper boundary is formed by the ionosphere [78]. The ionosphere lies between 70 km and 120 km and is marked by dramatic increases in plasma and ionized gas content. At VLF frequencies these plasmas are reflective and can refract radio waves. Due to the curvature of the Earth and the anisotropic conductivities of the boundaries, true models of wave propagation in the Earth-ionosphere waveguide can be complicated. It is, however, important that the boundaries are not perfect conductors for the sake of electromagnetic subsurface imaging, the topic of this thesis, which relies on wave propagation to a finite depth in the lower boundary, the Earth's surface. Ray tracing models [94], waveguide approximations [24], and discrete simulation models [96], [34], [64], have all been developed to model radio wave propagation in the Earth-ionosphere waveguide. Once key feature that we can see even in the spectrogram shown in Figure 1.1 is the existence of a cutoff frequency for TM mode propagation near 1.8 kHz. The null of signals between 500 Hz and 1.8 kHz is direct evidence of waveguide propagation.

We are interested in sferics as an illumination source for electromagnetic subsurface imaging for three reasons. Firstly, sferics are widely available in the data. The global rate of lightning discharges is estimated to be 45 discharges per second [40]. Each of these lightning discharges has the potential to produce a sferic. Due to the

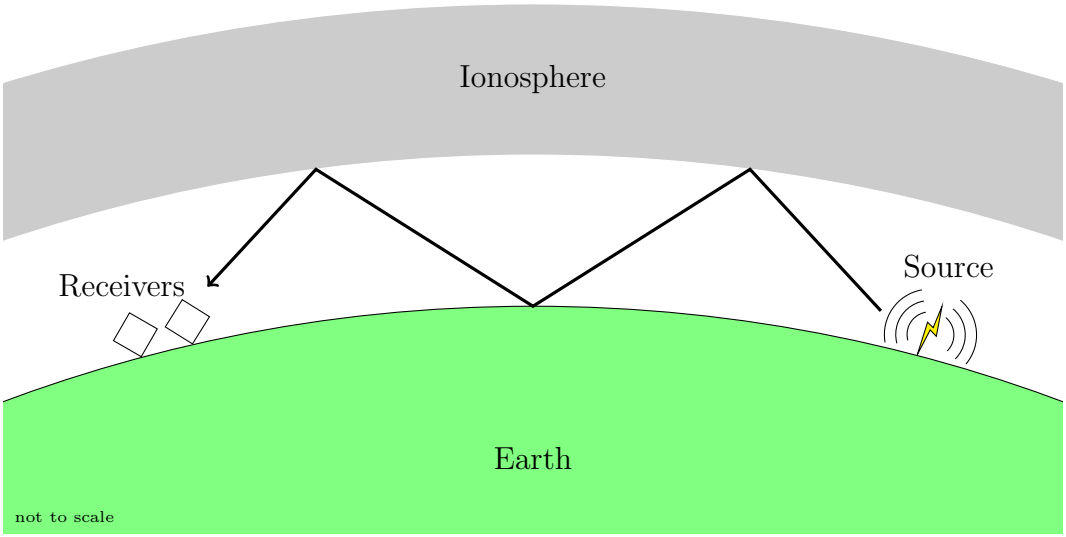

Figure 1.2: Propagation in the Earth-ionosphere waveguide. Lightning discharges radiate in the VLF band. These radio waves, known as sferics, propagate for thousands of kilometers in the Earth-ionosphere waveguide bounded by the surface of the semi-conductive Earth and anisotropic plasma in the ionosphere.

global propagation effects, and propagation in the Earth-ionosphere waveguide, we can observe sferics from lightning discharges thousands of kilometers away.

Secondly, sferics arrive from multiple different directions of arrival. Since sferics are a global phenomenon, it is possible to observe sferics that arrive from many locations, propagating in different directions in the Earth-ionosphere waveguide. We can estimate the direction of arrival of these sferics based on the simultaneous recordings of the stations and can compare these estimated directions of arrival with respect to the known locations of sferic discharges from national databases such as the Global Lightning Detection Network (GLD-360) and the National Lightning Detection Network (NLDN). The polarization of these electromagnetic waves is primarily transverse magnetic (TM), that is H is largely parallel to the surface of the Earth, however there is a known transverse electric component (TE) as well [123]. For both polarizations, both magnetic and electric components of the incident wave are known to penetrate into the subsurface and interact with subsurface conductivity anomalies [127], [45], [145].

Lastly, sferics are broadband in nature. They contain considerable energy in two distinct bands, the extremely low frequency band (ELF 3-300 Hz) and the VLF band (3-30 kHz). The ELF component, often described in the literature as the ELF slow tail [99], is the result of TEM mode propagation in the Earth-ionosphere waveguide. This component of the sferic gives the lowest frequency contributions which can help subsequent electromagnetic techniques to determine conductivities underground at greater depth. Higher modes of propagation in the Earth-ionosphere waveguide are present in the VLF band. These are primarily in transverse magnetic modes (TM). The broadband frequency content of each individual sferic gives additional information to the electromagnetic subsurface imaging problem.

1.2 Optimization

The two problems discussed in this thesis can be formulated as ill-posed, numerical optimization problems. We begin this section by describing the basics of optimization problems in general and then develop what we mean by ill-posed optimization problems. The techniques and strategies for finding solutions to sparse separation and electromagnetic subsurface imaging problems begin in a very similar place, but diverge due to the specific details of the problems themselves.

1.2.1 Numerical Optimization

The general goal of a numerical optimization problem is to find the minimum value and corresponding argument x, of an objective function $f_0(x)$. The variables x may be constrained by a set of inequalities and equalities, $f_i(x)$ and $h_i(x)$ respectively. All of $f(x), h(x)$ map the complex vector space to a single real value, $f_i(x), h_i(x) : \mathbb{C}^n \mapsto \mathbb{R}$. We generally state optimization problems in the form given in (1.1).

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0 \\ & h_i(x) = 0 \end{aligned} \tag{1.1}$$

The optimization problem is considered convex if the functions f, h are convex, that is they satisfy Jenson's inequality [18].

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y) \tag{1.2}$$

Some common convex functions include affine and linear transformations, $f(x) = Ax - b$, norms, $f(x) = \|x\|_2^2, f(x) = \|x\|_1$, and negative logarithm, $f(x) = -\log(x)$.

Many methods exist for obtaining the optimal solution to an optimization problem [18], [111]. Convex optimization problems have the added advantage that a unique optimum can be found using known polynomial-time algorithms, especially interior-point methods, which have been developed over the past decade. The ability to express an optimization problem as a convex optimization problem is thus computationally and theoretically attractive. Of course, not every problem is convex. For non convex optimization problems, the best we can achieve in a reasonable computational time is a local optimum, which may or may not be the global optimum of the problem. Adding additional constraints to the optimization problem can increase the difficulty of finding a feasible, optimal solution. The incorporation of non-smooth functions in optimization problems can cause both theoretical and computational difficulties.

Many methods have been developed for solving constrained and unconstrained optimization problems, and with the introduction of interior point methods, arbitrary convex optimization problems that include non-smooth constraints can be solved. Interior point methods replace non-smooth constraints with a tunable approximation that can be adjusted through iterations of the algorithm to be a better and better approximation until the desired numerical tolerance is reached. Other strategies involving projection methods have garnered increased research [20], [86], [92].

As data and problem sizes grow, there has been increasing interest in optimization problems and optimization routines that are distributed; that is the algorithm has components that can be run in parallel across multiple computers. Optimization strategies tailored to this specific, large dimensional challenge have received increasing attention in the literature including [20] and [116]. First order methods have typically been popular for distributed optimization because they are often able to achieve

tolerable accuracy with less computational effort. Gradient descent methods for unconstrained and constrained environments have been analyzed for multiple agents operating in a distributed environment [108], [107]. Gossip algorithms, or algorithms that reach consensus via repeated pair-wise interaction have also been analyzed for optimization routines [19]. Many other resources describing distributed optimization methods that extend these works [115], [13].

1.2.2 Ill-posed problems

We focus on ill-posed problems in this thesis; both electromagnetic subsurface imaging and sparse separation are in this difficult class of computational and mathematical problems. An ill posed problem can be identified if it violates one of the principles of a well posed problem which we define with three criterion:

1. The solution exists.
2. The solution is unique.
3. The estimates of the algorithm are a smooth function of the inputs.

An ill posed problem is an optimization problem that violates one or more of the principles of a well-posed problem [39]. For the optimization problems of interest in this thesis, it is the latter two points that are typically violated. We are not guaranteed that there is a unique solution; the systems of equations involved have a non-trivial null space. Furthermore many of the operators are poorly conditioned leading to the possibility that small changes in the inputs can yield dramatic changes in the output. For example, take the system of equations $y = Ax$, the basic linear system of equations. In the case that $y \in \mathbb{R}^m, x \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, m < n$, finding the solution $x \in \mathbb{R}^n$ is classically ill-posed; we have an underdetermined system of equations. There are a non-trivial number of vectors, $v : Av = 0$, in the nullspace of the matrix A. Therefore, we can always add a scaled version of a vector v to any solution x to achieve a new solution to this system of equations.

The last condition for well-posed problems can be summarized by stating that the solutions are highly sensitive to the input data. Small changes in the input

can lead to a large changes in the output. A very large condition number for the matrix A implies that at least some of the equations in the system are numerically linearly dependent. By "numerically" dependent, we mean that on a finite precision computing machine, the equations are linearly dependent within numerical precision. We define the condition number, κ, of a matrix A as the ratio of the largest singular value of the matrix A to the smallest singular value of A,

$$\kappa = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}. \tag{1.3}$$

[68]. If the condition number is large, small changes in the right hand side, y can lead to large changes in the solution x.

Ill-posed problems are typically treated with regularization methods to alleviate the "ill-posed" aspects of the problems. Adding regularization can make solution of a poorly conditioned system more stable; regularization can reduce the variability in solutions given small changes to the input. Regularization can also be seen as incorporating prior knowledge into the problem in order to prefer solutions with certain attributes or exclude solutions that may be numerically possible but are meaningless in the larger context of the problem. Many forms of regularization exist to promote certain characteristics of the solution. For example, preference for solutions with numerically small values is equivalent to adding an ℓ_2-norm regularization term to the objective function. Adding a total variation norm brings the additional knowledge that we expect the solution to favor constant values [22], [143], [122]. In sparse separation, we use regularization to select for solutions that are sparse, meaning the number of nonzero entries in the solution vector is at a minimum. In electromagnetic subsurface imaging, we regularize using a combination of constraining the solutions to feasible sets and mild ℓ_2 regularization. With ℓ_2 regularization we modify the system of equations such that we are solving $(A + \rho I)^{-1}$. Indeed, as ρ increases, the relative influence of the matrix A decreases and the system approaches a scaled identity matrix. The combination of these regularization terms gives us reasonable solutions that are also feasible in many circumstances [39].

1.3 Contributions

The contributions of this thesis are stated as follows:

1. Derived, implemented, and compared 7 methods for the non-convex, non-linear subsurface electromagnetic inverse problem so that these methods can incorporate information from multiple frequencies and directions of arrival as given by sferic illuminations. The 7 methods are:
   - Sequential Linear Approximation
   - Artificial Source
   - Split Field
   - Split-σ
   - Phase Split
   - Bi-Projection
   - Semidefinite Relaxation
2. Established relationship between the attributes of our model of Maxwell's equations and the accuracy to which we can estimate subsurface conductivities. We explore model parameters that are representative of real-world models.
3. Established subsampling principle for accurate conductivity recovery.
4. Developed sparse separation technique for time series filtering of morphologically distinct signals in electromagnetic waveform data.

1.4 Outline of Thesis

This thesis is divided into four basic parts, introduction, sparse separation, electromagnetic subsurface imaging, and discussion. The two main parts of the thesis, sparse separation and electromagnetic subsurface imaging are independent and self-sufficient. As described in this section, however, both of these parts draw on the principles of ill-posed optimization problems and electromagnetic wave propagation.

We present sparse separation, a method for decomposing data, such as VLF data into its morphologically distinct components in Part II. This decomposition is framed as an application of under-determined optimization and relaxation to a convex problem in order to do the decomposition. Sparse separation is a linear inverse problem and a clear description of this ill-posed problem sets the stage for the more complicated non-linear ill-posed problem of electromagnetic subsurface imaging discussed in Part III. The relationship between sparse separation and the specific data we have collected is described in detail in Chapter 2. Algorithms for achieving and solving the optimization problem for sparse separation are developed in Chapter 3. Numerical results are presented for these algorithms in Chapter 4.

Electromagnetic subsurface imaging is discussed in Part III. Electromagnetic subsurface imaging is developed as an optimization problem constrained by a set of partial differential equations (PDE-constrained optimization problem) in Chapter 6. Linear algebra and algorithms for solving the electromagnetic subsurface imaging problem are given in Chapter 7. Numerical results for electromagnetic subsurface imaging are given in Chapter 8, including overall performance of the algorithms and performance of the algorithms as various aspects of the model, background conductivity, and noise level, are varied. A subsampling principle which relates the number of measurements to the number of unknowns that can successfully be estimated is also discussed.

The final part of this thesis is discussion of the results and a brief outline of further work and extensions of the methods presented and developed in this thesis.

1.5 Notation

In this thesis, vectors are represented as lower case letters, such as the vector u. Matrices are represented as capital letters, such as the matrix A. Notationally, we use the transpose operation of a vector and a matrix to be both the transpose and the complex-hermitian transpose. These operations are often denoted as A^T and A^* respectively. In this work, we take A^T to mean both. We use the vector notation x^* to indicate the optimal, desired, or truth solution to various problems through the text.

The space of real numbers is $\mathbb{R}$, the space of complex numbers is $\mathbb{C}$, and the space of positive semidefinite matrices is $\mathbb{S}^+$. We use the standard definitions of p norms as

$$\|x\|_p = \left(\sum_i x_i^p\right)^{\frac{1}{p}}.$$

From this definition, it is clear that the 1-norm is the sum of the absolute values of the vector x and the ∞-norm is the maximum absolute value of the entries in x.

Some characters are used to represent static, physical constants, such as ϵ for electrical permittivity, σ for electrical conductivity, μ for magnetic permeability, ω for frequency dependence ($\omega = 2\pi f$). We use SI units and fix $\epsilon_0 \simeq 8.854 \times 10^{-12} (F/m)$ and $\mu_0 = 4\pi \times 10^{-7} (H/m)$. We use $e^{-i\omega t}$ time dependence where $i = \sqrt{-1}$.

We make frequent use of the indicator function, $I_{x \in \mathcal{C}}(x)$ which is a function that indicates set membership to the set $\mathcal{C}$.

$$I_{x \in \mathcal{C}}(x) = \begin{cases} 0 & x \in \mathcal{C} \\ \infty & x \notin \mathcal{C} \end{cases} \tag{1.4}$$

Part II

Sparse Separation

Chapter 2

Morphological Decomposition

Broadband observations of the VLF magnetic field at global field sites contain a mixture of diverse electromagnetic processes. Sferics and transmitters are the two main classes of signal in VLF data and are the focus of the data decomposition technique of sparse separation. Our decomposition of the data allows us to clearly and cleanly separate these two classes of signal solely based on the specific shape of the data; we use very little prior information about the exact content of the observed data. We solve for a sparse combination of basis vectors from the combination of the Fourier and wavelet dictionaries where wavelet components are used to efficiently represent sferics and Fourier components are used to efficiently represent transmitters. Methods for the clean and accurate identification of these signal types has been previously studied in many contexts [42], [67].

2.1 Representation of VLF Data

VLF data can be characterized by three types of content. All of these signals are directly digitized fluctuations of the electromagnetic field and are thus oscillatory in nature. The first type of signals are impulsive sferics which are short in time and, broad in frequency [42]. The second class of signals are those that are persistent in time and narrow in frequency. These signals are typically of man-made origin. Both transmitters such as NAA (24.0 kHz) and 60 Hz power line signals and associated

harmonics fall into this category. On occasion, a third class of geophysical emissions naturally generated in the Earth's high altitude plasma environment, *e.g.*, chorus, hiss, and whistlers, can be seen in VLF data [67].

The separation of these signals specifically in VLF data has many applications including the identification of the location of lightning discharges [123] and the isolation of transmitter signals to investigate disturbances in propagation conditions [117]. In [123], the exact structure of the sferic is used in order to estimate the propagation distance in the Earth-ionosphere waveguide. Sferic identification is done through cross-correlation with a predefined waveform bank derived from the data themselves. Isolation of sferics from interference is fundamental to the success of this method. In [123], linear bandpass filters are used to isolate the sferics. The focus of the work in [117] is to identify amplitude and phase perturbations in VLF transmitter propagation. Phase and amplitude of transmitter signals are estimated from a linearly filtered version of the data. Extracting the phase and amplitude is subject to interference specifically from sferics. Accuracy could be improved using a method that separates these disparate types of signal [82].

Prior attempts to isolate these signals assume specific knowledge about the exact frequency and modulation of these signals. Linear filters have been applied in addition to adaptive filters [41]. Neither of these methods specifically models the effect of the sferics in the data and thus these linear filter methods can often fail when strong sferics are present. The method of sparse separation is a nonlinear method for estimating the optimal separation between these disparate types of content in the data.

Sparse separation is a preprocessing step in order to understand the constituent components of the data. This separation allows us to isolate the interesting portions of the signal from the non-interesting portions of the data, potentially providing a means to excise other sources of interference, *i.e.*, noise from local power lines. Methods have been developed that track the amplitude and phase of the transmitter signals, a process that has encountered difficulty in the past due to the presence of many sferics in the data [41]. Methods for excising the sferics based on linear predictive models have also been used, but these methods can introduce new data artifacts [67]. Furthermore, characterization of average sferic waveforms is made difficult by

transmitters. Allowing for this explicit data decomposition based on morphological form, we are able to explicitly identify these components. The methods described in this Part were developed for separating different types of events in directly digitized electromagnetic data at, but not limited to, radio frequencies ($f < 300$GHz) based on sparse separation in redundant dictionaries.

2.2 Sparsity

The number of sferics observed in any appreciable segment of data (longer than 5 seconds) is sparse. On average, lightning discharges occur globally at a rate of 45 lightning discharges per second. Compared to the 100 kHz data rate used for digitizing VLF signals, the set of discharges is sparse.

Sferics are not assumed to be exactly sparse; that is, a single sferic is not known to have an exactly sparse representation in a pre-defined dictionary. Sferics are, however, approximately sparse. The majority of the energy of a sferic can be represented using a small number of wavelet coefficients. The sparsity we are primarily interested in, however, is that the number of sferic events in any window of time is sparse with respect to the quantity of sampled data. Even in a 10-second segment of data sferics are sparse. Due to the short time duration of each sferic they can be closely approximated by a small number of wavelets. We search for a sparse separation in the combination of the Fourier matrix and the wavelet matrix that explains the data. The sparse coefficients recovered in the wavelet transform well represent the sferics while the sparse collection of coefficients recovered in the Fourier transform well represent the transmitter signals and other steady state interference such as 60 Hz harmonics.

2.3 Wavelets and Fourier Series

The discrete wavelet transform projects a given signal onto a hierarchical family of orthogonal multi-scale atoms with specific time-frequency profiles [48], [101]. Sparse separation seeks to discriminate between events that have compact support in the time domain from events that have compact support in frequency domain. We used

the Daubechies family of discrete wavelets because they offer compact support in the time domain, bounded support in the frequency domain, and the Daubechies wavelet transform can be calculated in $O(n)$ time [48], [101]. Furthermore, the Daubechies-10 wavelets have a resemblance of the time domain sferics observed in our data.

The wavelet transform can be described as the convolution of an input data vector with a set of orthogonal prototype wave forms. A hierarchical decomposition involves projecting the data onto a collection of wavelets that have the same shape but are dilated and translated in time. This projection can be conceptualized as a matrix-vector operation, multiplying a data vector against the orthogonal wavelet projection matrix. The fast wavelet transform is an algorithm that accomplishes this matrix-vector operation in $O(n)$ time as opposed to the standard $O(n^2)$ time for matrix-vector multiplication. This time reduction is accomplished by convolving the same wavelet with successively down-sampled versions of the input data. Therefore only a few convolve and downsample operations are necessary to accomplish the orthogonal wavelet decomposition. More details on the fast wavelet transform can be found in [101]. The wavelet decomposition is best at capturing and representing local information in a dataset. The periodized dyadic wavelet transform is a one-to-one orthogonal projection.

The discrete Fourier transform projects an input signal onto the set of complex, orthogonal sinusoids (2.1). In this thesis, we use the Fourier transform in order to find a representation of our data in the frequency domain. Raw time-series data representing the sequential behavior of a signal in time can be transformed to a frequency domain representation using the Fourier transform.

$$\hat{x}_j = \sum_{k=1}^{N} e^{-i2\pi jk/N} x_k \tag{2.1}$$

The Fourier transform is a unique representation of the data that captures the harmonic content of the time series. Each Fourier term captures information about the entire input data vector; each Fourier component is non-local in time. Discrete, short time events will require a dense representation in the Fourier domain, while steady

sinusoidal signals in the time domain have a simple, sometimes single point, representation in the frequency domain. The classic example of this localization trade-off can be seen with the Dirac delta function. A delta function in the time domain is a short-time event that is represented by a single point in the time series. The Fourier transform of this simple time-domain signal is dense; the constant function in which every Fourier coefficient has a non zero value. To represent this short time event, combination of all possible orthogonal sinusoids is necessary. More details about the Fourier transform and associated transforms can be found in [21].

Like the wavelet transform, the discrete Fourier transform can be seen as explicit matrix-vector multiplication, however the Fast-Fourier Transform (FFT) algorithms are able to achieve this projection step in far less computational time ($O(n \log n)$) as opposed to naive dense matrix-vector multiplication ($O(n^2)$).

2.4 Prior Work in Sparse Separation

The literature in signal decomposition can be split into two general components: global denoising and content recognition. The method of sparse separation combines aspects of both. We are able to accurately estimate noise-reduced versions of the observed signals, but we accomplish these estimates based on modeling the content and identifying the constituent components of the data. Our approach to content separation is similar to previously published work [147].

Wavelet denoising has been an active area of research, and perhaps one of the first places where the idea of sparsity and ℓ_1 norms began to show value and promise. The early work on soft and hard wavelet thresholding opened new avenues for signal processing and denoising [54] [50]. Wavelet soft thresholding was later extended to combining multiple types of waveform "atoms" [33]. Continued work has led to the related idea of compressed sensing in which random dictionaries and random projections are used instead of highly structured dictionaries like wavelets and Fourier series [55]. These solutions may make us uneasy; we are able to recover more information, a unique higher dimensional representation, by solving for more unknowns than true equations. The theoretical implications of this discovery are still being explored [53],

[49], [31].

Compressed sensing has also led to decompositions known as Morphological Component Analysis (MCA) [17], [131], [23], [63]. The method of sparse separation is similar to MCA. Morphological separation rests on the heuristic that a particular data vector may not be sparse in one basis set but might have a sparse representation in a combination of basis sets. For 2D images, combinations of curvelets, discrete cosine transforms, and wavelets have all been used to separate differing types of image content [132], [130]. Applications to 1D time-series data such as those in this work have received less attention in the literature than 2D imaging processing.

Alternative methods for decomposing a set of signals into their constituent components include principal component analysis (PCA) [59] and independent component analysis (ICA) [76]. Principal component analysis discovers a rotation of the data into directions with maximal variance. Independent component analysis separates a set of vectors into components that have minimal mutual information. In contrast to PCA and ICA, which typically seek to find a lower dimensional space to represent the data, MCA separates signals into their constituent components within a redundant or overcomplete dictionary.

Data driven methods for learning sparse dictionaries such as K-SVD [6] pose computational difficulties for the size of the data used in this thesis. Methods such as K-SVD, PCA, and ICA are all difficult to apply to datasets like the VLF data set because of dimensionality and stationarity issues. In the data set, we have a collection of events with unique and disparate characteristic time scales. The sferics and other geophysical events are relatively short in time, thousands of sequential time-series points. The transmitters, in contrast, are persistent throughout the data set. The decomposition of the data must be able to capture the combination of this global and local behavior. Finite-dimension projections like PCA and ICA must break the data down into windowed segments and thus put a limit on the maximum length event that can be easily represented simply through the choice of the window length. Some data sets naturally have a partitioned structure of reasonable size and lend themselves naturally to this type of decomposition. No such reasonable partitioning of the data exists in our VLF datasets. Furthermore windowing methods like K-SVD, PCA, and

ICA, suffer in that they are not robust to translational variability in the dataset. For representing sferics, there is equal likelihood that a sferic will occur at any place within the windowed segment. Representation of this translation invariance would require a much larger, non-orthogonal dictionary of translated prototypes, which is not supported by these methods.

Chapter 3

Algorithms

The method of sparse separation searches for a parsimonious representation of our data by solving a linear ill-posed optimization problem. The method of sparse separation is a nonlinear method for estimating the optimal separation between these disparate types of content in the data. In this chapter we present the explicit mathematical description of the problem and several algorithms that allow us to efficiently compute the solution to this optimization problem.

3.1 Signal Model

We model each observed signal, $y \in \mathbb{R}^n$, using a noise corrupted linear model. The entries of the vector $x \in \mathbb{C}^N$ rescale the column vectors $\phi_i \in \mathbb{C}^n, i = 1, \ldots, N$. Taken together, these column vectors form the dictionary or the matrix $\Phi \in \mathbb{C}^{n \times N}$. Additive Gaussian noise, ν, with variance σ^2 corrupts our observations. We can express this succinctly as

$$y = \sum_{i=1}^{N} x_i \phi_i + \nu, \quad y = \Phi x + \nu, \tag{3.1}$$

$$\nu \sim \mathcal{N}(0, \sigma^2). \tag{3.2}$$

In this technique, our dictionary is overcomplete, meaning the number of columns exceeds the number of rows, $n < N$ in the matrix Φ. Therefore, the matrix Φ

has a non-trivial nullspace and a unique $x : y = \Phi x$ does not exist. To constrain the solution, we are interested in finding the sparsest x, *i.e.*, the x with minimum cardinality or number of non-zero entries. We define the cardinality function $\mathbf{card}(x)$ to be the count of non-zero entries in a vector x and a k-sparse vector to be one for which $\mathbf{card}(x) = k$. In the case that y does not contain noise, we wish to solve to obtain an optimal estimate for x, denoted $\hat{x}$:

$$\begin{aligned} \hat{x} = \underset{x}{\text{argmin}} \quad & \mathbf{card}(x) \\ \text{subject to} \quad & y = \Phi x. \end{aligned} \tag{3.3}$$

Obtaining a solution to this problem is NP-hard. It has been shown [57] that under suitable dictionary and sparsity constraints, the minimum cardinality solution can be recovered by minimizing the ℓ_1 norm which is the convex relaxation of the cardinality function [33]:

$$\begin{aligned} \hat{x} = \underset{x}{\text{argmin}} \quad & \|x\|_1 \\ \text{subject to} \quad & y = \Phi x. \end{aligned} \tag{3.4}$$

Given that our measurements contain Gaussian noise, we need not force equality of the data and dictionary description but enforce that the log likelihood of the error be contained by the standard deviation of the noise, σ, as stated in (3.2) [141] [142].

$$\begin{aligned} \hat{x} = \underset{x}{\text{argmin}} \quad & \|x\|_1 \\ \text{subject to} \quad & \|y - \Phi x\|_2 \leq \sigma. \end{aligned} \tag{3.5}$$

Problem (3.5) is difficult to solve exactly as well given that the true noise level σ is either unknown or hard to estimate. Therefore, a different form of (3.5) is solved that yields the same results for the appropriate choice of penalty parameter, λ, [56]:

$$\hat{x} = \underset{x}{\text{argmin}} \, \|x\|_1 + \lambda \|y - \Phi x\|_2^2. \tag{3.6}$$

There are many algorithms for solving the problems (3.3), (3.4), and (3.6) including orthogonal matching pursuit (OMP) [137], iterative weighted least squares

(IWLS) [27], interior point cone solvers [88], alternating direction method of multipliers (ADMM) [20], fast iterative shrinking and thresholding (FISTA) [9], and approximate message passing [51]. Each of these algorithms has its particular advantages and disadvantages, *e.g.*OMP is best if the sparsity is fixed and known while interior point cone solvers converge quadratically.

In the method of sparse separation, we choose a dictionary that is composed of morphologically distinct components. In this particular instance, we choose the concatenation of the orthogonal wavelet basis given by the Daubechies wavelets and the Fourier basis. Therefore, the dictionary Φ can be expressed as

$$\Phi = \begin{bmatrix} F^T & W^T \end{bmatrix}$$

where F corresponds to the orthogonal Fourier basis and W corresponds to the orthogonal, periodic wavelet basis.

After a solution to (3.6) has been found, we can separate the components of the solution, x, into the components that correspond to the Fourier basis vectors and the components that correspond to the wavelet basis vectors, *i.e.*, $x = \begin{bmatrix} x_{\text{Fourier}} & x_{\text{wavelet}} \end{bmatrix}^T$. From this, we get

$$\begin{aligned} y &= \Phi x + \nu \\ &= F x_{\text{Fourier}} + W x_{\text{wavelet}} + \nu \\ &= y_{\text{Fourier}} + y_{\text{wavelet}} + \nu. \end{aligned} \tag{3.7}$$

The separation of these two components of the data is the method of sparse separation and allows us to decompose the data in a parsimonious way based on morphologically distinct components.

Furthermore, we introduce a truncation of the wavelet expansion. The wavelet expansion of a signal is a hierarchical representation, therefore, it contains a mixture of both local and global information about the input signal. The lowest order wavelet coefficients represent trends in the data that are long in time and low in frequency. For completeness, these representations must exist in the full orthogonal dictionary. However, due to the signals of interest in our data, we know that lower frequency, long-time duration signals will be better represented by the Fourier components than by the

wavelet components. Therefore, we exclude the lowest frequency, lowest order wavelet coefficients from the wavelet representation. For example, if the wavelet transform contains P levels, we may only use the levels $3, 4, \ldots, P$ which correspond to the higher frequency, shorter time wavelet expansion coefficients. With this reduction, we minimize interaction and ambiguity between the two dictionaries at the lowest, broadest scales. These reduced wavelet transforms are referred to as the $p-$truncated wavelet transform, where we use the higher order wavelet coefficients down to the pth level.

3.2 Iterative Soft Thresholding

The FISTA algorithm is a first order optimization algorithm for solving the convex optimization Problem 3.6. It is an accelerated version of the simple iterative soft thresholding algorithm. We leave discussion of the acceleration steps to its original source, [9]. In the Iterative Soft Thresholding Algorithm (ISTA), new estimates of x are obtained by applying the proximal operator of the ℓ_1 norm, the soft thresholding operator, $\eta(x, \lambda)$, to the residuals in the transform domain,

$$x_{k+1} = \eta\left(x_k + \frac{1}{c}\left(\Phi^T(y - \Phi x_k)\right), \lambda\right). \tag{3.8}$$

In this algorithm, c is a constant that depends on the largest eigenvalue of the matrix Φ, and we define the soft threshold operator η as:

$$\eta(x, \lambda) = (x)(\max\{1 - \lambda/|x|, 0\}). \tag{3.9}$$

The thresholding parameter, λ^k, is chosen at each iteration as the best estimate of the standard deviation of the residual. By keeping only the outliers, *i.e.*, the points outside 2.5σ, we highlight and select the coefficients that stand out from the noise which supports the heuristic that we are separating content into two morphologically distinct dictionaries - what appears structured in one dictionary appears as noise in the other dictionary. To estimate the thresholding parameter λ at each iteration k, we

use the robust, unbiased estimator for the standard deviation based on the median:

$$\lambda_k = \text{median}\{\Phi(y - \Phi^T x_k)\}/0.6745. \tag{3.10}$$

3.3 ADMM Algorithms

We can alternatively solve the minimum ℓ_1 problem using the alternating directions method of multipliers (ADMM). With the proper choice of internal parameters, ADMM algorithms can have similar speed performance to iterative soft thresholding methods. Problem 3.6 is expressed in ADMM form by introducing a new variable, $\tilde{x}$ such that at optimality, $x = \tilde{x}$. We rewrite the problem as

$$\begin{array}{ll} \text{minimize} & \|y - \Phi x\|_2^2 + \gamma \|\tilde{x}\|_1 \\ \text{subject to} & x = \tilde{x}. \end{array} \tag{3.11}$$

We then can apply the alternating directions method of multipliers directly as outlined in [20]. The ADMM method for solving the minimum ℓ_1 problem is split between three steps, a projection step, a thresholding step, and a calculation of the residuals. These three steps can often be evaluated quickly and efficiently. We discuss an efficient way to implement these operators for large problems in the next section.

3.4 Implementation

For both of these optimization methods, we must look at the action of the fat matrix, Φ, where

$$\Phi = \begin{bmatrix} F^T & W^T \end{bmatrix}.$$

Some numerical solvers such as `CVX` and SeDuMi require explicit matrices in memory [70], [69], [135]. Explicitly forming these matrices is memory intensive when the vectors y or x are large, *i.e.*, larger than 2^{10}. The optimization methods described in Sections 3.2 and 3.3 do not require explicit matrices. We can solve optimization problems with large dimensions by using pipelined operators. Rather than explicitly

compute the entries of the matrices F, W, the symbols representing these objects are overloaded with methods that accomplish the Fourier transform and the wavelet transform respectively. Overloading the operators, however means that we can no longer perform direct factorizations of the matrix $\Phi\Phi^T$. Instead, inversions of this matrix must be accomplished using pipelined iterative methods. Since we are working with $\Phi\Phi^T$, we are guaranteed that the matrix is positive semidefinite and we can use conjugate gradients to invert this matrix. For conjugate gradients, we only need the functional ability to perform matrix-vector operations which can be done in this case using fast transform methods [84].

Chapter 4

Numerical Results

In Chapter 3, we formulated sparse separation as the solution to the convex relaxation of the minimum cardinality problem. We search for a sparse solution in the combined Fourier and wavelet dictionary that adequately represents the data. In the first part of this chapter, we establish a relationship through computational experiments that solving for the optimum of the convex problem (3.6) gives the solution for the non-convex optimization problem (3.3) when the descriptor vector, x, is less than 20% sparse, *i.e.*, fewer than 20% of the coefficients are non-zero. In the second part of the chapter, the results from applying the method of sparse separation directly to data are presented.

4.1 Subsampling

Recovery of an unknown high dimensional vector x from a lower dimensional projection, $y : y = \Phi x$, is mathematically equivalent to compressed sensing, a method for uniquely solving underdetermined systems of equations under an assumption of sparsity [55]. Discovering the problem parameters (k, n, N) for which solving (3.4) recovers the k-sparse model vector x from a projection of of the $n \times N$ matrix Φ taken from an ensemble of matrices has been an active area of research [57], [55], [56], [137], [144], [150], [138], [25], [26], [23], [52], [65]. In most cases, analytical bounds have failed to adequately predict the location of observed phase transitions in the

performance of these algorithms [53].

The relationship between the probability of successful recovery and the tuple (k, n, N) is related to high-dimensional geometry, specifically the face counts of the projections of high-dimensional polytopes [49], [53]. Results from computational experiments support this hypothesis for several matrix collections including random normal Gaussian matrices and partial Fourier matrices [53].

As described in Section 3.1, we are concerned with discovering a high dimensional mixture of atoms with particular morphological shape that well represent observed data. To demonstrate that solving (3.4) recovers the exact, unique, sparse vector x^*, we run computational experiments similar to those proposed in [52] over the truncated-wavelet and Fourier dictionary used in this thesis and described below.

We focus on the dictionary Φ constructed from the concatenation of the discrete Fourier transform (DFT) and the discrete periodized wavelet transform (DWT), $\Phi = \begin{bmatrix} F^T & W^T \end{bmatrix}$. We truncate the discrete wavelet transform to minimize overlap between the wavelet and Fourier domains. The discrete wavelet transform is a dyadic transform that produces 2^p wavelet coefficients for each level p, $0 \leq p \leq P = \log_2(n) - 1$. For small values of p, *i.e.*, $p = 0, 1, 2$, coarse scale wavelets are obtained. These coarse wavelets have long time duration and low frequency content. Heuristically, these wavelet levels are similar in character to low frequency components of the Fourier transform. For higher values of p, *i.e.*, $p \simeq \log_2(n)$, fine scale wavelets are obtained; these wavelets have short time duration and higher frequency content. Given the datasets of interest, we are interested in separating our data into events with sparse support in the time domain well represented by fine-scale wavelets and events with sparse support in the frequency domain represented by the Fourier matrix. Eliminating the coarser wavelet levels gives a matrix with a lower redundancy ratio. For a lower redundancy ratio, we predict the phase transition will occur for higher sparsity levels. By combining the full Fourier matrix and wavelet decomposition truncated below the pth level we obtain a family of matrices that range from a square $n \times n$ matrix that consists only of the Fourier matrix to an $n \times 2n$ matrix that consists of the full Fourier and full wavelet matrices. By decomposing down to the pth level, our matrices have dimensions fixed at $n \times (n + \sum_{j=p}^{J} 2^j)$. Thus we cannot explore

the entire range of (k, n, N), only a fixed set as dictated by the dyadic nature of the wavelet transform.

4.2 Testing

For each choice of matrix Φ from the ensemble, we ran 70 Monte Carlo simulations to verify that solving (3.4) recovers the sparsest descriptor. In each of the Monte Carlo trials, a test vector, x^* is generated that has k values of ± 1. Test data, $y = \Phi x^*$ is generated and problem (3.4) is solved, to obtain an estimate, $\hat{x}$. We use a solver based on the alternating directions method of multipliers (ADMM) [20] to solve problem (3.4) to high tolerance. If the mean squared error, $MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{x}_i - x_i^*)^2$, is less than our error threshold, 1×10^{-4}, our solution to the problem is deemed a success. In Figure 4.1, we present the results from these experiments. Matrices were chosen such that the number of rows remained constant at $n = 512$. The number of columns in these matrices varied depending on the number of wavelet levels included in the dictionary. At each choice of dictionary, we chose 30 equally spaced values for k, the sparsity of the true solution x^*.

The probability of success for each triplet (k, n, N) estimated from 70 Monte Carlo trials is a unique point on the relative cardinality, k/n, vs redundancy, N/n, grid. Red dots indicate a high probability of recovering the true vector x^* via solving (3.4) for that triplet. The figure indicates that once wavelets are added to the dictionary, the probability of recovering vectors with less than $n/2$ non-zero entries is high. As more wavelet levels are added to the dictionary, *i.e.*, the dimension of the dictionary N, this relative sparsity value separating success from failure remains relatively constant. Therefore, these computational experiments indicate that if a sparse vector, x, with approximately $n/2$ non-zero entries is recovered by solving (3.4), we have uniquely obtained the sparsest solution with high probability. In practice, relative sparsity ratios, $k/n < 0.1$ are observed in the datasets described in Section 2.1 and presented in Section 4.3.

The numerical subsampling tests provide confidence that we have recovered the sparsest solution if we have recovered a solution with sparsity less than 0.2. In other

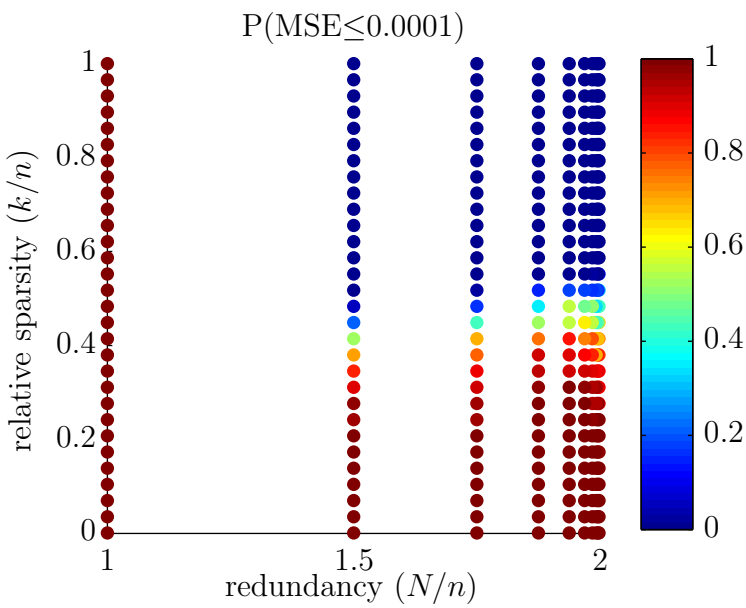

Figure 4.1: Probability that solving (3.4) given measurements y uniquely recovers the true sparse vector x^* for different choices of dictionary size, N, and sparsity k, with dictionary Φ taken from the truncated wavelet-Fourier combination as described in Section 4.2. Each dot represents the probability that the mean squared error over 70 Monte Carlo trials is less than 10^{-4}. High probability (red dots) mean the sparse vector can be uniquely found, low probability indicates that the sparse vector cannot be found by solving (3.4).

words, if we solve the sparse separation problem using real data, and we recover a solution with sparsity less than 0.2, we have, with high probability, recovered the sparsest solution to the optimization problem. The minimum cardinality problem and the ℓ_1 approximation give the same result in these cases.

4.3 Data Results

Stanford University operates VLF radio receivers at field sites around the globe, each with a varying degree of man-made interference such as power line harmonics and harmonics from various electro-mechanical systems. Mitigating this unwanted interference is often necessary before scientific analysis can proceed and is particularly important for maintaining the integrity of long-term datasets. Even though much of the data obtained during the course of this thesis were recorded using autonomous, battery powered systems, interference from nearby power systems is not negligible. These low frequency sources have the potential to obfuscate low frequency components of sferics commonly called the ELF tail. In the subsequent data, coarser wavelets are included so as to represent this low-frequency but sparse behavior. The finest 12 levels of the wavelet transform are used, *i.e.*levels $j = 9, 10, \ldots 20$. We separate the VLF data into short impulsive bursts, *i.e.*sferics, and steady tones, *i.e.*transmitters, power-lines, and strong interference using a dictionary composed of the Fourier basis and a truncated Daubechies-10 wavelet basis. Data vectors of length 2^{20} are used.

Figure 4.2 shows an example of sparse decomposition in VLF data recorded in Northern Idaho on August 27, 2009, at 03:20 UT. Spectrograms of the raw data, wavelet components, y_{wavelets}, and the Fourier components, y_{Fourier}, are shown in the top, middle, and bottom panels respectively. To illustrate and highlight the effectiveness of the separation, a shorter clip from the spectrograms is shown in the time domain in Figure 4.3. Again, the raw data, wavelet, and Fourier reconstructions are shown in the top, middle, and bottom panels respectively. The algorithm separates the sferics and the tones with minimal distortion to the underlying waveforms.

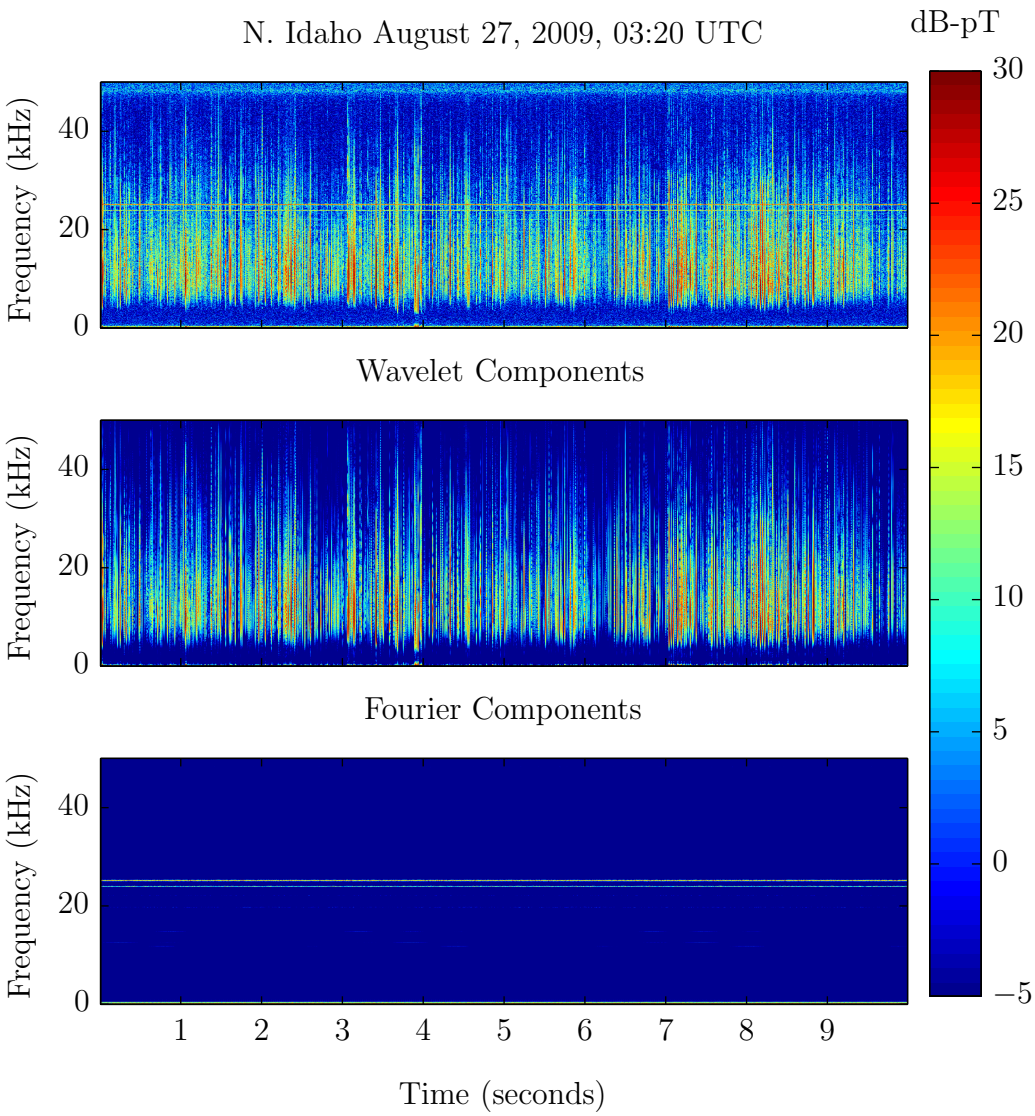

Figure 4.2: Separation of sample data from Northern Idaho, August 27, 2009. At the top is the spectrogram representation of the components of the data sparsely represented by wavelets. At the bottom is the spectrogram representation of the components of the data sparsely represented by Fourier series. Color indicates the amplitude of the signals in time-frequency space.

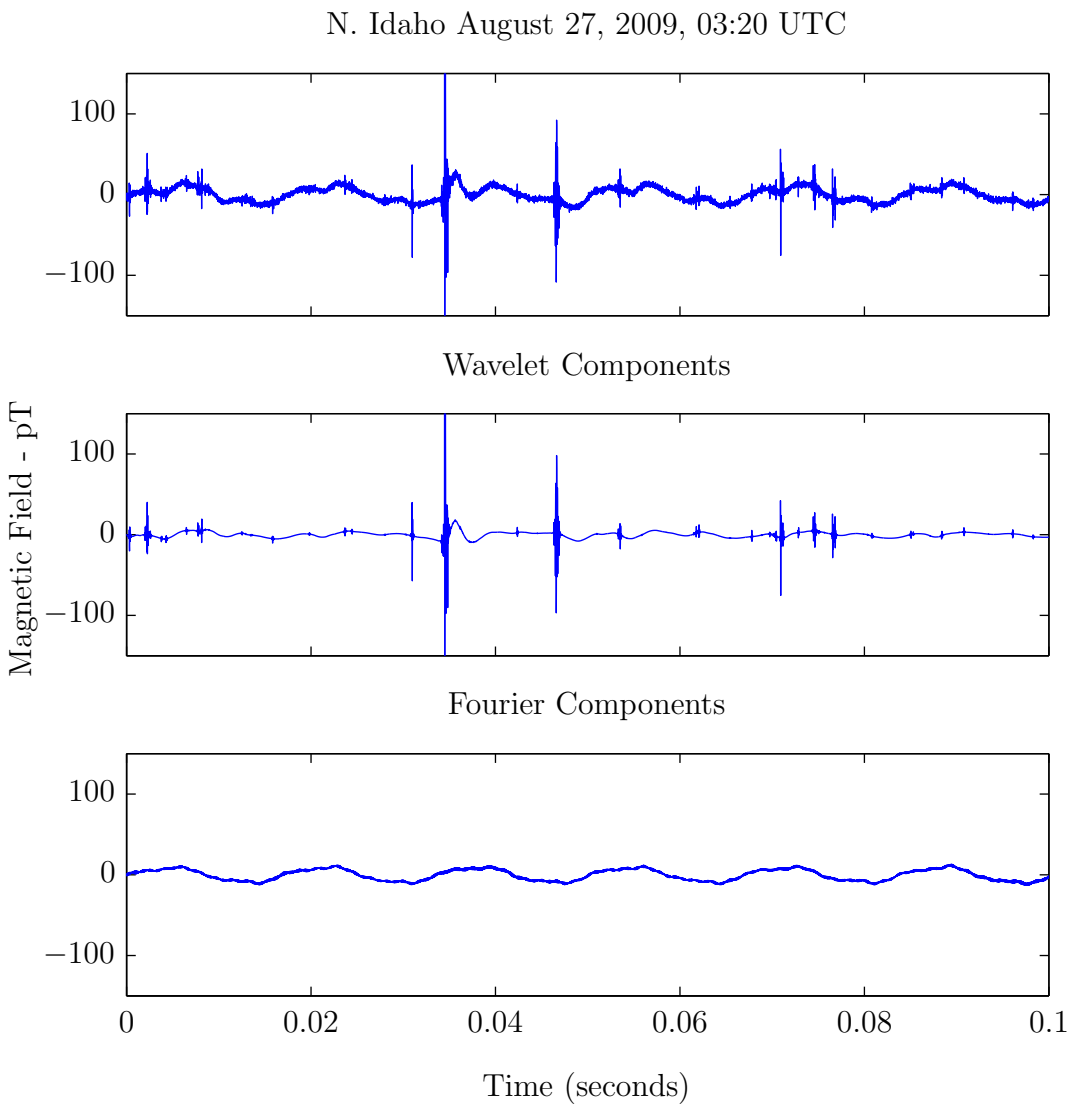

Figure 4.3: Separation of sample data from Northern Idaho, August 27, 2009. At the top is a clip of the calibrated time domain representation of the data. In the middle is the time domain representation of the recovered wavelet components. At the bottom is the time domain representation of the Fourier components.

Part III

Electromagnetic Subsurface Imaging

Chapter 5

Background

The third part of this thesis describes electromagnetic subsurface imaging. In this chapter, we present background information on wave imaging methods and describe the advantages for using VLF radio waves for subsurface conductivity imaging.

5.1 Noninvasive Wave Imaging

Techniques for imaging material structures by non invasive means have seen remarkable progress over the last few decades. Non invasive methods have been developed across the entire spectrum and use many different physics principles. Methods as diverse as X-Ray tomography and seismic prospecting can be seen as applications of non-invasive imaging methods. The topic is too broad to be summarized by any single source. More information on a large number of these topics can be found in sources [100], [83], [151], [106].

Geophysical exploration is a branch of remote sensing where measurements near the surface are used to determine the subsurface parameters in a non-destructive, non-invasive way. Geophysical exploration specifically refers to finding valuable resources underground such as mineral deposits and hydrocarbon deposits. These methods have been investigated and explored for many years hand have achieved a wide variety of results and yielded many computational challenges. While geophysical exploration is important in its own right, the methods of electromagnetic subsurface imaging have

applications in fields beyond geophysical exploration, and often literature in these other areas is general enough to be applied to geophysical exploration. We describe a few of the different aspects of electromagnetic imaging problems below.

5.2 Electromagnetic Inverse Problems

Electromagnetic methods are often split into different methods depending on the frequency of illumination. Table 5.1 shows a comparison of these various techniques based on frequency. A trade-off relationship exists between the frequency used for imaging, the depth of penetration of the method, and the resolution that can be obtained.

The skin depth, δ, of a material is given as the depth that the wave amplitude has decayed by a factor of e. The skin depth is a function of the frequency of the waves, ω, the conductivity of the material, σ (assumed to be homogenous), and the magnetic permeability, μ (also assumed to be homogenous and constant). The definition of skin depth, given in (5.1), shows the inverse relationship between the conductivity and skin depth and frequency and skin depth. Lowering the frequency increases the skin depth, while increasing the conductivity reduces the skin depth.

Electrical impedance tomography is a method that measures DC resistivity across a network of sensors in order to determine the resistivity of an object. Electrical impedance tomography has found applications both in medicine and geophysical exploration. These imaging problems are known to be severely ill-posed and often result in difficult systems of equations.

For the lowest frequencies, great penetration depths can be achieved. Indeed magnetotelluric methods have been used to image resources at depths of greater than 1 kilometer. In contrast, these methods cannot typically resolve objects smaller than half a kilometer in size. These methods have been very successful for understanding geologic processes [85].

Many ground penetrating radar techniques rely on knowledge of wave propagation in the subsurface. Echograms, or plots of the amplitude of the reflected waves as a function of depth, can be valuable for imaging the location of scatterers in a media.

Method	Frequency Range
Electrical Impedance Tomography	DC
Magnetotellurics	< 1 Hz
Audio Frequency Magnetotellurics	$1 - 50$kHz
Electromagnetic Subsurface Imaging (this work)	$3 - 30$ kHz
Ground Penetrating Radar	$3 - 3000$ MHz
Microwave Imaging	$1 - 3$GHz
Optical Coherence Tomography	optical frequencies

Table 5.1: Table of electromagnetic methods for non-invasive imaging and applications.

Ground penetrating radar techniques, in contrast, are able to image small, centimeter sized objects, but are constrained to imaging the first few meters of the surface of the Earth.

Electromagnetic wave inversion has also found application at microwave frequencies. Microwave methods (300 MHz-300 GHz) have also been developed. Much of the mathematical formality developed in subsequent chapters parallels the work at higher frequencies. Microwave imaging has been developed for medical imaging purposes and is designed to resolve structures on the order of millimeters, but only within a few centimeters of the skin surface [146, 125]. Using a linear absorption model of propagation, attenuation of microwave links has also been used for tomography specifically for rain detection [104].

At optical frequencies the same principle used for imaging has been applied to design. Nanophotonic components are increasingly difficult to design efficiently especially given space and material constraints. The electromagnetic photonics design problem seeks to find a realizable structure that matches electromagnetic fields modeled to the electromagnetic fields desired for the device. Rather than work with true observations, the optimization matches design criterion. Two dimensional and 2.5D structures have been the focus of much recent work [97], [66], [98].

5.3 VLF Imaging

Electromagnetic wave propagation in a conductive media is primarily characterized by loss. The skin depth is a common parameter that describes the depth in a homogeneous conductive medium at which a plane wave of frequency ω decays in amplitude by a factor of $e \sim 2.718$ and is given in equation 5.1.

$$\delta_{\text{skin depth}} = \sqrt{\frac{2}{\mu\sigma\omega}} \tag{5.1}$$

For $\sigma = 0.0018$ S/m and $f = 1$ kHz, the skin depth is 370 meters. At 13 kHz, the skin depth is 100 meters. These skin depths give an estimate of the depth at which we can accurately image underground. The exact depth is dictated by signal to noise ratios, the exact geometry of the subsurface materials, and the method for parameter inference.

The skin depth is derived from approximating the real component of the square root of the propagation constant,

$$k = \sqrt{\omega^2\mu\epsilon + i\omega\mu\sigma}. \tag{5.2}$$

By analyzing the propagation constant, k, over a variety of frequencies and conductivities, we can establish two distinct regimes. First is a regime where the term, $\omega^2\mu\epsilon$, dominates and electromagnetic propagation is wave-like. Second is a regime where the term, $i\omega\mu\sigma$, dominates which leads to exponential decay and diffusion-like behavior for the electromagnetic waves.

At VLF frequencies and conductivities representative of the surface of the Earth, propagation is mostly diffusive. The diffusive characteristic means that our ability to image objects at depth is greatly influenced by the loss of information imposed by attenuation. At some depth, radio waves are attenuated to a level below our minimum detection threshold and information from those depths is not available.

The split between diffusion and wave propagation can clearly be seen if we plot the real portion of the propagation constant as a function of both conductivity and frequency. A plot of this variation is shown in Figure 5.1.

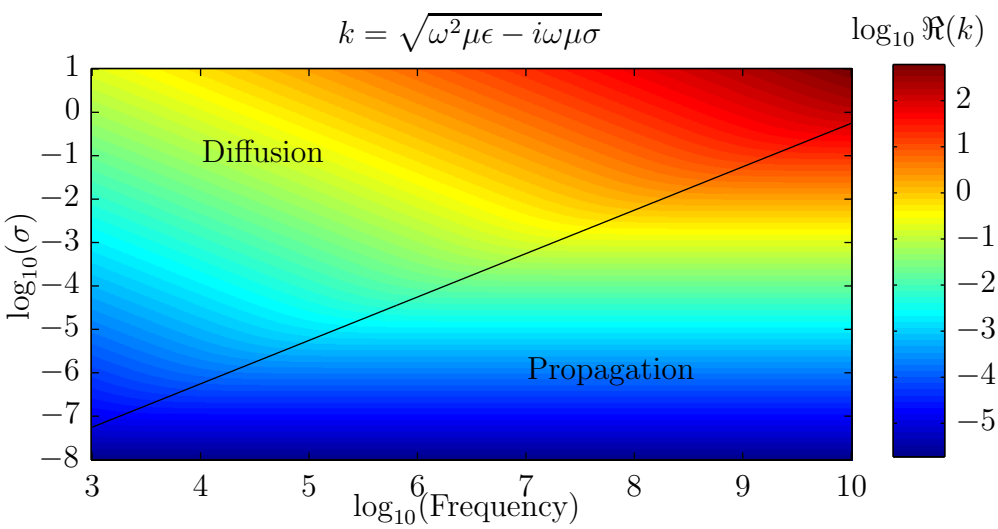

Figure 5.1: Real component of the wave propagation coefficient, $k = \sqrt{\omega^2\mu\epsilon - i\omega\mu\sigma}$. The division between propagation regimes and diffusion regimes can be seen as a function of both frequency and background conductivity.

Chapter 6

PDE Constrained Optimization

Electromagnetic subsurface imaging is an application of PDE constrained optimization. We search for a set of parameters in a set of partial differential equations such that the solution to the PDE system closely matches observations. In electromagnetic subsurface imaging, the parameters are subsurface electrical conductivities and the PDEs are Maxwell's equations which govern electromagnetic wave propagation.

PDE constrained optimization problems are typically split into two components, the forward solver and the inversion method. The forward solver is a routine that calculates the state of the PDE system given a fixed set of parameters. In this case, the forward solver is a solver for Maxwell's equations given a particular frequency, incident wave, and set of subsurface conductivities. The inverse problem, is the process of finding the unknown model parameters, *i.e.*, the conductivities. We set up our inverse problem as an optimization problem; we look for the optimal set of conductivity parameters in a model of Maxwell's equations such that the predictions of our model match observations. We explicitly incorporate two pieces of prior knowledge on the magnitude and shape of the conductivity parameters; the unknown conductivity parameters are restricted to lie in a known, bounded region in the subsurface and the values of the conductivity parameters are bounded, *i.e.*, we know the maximum and minimum values of conductivity perturbations.

An illustration of the PDE constrained optimization domain is given in Figure 6.1. We will refer to the variables in this diagram consistently throughout our development

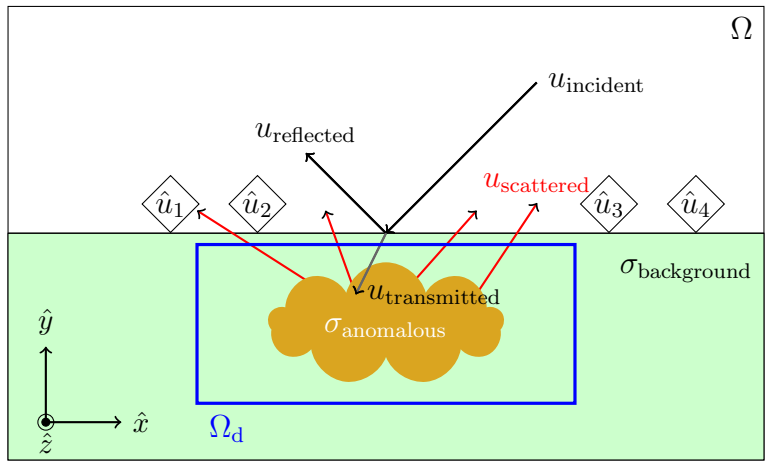

Figure 6.1: An illustration of electromagnetic subsurface imaging. Below the surface of the conductive Earth, $\sigma_{\text{background}}$, is a conductivity anomaly, $\sigma_{\text{anomalous}}$. Incident waves, u_{incident}, reflect off the surface, $u_{\text{reflected}}$, and penetrate into the subsurface, $u_{\text{transmitted}}$. The transmitted wave interacts with the anomalous material and produces the scattered wave $u_{\text{scattered}}$. Measurement of the electromagnetic waves at the surface provides the input for electromagnetic subsurface imaging algorithms which seek to estimate the magnitude and distribution of $\sigma_{\text{anomalous}}$.

of electromagnetic subsurface imaging.

We approximate the solution to Maxwell's equations in a bounded, finite domain, Ω, chosen to represent the space illustrated in Figure 6.1. The lower half of the space represents the conductive Earth; we assume a uniform background conductivity and homogeneous electromagnetic permittivity ϵ_0. While we have chosen a uniform background model, the finite difference frequency domain formulation is expressive enough to include any arbitrary background model that might include multiple subsurface layers and variable terrain. Given a set of anomalous conductivities underground, $\sigma_{\text{anomalous}}$, a governing frequency, ω, and a known incident wave, u_{incident}, we can calculate the electromagnetic fields everywhere with the model, specifically, the combination of $u_{\text{reflected}}$, $u_{\text{transmitted}}$, and $u_{\text{scattered}}$. In PDE constrained optimization, we search for the conductivities $\sigma_{\text{anomalous}}$ in the model that predict measurements made at the surface $\hat{u}$. The anomalous conductivities are restricted to a known reconstruction domain, $\sigma_{\text{anomalous}} \in \Omega_{\text{d}}$, $\Omega_{\text{d}} \subset \Omega$. The operator $\mathcal{M}_{\text{d}}$ is defined to be a linear

operator that maps from Ω to Ω_{d}.

The modeling domain is illuminated by a sequence of plane waves that we use as a first order approximation of sferics. Each sferic comes from a unique direction of arrival and contains broadband frequency content. We take the observation of a single sferic in the time domain to represent steady-state wave propagation. The spatial scales of this work are small with respect to the free-space wavelength of the illumination. Given that the model space is approximately $\frac{1}{12}$ of a wavelength, we assume that the steady state behavior is established quickly. We split each of the broadband sferic illuminations into its frequency components via the Fourier transform. We index the different frequency components and directions of arrival jointly by a single index, j. Each of the J incident conditions represents illumination by a plane wave from a specific incident direction and frequency. Each of these illumination conditions leads to an independent model of the complex electromagnetic field over the entire domain, $u_j \in \mathbb{C}^{\Omega}$. Furthermore, we assume that each of the frequency components is independent and there are no non-linear interactions in the physics model between frequencies and in the materials themselves.

Measurements of the field are taken at a set of R locations disjoint from the subsurface materials $\Omega_{\mathrm{r}} \subset \Omega, \Omega_{\mathrm{r}} \cap \Omega_{\mathrm{d}} = \emptyset$. The operator $\mathcal{M}_{\mathrm{r}}$ is defined as a linear operator that maps the full domain Ω to the receiver set Ω_{r}. We confine these observational locations to the surface of the conductive half space in our model. From each illumination condition, j, we make K independent, noise corrupted observations of the electromagnetic field, $\hat{u}_{jk} \in \mathbb{C}^{\Omega_{\mathrm{r}}}$. We assume an additive Gaussian noise model:

$$\hat{u}_{jk} = \mathcal{M}_{\mathrm{r}} u_{jk} + \nu, \quad \nu \sim \mathcal{N}(0, \Sigma^2). \tag{6.1}$$

6.1 The Discrete Model

There are many methods for solving or approximating the solution to Maxwell's equations for inhomogeneous environments. Two distinct classes exist for modeling electromagnetic fields. Time domain methods find the solution to Maxwell's equations as a direct function of time and include methods such as the finite diference time

domain method (FDTD) [136], [79], finite volume (FV) methods [75], and discontinuous Galerkin methods [64]. Frequency domain methods solve Maxwell's equations for a steady-state configuration at a single frequency and include Green's functions [36], finite difference frequency domain (FDFD) [30], finite element methods [148], and multi-pole methods [129]. These frequency domain methods involve solving typically large systems of linear equations in order to predict the electromagnetic waves at a particular set of locations given a specific incident wave and material configuration. All of these methods have been applied to model electromagnetic processes at VLF wavelengths in conductive media [64], [34], [102], [30].

We select FDFD methods for approximating the solution to Maxwell's equations for two reasons: (1) the systems of equations for finite difference frequency domain methods are simple to set up allowing us to focus on aspects of the inverse problem and (2) FDFD methods are more flexible than Green's function methods for incorporating arbitrary inhomogenous media.

6.1.1 Maxwell's Equations

We discretize the domain, Ω, in a regular grid and use the FDFD method to approximate the solution to Maxwell's equations (6.5). After substituting in the constituent relationships $D = \epsilon E$ and $\mu H = B$, we solve Maxwell's equations on a regular, charge free ($\rho = 0$) grid.

$$\nabla \cdot E = \frac{\rho}{\epsilon} \tag{6.2}$$

$$\nabla \cdot H = 0 \tag{6.3}$$

$$\nabla \times E = -\mu \frac{\partial H}{\partial t} + M_s \tag{6.4}$$

$$\nabla \times H = J_s + \sigma E + \frac{\partial \epsilon E}{\partial t} \tag{6.5}$$

We transform the last two equations to the frequency domain by taking their Fourier transform, with $e^{-i\omega t}$ time dependence:

$$\nabla \times E = i\omega\mu H + M_s \tag{6.6}$$

$$\nabla \times H = \sigma E + i\omega\epsilon E + J_s. \tag{6.7}$$

On a regular grid, and in charge-free simulations, it can be shown that the divergence equations are satisfied in the solution of Maxwell's curl equations [136].

Using the FDFD method, we discretize in space using a simple finite difference stencil:

$$\frac{\partial u}{\partial x} \simeq \frac{u_i - u_{i+1}}{\Delta x} \tag{6.8}$$

In one dimension, we can represent this operator as a matrix-vector operation. We represent the operator ∇ with a matrix:

$$\nabla_x = \frac{1}{\Delta x}\begin{bmatrix} -1 & 0 & 0 & 0 & \dots & & \\ 1 & -1 & 0 & 0 & \dots & & \\ 0 & 1 & -1 & 0 & \dots & & \\ \vdots & & & & \ddots & & \\ & & & & & 1 & -1 \\ & & & & & 0 & 1 \end{bmatrix} \tag{6.9}$$

Dirichlet boundary conditions have been implemented; the boundary values of the operand are set to zero.

6.1.2 Boundary Conditions

To truncate the simulation domain, we use the perfectly matched layer (PML) technique [10]. The original formulation of the PML introduces a special material at the boundary of the domain to absorb outgoing waves and is commonly known as the uniaxial PML. For the FDFD models in this work, the stretched coordinate PML was used instead [126]. For the stretched coordinate PML, we modify the derivative

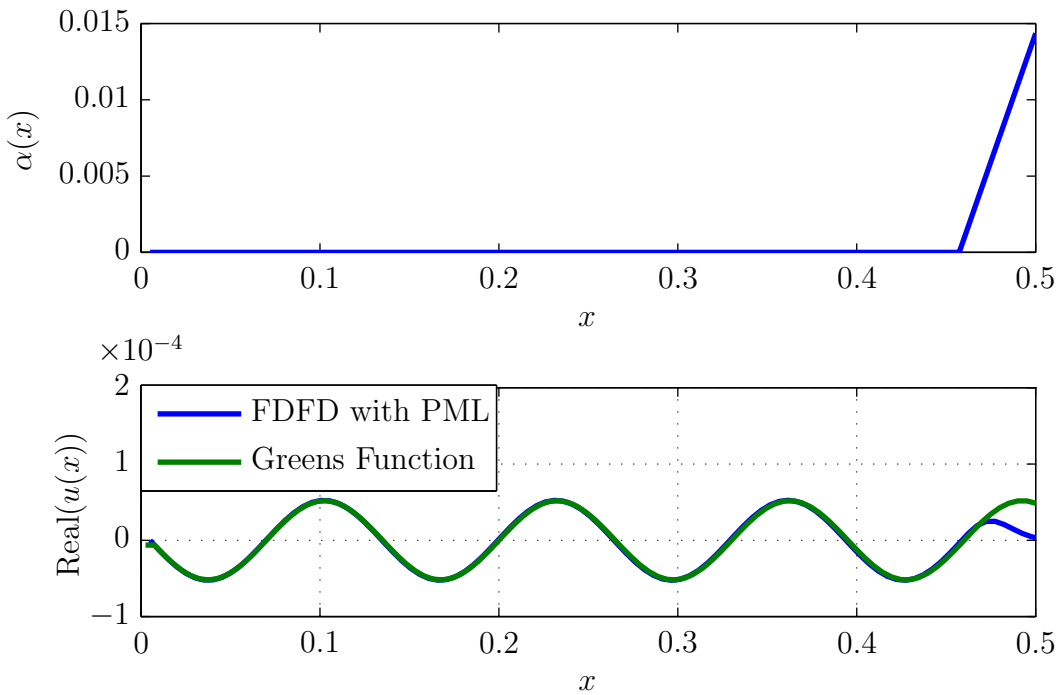

Figure 6.2: At the top, a plot of the space stretching parameter α as a function of location in the right half of the simulation domain. Only the right half of the simulation domain is shown; the left half is a mirror image. At the bottom, a comparison of the analytic Green's function solution and the FDFD solution of the 1D electromagnetic field from a point source is plotted. The affect of adding the space-dependent parameter α can be seen in the solution of the FDFD method near the domain boundary.

operator [113], [38]. We define the stretched coordinate derivative operator as

$$\tilde{\nabla} = \frac{1}{1+i\alpha(x)}\frac{\partial}{\partial x} + \frac{1}{1+i\alpha(y)}\frac{\partial}{\partial y} + \frac{1}{1+i\alpha(z)}\frac{\partial}{\partial z}. \tag{6.10}$$

Adding this additional term to the derivative has the effect of introducing an attenuation term that decreases the wave amplitude near the boundary of the computational domain. If the loss term is chosen correctly, the wave is attenuated to zero at the boundary and Dirichlet boundary conditions can be enforced with no reflections.

An example 1D simulation that compares the analytical Green's function solution

for an outgoing plane wave to the outgoing wave computed using the FDFD technique with a PML absorbing boundary condition is shown in Figure 6.2. The FDFD solution closely matches the analytic solution given by the Green's function. Near the boundary, the absorbing layer attenuates the outgoing wave so that the amplitude is zero at the boundary and there are no reflections.

The PML parameter α increases linearly at a fixed rate and is chosen via a brute force search. The rate of increase is a strong function of both the frequency of simulation ω, and the grid size Δx. A database of these parameters was computed and kept for reuse between simulations.

6.1.3 Extension to Multiple Dimensions

The discussion of finite difference operators has been restricted to one dimensional models thus far. Extending the model to 2D, and even 3D can be done in a straightforward manner, however with each increase in dimension, the size of the system increases exponentially.

To reach more dimensions, we use simple matrix operators. Conceptually it is easy to think of the 2D grid of electromagnetic field values as a doubly-indexed array, `x[i,j]`. Unwrapping this 2D array, *i.e.*, stacking all of the columns of the array, gives us a single 1D vector representing a 2D field. In other words, we represent the electromagnetic fields on our grid as an $n^2 \times 1$ vector rather than an $n \times n$ array. We express the derivative operators then as a matrix operation acting on this vector representation of the grid.

$$\nabla_{xy} = I^n \otimes \nabla_x + \nabla_y \otimes I^n \tag{6.11}$$

With these operations, it is simple to build large matrices for representing Maxwell's two curl equations. The PML described previously can be incorporated into the 1D operators and thus extended to two dimensions in a straightforward manner. Constructing this gradient operator produces a large, sparse, asymmetric matrix.

6.1.4 Plane Wave Excitation

We model sferics as a superposition of incident plane waves at multiple frequencies. A coordinated set of sources can be introduced to the domain as J_s and M_s as given in equation 6.7. The total field, scattered field technique (TF/SF), [136], [79], provides a method for introducing a known source field, a plane wave in this case, to the interior of the simulation space. The TF/SF method also subtracts the field from the exterior of the scattering domain to minimize reflections and interactions with the boundary layers. Arbitrarily directed waves can be introduced of either transverse electric (TE) or transverse magnetic (TM) polarization using this method.

6.1.5 Abstract Model

The FDFD model results in a linear system of equations for any given choice of $\omega, \epsilon, \sigma, \mu$, wave polarization, and incident direction. Explicitly, the model is a matrix equation:

$$\begin{bmatrix} i\omega\epsilon + \sigma & \nabla\times \\ \nabla\times & i\omega\mu \end{bmatrix} \begin{bmatrix} E \\ H \end{bmatrix} = \begin{bmatrix} J_s \\ M_s \end{bmatrix}. \tag{6.12}$$

It is easier to represent this complicated set of terms as an abstract model parameterized by the conductivities σ and indexed by the frequency and incident wave direction, j:

$$A_j(\sigma)u_j = s_j. \tag{6.13}$$

We combine the representation of the electric and magnetic fields into a single term that represents both, u. The collection of sources used to describe an incident plane wave collapses into a single term s, and the matrix relationship involving spatial derivatives and material properties is simply represented by a matrix A.

6.1.6 Scattered Fields

The parameter σ is only partially known in these systems of equations. The goal of subsurface electromagnetic imaging is to recover the proper values. If the matrix A

incorporates the background, known model of the space, then we can add a conductivity perturbation with a diagonal matrix. We simplify this equation as shown in Equation 6.14

$$\begin{aligned} \left[A + \mathbf{diag}\left(\begin{bmatrix}\sigma \\ \mathbf{0}\end{bmatrix}\right)\right] u &= s \\ Au + \sigma \circ u &= s \end{aligned} \tag{6.14}$$

The solution to this equation is an electromagnetic field that we describe as the superposition of two fields, the background field, u_{b} and the scattered field, u_{s}, where u_{b} satisfies the relationship of Equation 6.15. The circle ($\circ$) represents the Hadamard product or entry-wise multiplication between two vectors.

$$Au_{\mathrm{b}} = s. \tag{6.15}$$

By combining (6.14) and (6.15) and our definition of the background and scattered fields, we obtain Equation 6.16:

$$A_{\mathrm{s}}u + \sigma \circ u_{\mathrm{s}} = 0. \tag{6.16}$$

By Equation 6.16, we have rearranged and simplified the linear algebra to the point that very few terms are involved so that we can highlight the key mathematical form of the constraints imposed by Maxwell's equations.

6.2 Optimization

In order to solve the subsurface imaging problem, we are after the optimal parameters σ that match the observations $\hat{u}$ to the predictions of the model u. These relationships must be satisfied for a host of models at different frequencies and incident directions of arrival which are jointly indexed by j. Equation 6.17 gives electromagnetic subsurface

imaging as an optimization problem in standard form.

$$\begin{aligned} \underset{\sigma}{\text{argmin}} \quad & \sum_{j}^{J} \|\hat{u}_j - \mathcal{M}_{\mathrm{r}} u_j\|_2^2 \\ \text{subject to} \quad & A_j u_j + (\mathcal{M}_{\mathrm{d}}^T \sigma) \circ u_j = 0 \quad \forall j = 1, \ldots, J \end{aligned} \tag{6.17}$$

This optimization problem is not convex because of the $\sigma \circ u$ term in the constraint equation. Since we do not have a guarantee on the sign of both σ and u, the set described by this equation is not convex as can be easily seen by evaluating the Hessian of the equation to reveal a indefinite matrix. Due to the non-convexity of the problem, we cannot demonstrate in polynomial time that a unique optimum has been obtained. Equation 6.17 is an optimization problem, specifically a least squares problem, constrained by a set of PDEs that describe Maxwell's equations.

6.3 Prior Work

Algorithms for approximating the solution to the electromagnetic inverse problem are not new. PDE constrained optimization problems have been studied in many contexts from optimal control of fluids to airplane design. The topic of PDE constrained optimization is too vast to review in great depth here, but we give an introduction to the key concepts and general trends specifically as they apply to electromagnetic subsurface imaging. Many aspects of large PDE constrained optimization problems are covered in [14]. Further reviews of electromagnetic inverse problems can be found in [3], while [61] provides a more detailed look at the frameworks necessary for electromagnetic inversion, data aggregation, the forward solver, and the inverse method. A mathematical review of inverse problems and regularization can be found in [81].

The main challenges of PDE constrained optimization problems are that they often result in large, poorly-conditioned systems of equations. As mentioned before, these methods must estimate the parameters describing the system in addition to the state of the system. In inverse problems such as electromagnetic subsurface imaging and design problems, both the electromagnetic field in the simulation space must be estimated in addition to the material parameters of interest. In optimal control

problems, the driving sources must be determined in addition to the subsequent state of the optimization problem. Furthermore, multiple criterion, multiple aspects of the model may be of importance including frequency and or time dependence which requires that the state of the PDE must be expanded to include the full frequency or full time domain of interest which often leads to large spaces of unknown variables.

6.3.1 Model Selection

One of the crucial aspects of the electromagnetic inverse problem is the computational model of the physics and the model of the Earth. The simplest model of the Earth is a layered 1D model. This model has been used for magnetotelluric inverse problems for many years [106]. Despite its simplicity, the 1D model provides many reasonable computational advantages and continue to be used and developed [121], [118], [43]. The electromagnetics of a 1D model are simple compared to the more complicated 2D and 3D models and the electromagnetic fields can be computed analytically. If the number of layers is known ahead of time, these models can be solved with relatively low computational complexity. However, determining the number of subsurface layers in addition to the conductivities and depths is computationally difficult.

Pixel based models are a class of models where the model parameters at each spatial point in the computational domain are modeled independently. These points could be grid points in a finite difference method, polygons in a finite element method, or grid points in a Green's function method. These models offer the most flexibility in terms of specifying the geometry of subsurface materials. As described in Section 6.1, we opt for a pixel based model.

The majority of pixel based methods are frequency domain methods; the forward solvers operate at a single frequency and a single excitation. Time domain physics models have also been considered and developed for electromagnetic inverse imaging [75]. Time domain methods are difficult because they lead to very large systems of coupled equations that change at every iteration. While these matrices often have favorable block structure, the size of the systems may still be computationally prohibitive. Many of these difficulties are addressed by using limited memory methods

such as L-BFGS [75], [111].

6.3.2 Born Approximation

Scattering and perturbation problems have long been associated with approximate methods. In electromagnetics, in cases where the magnitude of the scattering perturbation is small, the Born approximation is often used to approximate the solution to the electromagnetics [36]. The Born approximation is a linearization of the non-linear interaction term between the material parameters and the electromagnetic field. It is typically used in situations where approximate solutions are desired quickly [71]. Pre-computed solutions of the Born approximation can also be used in reduced order methods to compress the set of equations necessary for electromagnetic imaging [29].

The Born approximation is often expanded with respect to the Green's function integral equation approach to electromagnetic modeling, but we can easily apply the same concept to our implicit finite difference model of Maxwell's equations.

$$Au + \sigma \circ u = b \tag{6.18}$$

We expand our generalized form of Maxwell's equations (6.18) about a perturbation δ_σ, δ_u:

$$\begin{aligned} A(u + \delta_u) + (\theta + \delta_\sigma) \circ (u + \delta_u) &= b \\ Au + A\delta_u + \sigma \circ u + \sigma \circ \delta_u + \delta_\sigma \circ u + \delta_\sigma \circ \delta_u &= b. \end{aligned} \tag{6.19}$$

Assuming that both the magnitude of the field perturbation δ_u and the parameter perturbation δ_σ are small, we can ignore the influence of the second order term, $\delta_\sigma \circ \delta_u$. Without this term, the equation is linear in δ_σ, δ_u. We restate the Born approximation concisely:

$$Au + A\delta_u + \sigma \circ u + \sigma \circ \delta_u + u \circ \delta_\sigma \simeq b. \tag{6.20}$$

Three variants of the Born approximation have been used for electromagnetic subsurface imaging, specifically the Born Approximation (BA), Born Iterative Method (BIM), and Distorted Born Iterative Method (DBIM) [47], [35]. These three methods rely on the Green's function formulation of Maxwell's equations. A full description of

Green's functions in electromagnetics is beyond the scope of this thesis; more details can be found in [36]. The Born approximation (BA) method refers to solving the inverse problem once; to simply use the linear model to predict the scatterers. The Born Iterative method solves the linear problem multiple times in order to get the best approximation for the scatterers. The Distorted Born Iterative Method solves for small perturbation iteratively and updates the model equations between iterations. This method is more computationally difficult than the other two, but is known to produce better results [46], [72].

6.3.3 Newton Based Methods

General non-linear optimization techniques have also been applied to the electromagnetic inverse problem. These methods form a non-linear set of equations that define the optimal point of the optimization problem and proceed by using Newton's method or an approximate Newton's method in order to solve the non-linear equations.

Electromagnetic subsurface techniques involving Newton methods fall into two generic categories: unconstrained and constrained optimization problems. In unconstrained optimization problems, the system of equations that governs electromagnetic radio propagation are folded into the objective function. The optimization problem then becomes a non-linear least squares problem which can be approached using Gauss-Newton methods. A minimum is reached by taking steps in the negative gradient direction of the objective function [61]. The function relating the conductivities to the field measurements is often stiff and its derivatives which are necessary for Gauss-Newton methods, both the Jacobian and the Hessian, are costly to evaluate computationally [109], [110], [152], [120], [2], [128]. Methods for calculating the gradients using adjoint operators have also been developed [58].

The model can also remain as part of the constraint set leading to constrained optimization methods [73, 74, 15, 1, 87, 140]. A wide variety of numerical techniques have been applied to both constrained and unconstrained problems including non-linear conjugate gradients [120], Gauss-Newton iterations [2], and augmented Lagrangians [1]. [73] develops many techniques for preconditioning the system of equations solved

at each iteration based on previous factorizations and Hessian approximations. These preconditioning methods were explored in great detail for application to the methods we develop in Section 7.3.2 but were ultimately unsuccessful for numerical reasons. [86] adapted Gauss-Newton to include box-constraint sets similar to the constraint sets explained in Section 7.1. [92] introduce a method that combines the simplicity of proximal functions for non-smooth constraints and the efficiency and fast-convergence properties of a Newton method. Full discussion and development of these techniques is beyond the scope of this thesis.

Methods with increasing sophistication have also been developed such as the parallel Lagrange-Newton-Krylov-Schur method for PDE constrained optimization [16, 15]. This method has only been applied to flow control problems, estimating the proper inputs to a PDE constrained system, not finding internal parameters of the system. The method is general and parallel, but does not explicitly exploit problem structure in its parallelism. These methods have been extended to parameter estimation in time domain wave inverse problems as well [7]. Again, these methods do not exploit problem structure for parallel decomposability.

6.3.4 Contrast Source Method

The contrast source method is an extension of the unconstrained techniques [11]. The contrast source technique introduces a new term, the contrast-source term, χ, that represents the scaled local product of the subsurface fields, u, and the subsurface material properties, σ and ϵ. The objective function is split into two terms, one governing the state of the PDE, the other governing the state of the contrast source. Minimization of the objective function proceeds by alternately updating u and χ using conjugate gradients. Gradient vectors are computed by solving the adjoint electromagnetics system. The contrast source inversion technique has been extended to include finite difference methods [4], finite element methods [125], and multiplicative regularization costs [149]. The contrast source method has been adapted for low-frequency subsurface imaging applications as well [5].

6.3.5 ADMM in PDE-Constrained Optimization

Many of the techniques that we develop in Section 7.3.2 use the Alternating Direction Method of Multipliers, a convergent technique for solving convex optimization problems [20]. ADMM methods have been adapted to optimal control problems that require non-smooth constraints in an elegant way [112]. These techniques offer decomposability and parallelization based on the structure of the problem. ADMM techniques have been applied to parameter estimation problems in diffusion equations [80], [22] and optical diffusion equations [1].

Chapter 7

Algorithms

In the last chapter, we motivated electromagnetic subsurface imaging as a PDE constrained optimization problem. In this chapter, we will analyze the form of the optimization problem and describe several algorithms that can be used to approximate the solution to this optimization problem. Finding the subsurface parameters requires jointly estimating the material parameters $\sigma \in \mathbb{R}^{\Omega_{\mathrm{d}}}$ and the full state, $u \in \mathbb{C}^{\Omega}$. We state the optimization problem again for convenience.

$$\begin{aligned} \underset{\sigma}{\text{argmin}} \quad & \sum_{j}^{J} \|\hat{u}_j - \mathcal{M}_{\mathrm{r}} u_j\|_2^2 \\ \text{subject to} \quad & A_j u_j + (\mathcal{M}_{\mathrm{d}}^T \sigma) \circ u_j = 0 \quad \forall j = 1, \ldots, J \end{aligned} \tag{7.1}$$

7.1 Noise and Regularization

For each of the J illumination conditions, many independent, noisy measurements are made (6.1), $\hat{u}_{jk}$, where j indexes the illumination condition and k indexes the acquisition number. Minimizing the least squares distance between the observations $\hat{u}$ and the model predictions u, can also be interpreted as minimizing the negative expected log-likelihood of the error. We define the objective function as:

$$f(u_j) = \mathop{\mathbf{E}}_{k} \left[\|\mathcal{M}_r u_j - \hat{u}_{jk}\|_2^2 \right]. \tag{7.2}$$

This expression simplifies; incorporating information from k different independent observations of the same illumination can be handled by simple averaging. Therefore, random noise associated with the measurement process is reduced as more independent observations are made. Henceforth, we let $\hat{u}$ generally represent the mean of many observations made at a particular incident wave condition.

It is important to note that a single set of conductivities, σ, is common to all of the J electromagnetic models. All of the different frequencies and illuminations agree on one set of frequency-independent parameters, σ, for the subsurface materials. As stated, this problem is ill posed, meaning there may not be a unique set of parameters σ that will satisfy this equation. It is typical to include a regularization penalty in the objective function. Quadratic regularization or ℓ_2-norm regularization is typical, but total variation norms and ℓ_1 norms have also been considered, [4], [146]. Our form of regularization takes the restriction that σ is uniformly distributed in a bounded region, $\sigma \sim \mathcal{U}[0, \gamma]$, a non-linear, non-quadratic constraint.

$$\begin{aligned} \sigma_{\text{opt}} = \underset{\sigma \in \mathbb{R}}{\text{argmin}} \quad & \textstyle\sum_k \|\mathcal{M}_{\text{r}} u_j - \hat{u}_j\|_2^2 \\ \text{subject to} \quad & A_j u_j + (\mathcal{M}_{\text{d}}^T \sigma) \circ u_j = s_j \quad \forall\, j = 1, \ldots, J \\ & 0 \leq \sigma \leq \sigma_{\max} \end{aligned} \tag{7.3}$$

The optimization problem given in (7.3) is the most complete form of electromagnetic subsurface imaging. All subsequent algorithms directly address this optimization problem.

7.2 Analysis of Optimization Problem

The electromagnetic subsurface imaging problem is difficult for two reasons: first, it is not convex, second, it is large and poorly conditioned.

This optimization problem (7.3) is not convex since the model constraint is not jointly convex in the model parameters, σ, and the model state, u. If we represent this linear equation as a function, $g(u, \sigma) = Au + \sigma \circ u$, we can derive the Hessian with respect to u and σ. The Hessian of this function is indefinite which means it has both positive, negative, and complex eigenvalues which ensures that it is not a

convex constraint.

$$\nabla^2 g(u,\sigma) = \begin{bmatrix} \frac{\partial^2 g}{\partial u^2} & \frac{\partial^2 g}{\partial u \partial \sigma} \\ \frac{\partial^2 g}{\partial \sigma \partial u} & \frac{\partial^2 g}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} 0 & \mathcal{M}_{\mathrm{d}}^T \\ \mathcal{M}_{\mathrm{d}} & 0 \end{bmatrix} \tag{7.4}$$

Therefore, we cannot provide a guarantee that a local optimizer is a global optimizer without finding each of the local optima.

7.2.1 Reduction

Equation 7.1 can be reduced to an even simpler form, allowing us to focus on the salient features of the structure of the problem. Elimination clarifies the problem, but in many practical implementations of the problem, such elimination is not efficient and leads to longer solve times and increased memory requirements.

Maxwell's equations are incorporated into the problem as constraint equation,

$$Au + (\mathcal{M}_d \sigma) \circ u = s.$$

The operator $\mathcal{M}_d$ maps the parameters σ to their proper locations in the full discretization grid, Ω. The number of parameters, d, is typically less than the number of variables that describe the field solutions of the PDE, u. We can separate u into two components, locations where there are parameters to estimate, u, and locations where the parameters are fixed $\bar{u}$,

$$A \begin{bmatrix} u \\ \bar{u} \end{bmatrix} + \begin{bmatrix} \sigma \circ u \\ 0 \end{bmatrix} = \begin{bmatrix} s \\ \bar{s} \end{bmatrix}.$$

Partitioning Maxwell's equations in to four components, we get

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} u \\ \bar{u} \end{bmatrix} + \begin{bmatrix} \sigma \circ u \\ 0 \end{bmatrix} = \begin{bmatrix} s \\ \bar{s} \end{bmatrix}. \tag{7.5}$$

Via the Schur complement, we can solve for $\bar{u}$ in terms of u and eliminate the sampling

operator from the problem,

$$\begin{aligned}(A - BD^{-1}C)u + \sigma \circ u &= s - BD^{-1}\bar{s} \\ \bar{u} &= D^{-1}(\bar{s} - Cu)\end{aligned} \tag{7.6}$$

Given any subset of conductivities, σ, that does not span the full space, Ω, we can rearrange the equations to obtain a set of equations that removes the dependency of the reconstruction domain sampling operator, $\mathcal{M}_{\mathrm{d}}$. Taken together, we obtain the reduced form of the optimization problem:

$$\begin{aligned}\text{minimize} \quad & \|\hat{u} - \mathcal{M}_r D^{-1}(\bar{s} - Cu)\|_2^2 \\ \text{subject to} \quad & (A - BD^{-1}C)u + \sigma \circ u = s \\ & 0 \leq \sigma \leq \sigma_{\max}\end{aligned} \tag{7.7}$$

We can finally abstract (7.7) to the most generic form of the problem, where we can incorporate $D^{-1}s$ into $\hat{u}$ and define the measurement operator, $\mathcal{M} = \mathcal{M}_r D^{-1}C$

$$\begin{aligned}\text{minimize} \quad & \|\hat{u} - \mathcal{M}u\|_2^2 \\ \text{subject to} \quad & Au + \sigma \circ u = s \\ & 0 \leq \sigma \leq \sigma_{\max}\end{aligned} \tag{7.8}$$

In the cases of interest, the imaging domain and the measurement domain are disjoint; we do not directly measure the fields in the grid locations we are imaging. Therefore, the measurement matrix $\mathcal{M}$ is not equivalent to the identity matrix, and is frequently a dense matrix. This analysis makes the relationship between the unknown model parameters, σ and the fields, u very clear; furthermore it allows us to rewrite any of the imaging problems in general form. In the implementation of many solvers, it is not necessary to explicitly take this elimination step. Evaluating $D^{-1}C$ gives rise to a dense matrix which can be less efficient to work with than the sparse matrices A.

7.2.2 Signs

As mentioned, in the general form of the problem (7.3) is not convex, however in the reduced form, (7.8) we can see that if the sign of u is known, then the problem becomes convex. If both u and σ are positive, the constraint $Au + u \circ \sigma = s$ is equivalent to the linear-fractional family of functions which are known to be convex [18].

At this point, we make the distinction between two families of problems, those that exist entirely on the real plane, where $u, A, \sigma \in \mathbb{R}$ and those problems that exist on the complex plane, where $u, A \in \mathbb{C}, \sigma \in \mathbb{R}$. In the real-only case, knowing the sign of u allows us to rewrite the problem as a convex problem. In the complex case, knowing the phase angle, $\arg(u)$, allows us to rewrite the problem as a convex problem.

In the real-only case, we introduce the new variable, $x_k \in \mathbb{R}, x_k = u_k\sigma_k$ so that we have $Au + x = s$. Given $u_k > 0$, we can see that $0 \leq x_k/u_k = \sigma_k \leq \sigma_{\max}$. Finally, this can be split into two affine relations:

$$0 \leq x_k \quad x_k \leq \sigma_{\max} u_k.$$

As is necessary, if the sign of u_k is known to be negative, the relations become

$$0 \geq x_k \quad x_k \geq \sigma_{\max} u_k.$$

Incorporation of this knowledge gives us a convex problem. Of course if the number of unknowns exceeds the number of measurements, the optimal solution, but the optimal value of the objective function does have a unique value.

In the complex case, we need to know the phase of u, *i.e.*$\arg(u)$ to make the problem convex. We let $\phi = \arg(u)$ and introduce $r_k \in \mathbb{R}$

$$r_k = \sigma_k |u_k| = \sigma_k \Re(u_k e^{-i\phi_k}). \tag{7.9}$$

Then, using complex phases, we have that

$$0 \leq e^{-i\phi_k} r_k / |u_k| \leq \sigma_{\max},$$

which can be split into two inequalities

$$0 \leq r, \quad r \leq \Re(e^{i\phi} \circ u)\sigma_{\max}.$$

By incorporating these inequalities, we have convex constraints in the subsurface electromagnetic problem and therefore can solve for a unique optimal value of the problem.

7.2.3 Uniqueness

Even if the phase is determined in the imaging domain, we are still not guaranteed to obtain a unique solution. Proper sampling of the field u is necessary in order to uniquely determine the parameters σ. In the real-only case, each unknown σ needs to be matched with a measurement. In the complex case, each complex measurement $\hat{u}$ matches with two real-only values of σ. Even when the linear bounds are imposed on the parameter σ, these subsampling ratios are required. If the number of observations is less than the number of model parameters, we can identify a polytope of feasible solutions that yield identical objective functions. In other words, we can identify a feasible null space of the measurement operator.

Imagine the case where the number of observations matches the number of unknown model parameters, $m = n$, and the measurement operator, $\mathcal{M}$ is full-rank and invertible. In this case, we can chose $u = \mathcal{M}^{-1}\hat{u}$ to obtain the minimal objective function value of zero. The unknown parameters σ can be found directly after substitution back into the set of constraint equations. Provided the observations come from a strictly feasible solution, an optimal and unique set of parameters σ can be obtained.

If more observations than unknowns are available, $m < n$, then we can obtain all of u via a least squares projection: $u_{\text{opt}} = (\mathcal{M}^T\mathcal{M})^{-1}\mathcal{M}^T\hat{u}$, which we assume to be

full rank, therefore, there are no vectors u_{null} such that $\mathcal{M}u_{\text{null}} = 0$ other than the trivial vector, $u = 0$.

In the case that there are fewer observations than parameters, the measurement operator $\mathcal{M}$ contains a non-trivial null space that contains feasible solutions. In this case, $\mathcal{M}$ is not invertible, but it has a pseudo inverse, $\mathcal{M}^\dagger$. Still, there are vectors $\nu : \mathcal{M}\nu = 0, \nu \neq 0$. If we choose $u = \mathcal{M}^\dagger \hat{u}$, we have a family of solutions, $u + \nu$ that give an equivalent objective function. For any of these solutions, we can find a set of parameters σ that satisfy the following relationships:

$$0 \leq \theta_i = b_i - \frac{a_i^T \mathcal{M}^\dagger \hat{u} + a_i^T \nu}{(\mathcal{M}^\dagger \hat{u})_i + \nu_i} \leq \gamma$$

We obtain a feasible, bounded region for equivalent solutions that solve the optimization problem.

7.2.4 Quadratically Constrained Quadratic Programming

Further analysis of the problem allows us to represent the problem in a different form which opens up more solver possibilities. The analysis begins with the constraint relationship:

$$\sigma \circ u = w.$$

This relationship can equivalently be written in quadratic form,

$$\begin{bmatrix} u \\ \sigma \end{bmatrix}^T \begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} u \\ \sigma \end{bmatrix} - w = 0. \tag{7.10}$$

The optimization problem given by Equation 7.1 can be rewritten as an indefinite quadratically constrained quadratic program. Indefinite QCQP's are known to come from a family of NP-hard problems. Thus the true difficulty of solving the electromagnetic subsurface imaging problem is clear; at best we can only hope to approximate the solution in polynomial time.

The constraint equation representing Maxwell's equations, $Au + \sigma \circ u = s$, can be

re-expressed as a sequence of quadratic constraints over the joint space $\begin{bmatrix} u & \sigma \end{bmatrix}^T$:

$$a_k^T u + \begin{bmatrix} u_k \\ \sigma_k \end{bmatrix}^T \begin{bmatrix} 0 & 1/2 \\ 1/2 & 0 \end{bmatrix} \begin{bmatrix} u_k \\ \sigma_k \end{bmatrix} = s_k.$$

The matrix, $Q = \begin{bmatrix} 0 & 1/2 \\ 1/2 & 0 \end{bmatrix}$ is a indefinite matrix; it has eigenvalues $\{1/2, -1/2\}$. Identifying these constraint equations as quadratic constraints allows us to develop this problem in the context of quadratically constrained quadratic programs (QCQP).

The subsurface imaging problem can then be rewritten using these indefinite quadratic constraints.

$$\begin{aligned} \text{minimize} \quad & \|\hat{u} - \mathcal{M}u\|_2^2 \\ \text{subject to} \quad & \begin{bmatrix} u \\ \sigma \end{bmatrix}^T \begin{bmatrix} \mathbf{0} & 1/2_k \\ 1/2_k & \mathbf{0} \end{bmatrix} \begin{bmatrix} u \\ \sigma \end{bmatrix} + a_k u - s_k = 0 \quad k = 1, \ldots, K \\ & 0 \leq \sigma \leq \sigma_{\max} \end{aligned} \tag{7.11}$$

The row vectors a_k correspond to the kth rows of the matrix A. The matrices used for the quadratic constraint are described using a shortened notation; the matrices are of full size, and the $1/2_k$ terms are in the appropriate location to multiply u_k and σ_k and $\mathbf{0}$ are zero matrices of appropriate size.

As stated, (7.11) is a non-convex optimization problem which is clear from the indefinite quadratic constraints, as such, it is an NP-complete problem for which polynomial time algorithms with optimality guarantees have not been discovered.

7.3 Solution Techniques

In this section we describe seven solution techniques that can be used to approximate the solution to optimization problems with identical mathematical form as the non-convex subsurface electromagnetic imaging problem given in (7.3). These methods include the sequential linear approximation, artificial source, split field, split-σ, phase

split, bi-projection, and semidefinite relaxation methods. The artificial source, split field, phase split, and bi-projection methods are all based on the alternating directions method of multipliers, a technique for solving convex optimization problems [20]. The sequential linear approximation and semidefinite relaxation methods are other techniques for approximating the solution to non-convex optimization problems.

7.3.1 Sequential Linear Approximation

In the sequential linear approximation, we solve a sequence of convex problems in which linear approximation of the non-convex constraint $Au+\sigma\circ u = s$ is substituted. Incremental updates for u and σ are found by solving the optimization problem with linear, convex constraints; the linear model is updated after each iteration.

We implicitly define the constraint equation as a function:

$$g(u,\sigma) = Au + (\mathcal{M}_{\mathrm{d}}^T\sigma)\circ u = s. \tag{7.12}$$

The first order Taylor expansion of $g(u,\sigma)$ is given as:

$$g(u+\delta_u,\sigma+\delta_\sigma) \simeq g(u,\sigma) + \nabla_u g(u,\sigma)^T\delta_u + \nabla_\sigma g(u,\sigma)^T\delta_\sigma. \tag{7.13}$$

More explicitly:

$$g(u+\delta_u,\sigma+\delta_\sigma) \simeq \left(A + \mathbf{diag}(\mathcal{M}_{\mathrm{d}}^T\sigma)\right)u + (A+\mathbf{diag}(\mathcal{M}_{\mathrm{d}}^T\sigma))\delta_u + \mathbf{diag}(\mathcal{M}_{\mathrm{d}}^T\delta_\sigma)u \tag{7.14}$$

For the Helmholtz wave equation, this linearization is often called the Born approximation [35] and gives small error if the magnitude of the perturbation, δ_σ, and the ratio between the scattered wave and the background wave, $\frac{\delta_u}{u}$, are small. It is convex in the incremental update terms δ_σ and δ_u. More discussion of the Born approximation and related methods is given in Section 6.3.2.

In each iteration, an update to the subsurface materials, δ_σ, is found for each of the J illumination conditions. Each of these updates is aggregated, averaged, and added to the current estimate of the conductivity parameters, σ^m. Updates to the

fields, δ_u, are not added to the model. Instead, we compute new background fields at each iteration based on the current set of parameters σ. To ensure that the Born Approximation is not violated, δ_σ is restricted to a trust region, enforced by an ℓ_∞ ball, $\|\delta_\sigma\|_\infty \leq \gamma$. The update δ_σ is scaled by the parameter α and added to the current estimate of the subsurface materials σ. The full algorithm is given in Algorithm 7.3.1. The algorithm iterates over the data several times to reach a stable estimate of the subsurface materials.

Algorithm 7.3.1 Sequential Linear Approximation (SLA)

Require: $\sigma^1 \leftarrow 0$

1: **for** $m = 1, 2, \ldots$ **do**

2: $\quad u_j^m \leftarrow \left[A_j + \mathbf{diag}(\mathcal{M}_\mathrm{d}^T \sigma^m)\right]^{-1} s_j$

$$\left[\delta_{u_j},\ \delta_{\sigma_j}\right] \leftarrow \begin{array}{rl} \text{argmin} & \|\hat{u}_j - \mathcal{M}_\mathrm{r}\left(u_j^m + \delta_{u_j}\right)\|_2^2 \\ \text{subject to} & (A_j + \mathbf{diag}(\mathcal{M}_\mathrm{d}^T \sigma^m))(u_j^m + \delta_{u_j}) \\ & \quad + \mathbf{diag}(u_j^m)\left(\mathcal{M}_\mathrm{d}^T \delta_{\sigma_j}\right) = 0 \\ & 0 \leq \sigma^m + \delta_{\sigma_j} \leq \sigma_{\max} \\ & \|\delta_{\sigma_j}\|_\infty \leq \gamma \end{array}$$

3:

4: $\quad \sigma^{m+1} = \sigma^m + \alpha \frac{1}{J} \sum_{j=1}^{J} \delta_{\sigma_j}$

5: **end for**

A custom solver was written for the internal convex optimization problem that appears in step 3 of Algorithm 7.3.1 using ADMM techniques. A convex solver based on these techniques permits accurate solution of the optimization problem with fewer matrix factorizations than an interior point solver. The ADMM solver developed for Algorithm 7.3.1 differs from the ADMM algorithms discussed in the next section. Each step of the sequential linear approximation is convex, therefore each subproblem will converge to its optimal solution. In general, the sequential linear approximation method is not guaranteed to converge to the global optimizer; at best it will find a local optimum.

7.3.2 ADMM Methods

The alternating directions method of multipliers is a convergent technique for solving convex optimization problems [20]. Optimization problems with decomposable structure are well suited for ADMM solvers. The ADMM algorithm blends augmented Lagrangians and dual ascent methods yielding convergent solvers that often have components that can be solved in parallel. Compared to an interior point solver, each iteration in the ADMM algorithm can be computationally cheaper, but more iterations are typically needed to obtain satisfactory error.

The subsurface imaging problem is not convex, but ADMM algorithms have been applied to non-convex optimization problems that are similar in form to the subsurface imaging problem, namely, non-negative matrix factorization [90], which has applications in medical image registration and computer vision [12, 62].

Since ADMM solvers can exploit decomposable structure, it is often necessary and beneficial to introduce new variables that do not change the optimal solution but change the decomposability of the problem. That is, by introducing new constraints that ultimately yield the same optimization problem, we can produce a solver that requires fewer matrix factorizations. Such reduction of computational complexity per iteration is attractive. If a method requires solving a new $N \times N$ matrix equation at each step, then it requires $O(mN^3)$ operations for m iterations. If an algorithm can reuse the factorization, then for m iterations the cost is $O(N^3 + mN^2)$ operations. While this is generally not considered a change in big-O notation, it can still yield substantial computational savings.

In this section, we explore five different ways of applying ADMM to the subsurface imaging problem. The non-convex nature poses new challenges and requires careful structuring to develop an efficient method.

While the subsurface imaging problem is not convex, it is "biconvex." In the non-convex constraint function, $g(u, \sigma)$, when u is held constant, the function is linear in σ and if σ is held constant, the function is linear in u. ADMM has been applied to similar "biconvex" problems by splitting the optimization into two parts to exploit this property. Three of the subsequent techniques split the optimization problem into two general steps: first, σ is held constant while an optimal u is found, next u is held

constant while an optimal σ is found.

Artificial Source

The artificial source method is named for the "artificial source" term introduced to the constraint equations. We introduce a new variable, $w \in \mathbb{C}^{\Omega_\mathrm{d}}$:

$$w_j = \sigma \circ (\mathcal{M}_\mathrm{d} u_j). \tag{7.15}$$

The new variable w_j behaves as a source function in the subsurface domain, Ω_d. With this new variable, w_j, our optimization problem can be stated as:

$$\begin{aligned} \sigma_\mathrm{opt} = \quad & \underset{\sigma \in \mathbb{R}}{\operatorname{argmin}} \quad \textstyle\sum_j \|\mathcal{M}_\mathrm{r} u_j - \hat{u}_j\|_2^2 \\ & \text{subject to} \quad A_j u_j + \mathcal{M}_\mathrm{d}^T w_j = s_j \\ & \qquad\qquad\quad w_j - \sigma \circ (\mathcal{M}_\mathrm{d} u_j) = 0 \\ & \qquad\qquad\quad 0 \leq \sigma \leq \sigma_\mathrm{max}. \end{aligned} \tag{7.16}$$

This problem is mathematically equivalent to the original optimization problem, but lends itself to different algorithmic structure. The constraints are such that the problem is convex in u and w if σ is held constant, and convex in σ if u and w are held constant.

Introducing the variable w allows us to find a true solution of Maxwell's equation at each iteration. The term w can be interpreted as a set of complex sources in the subsurface domain Ω^d. By the principle of reciprocity, we can always find a set of complex sources that will reproduce the measurements observed at the surface. Through the course of the optimization problem, we enforce that these complex sources can be described as the local product of the electromagnetic field and a positive, real conductivity, σ. The introduction of this term is very similar to the contrast source method [4, 11]. Our ADMM algorithm is different from the contrast source method in that it finds the optimum by solving a sequence of optimization problems, while in the contrast source method, the objective function is typically minimized using gradient or projected gradient techniques.

To solve this optimization problem with ADMM requires solving a sequence of two optimization problems, each with an analytic solution obtained by solving a single linear system of equations. First, an update for u and w is calculated such that Maxwell's equations with the artificial sources w are satisfied. A penalty is added to the objective function for violation of the definition of w with respect to u and σ. Next, a new value of σ is computed so that it matches the current estimates of u and w and is in the constraint set, $\mathcal{C}$:

$$\{\sigma \in \mathcal{C} | 0 \leq \sigma \leq \sigma_{\max}, \sigma \in \mathbb{R}\} \tag{7.17}$$

Finding a solution for σ in the bounded, real, constraint set is done by applying the saturation function, $\mathbf{sat}_{0,\sigma_{\max}}$, defined in (7.18).

$$\mathbf{sat}_{a,b}(x) = \begin{cases} b & x \geq b \\ a & x \leq a \\ x & \text{otherwise} \end{cases} \tag{7.18}$$

Finding a single σ enforces consensus among the J illumination conditions. Last, the dual variable for the materials constraint, λ, is updated. The full algorithm is given in Algorithm 7.3.2.

Algorithm 7.3.2 Artificial Source

Require: $\sigma^1 \leftarrow 0, \lambda^1 \leftarrow 0$

1: **for** $m = 1, 2, \ldots$ **do**

2: $$[u_j^{m+1}, w_j^{m+1}] \leftarrow \underset{u,w}{\operatorname{argmin}} \tfrac{1}{2}\|\mathcal{M}_{\mathrm{r}} u_j - \hat{u}_j\|_2^2 + \tfrac{\rho}{2}\|w - \sigma^m \circ (\mathcal{M}_{\mathrm{d}} u) + \lambda_j^m\|_2^2 \quad \text{subject to} \quad A_j u + \mathcal{M}_{\mathrm{d}}^T w = s_j$$

3: $$\sigma^{m+1} \leftarrow \underset{\sigma}{\operatorname{argmin}}\, I_{\mathcal{C}}(\sigma) + \tfrac{\rho}{2} \textstyle\sum_j \|w_j^{m+1} - \sigma \circ (\mathcal{M}_{\mathrm{d}} u_j^{m+1}) + \lambda_j^m\|_2^2$$

4: $$\lambda_j^{m+1} \leftarrow \lambda_j^m + w_j^{m+1} - \sigma_{m+1} \circ (\mathcal{M}_{\mathrm{d}} u_j^{m+1})$$

5: **end for**

Algorithm 7.3.2 has two favorable properties: the update steps for u_j, w_j, and λ_j can be performed in parallel and we can obtain the solution to each of the optimization steps by solving a system of linear equations. The optimization problem in line 3 can

be solved by first solving the unrestricted least squares problem, then projecting into the convex set $\mathcal{C}$, which can be done by simple thresholding. The downside of this algorithm is that in each iteration, m, a matrix factorization is required to solve the optimization problem of line 2.

Split σ

We can improve the performance of the artificial source method by enforcing consensus at a different location in the algorithm. In the formulation given above, it is possible that each illumination condition will produce a different estimate of the optimal conductivity parameter, σ, outside the constraint set. When consensus is established, one illumination condition that has an outlier estimate could substantially skew the global estimate of the conductivity parameter σ. Since the consensus step happens before the thresholding step, it is possible that values significantly outside the constraint set can skew the estimate. Therefore, we introduce the split-σ method as a method that enforces consensus after thresholding.

We introduce a new global variable, $\tilde{\sigma}$, that will enforce consensus after thresholding on the local, model dependent estimates of the conductivity, σ_j. The split-σ method has shown improved accuracy with respect to other methods.

Algorithm 7.3.3 Split σ

Require: $\tilde{\sigma}^1 \leftarrow 0, \sigma_j^1 \leftarrow 0, \lambda^1 \leftarrow 0$

1: **for** $m = 1, 2, \ldots$ **do**

2: $$[u_j^{m+1}, w_j^{m+1}] \leftarrow \underset{u,w}{\operatorname{argmin}} \tfrac{1}{2}\|\mathcal{M}_{\mathrm{r}} u_j - \hat{u}_j\|_2^2 + \tfrac{\rho}{2}\|w - \sigma_j^m \circ (\mathcal{M}_{\mathrm{d}} u) + \lambda_j^m\|_2^2 \quad \text{subject to} \quad A_j u + \mathcal{M}_{\mathrm{d}}^T w = s_j$$

3: $$\sigma_j^{m+1} \leftarrow \underset{\sigma}{\operatorname{argmin}}\, I_{\mathcal{C}}(\sigma) + \tfrac{\rho}{2}\sum_j \|w_j^{m+1} - \sigma \circ (\mathcal{M}_{\mathrm{d}} u_j^{m+1}) + \lambda_j^m\|_2^2 + \tfrac{\xi}{2}\|\sigma - \tilde{\sigma} + \theta_j^m\|_2^2$$

4: $$\tilde{\sigma} = \tfrac{1}{J}\sum_j (\sigma_j - \theta_j)$$

5: $$\theta_j^{m+1} \leftarrow \theta_j^m + \sigma_j - \tilde{\sigma}$$

6: $$\lambda_j^{m+1} \leftarrow \lambda_j^m + w_j^{m+1} - \sigma_{m+1} \circ (\mathcal{M}_{\mathrm{d}} u_j^{m+1})$$

7: **end for**

Split Field

If we introduce a new variable, $v_j \in \mathbb{C}^{\Omega}$, we can eliminate the necessity to factor a large matrix at every iteration. At the optimal point, we will enforce that $u = v$. Introducing an additional estimate of the electromagnetic field, v, allows us to rearrange the algebra so that at each iteration changes in the material parameter σ show up as changes in a diagonal matrix which is simple to solve. We introduce the variable v into Maxwell's constraint equation:

$$A_j u_j + v_j \circ (\mathcal{M}_{\mathrm{d}}\sigma) = s_j. \tag{7.19}$$

We rewrite the optimization problem as:

$$\begin{aligned} \sigma_{\mathrm{opt}} = \underset{\sigma}{\operatorname{argmin}} \quad & \textstyle\sum_j \|\mathcal{M}_{\mathrm{d}} u_j - \hat{u}_j\|_2^2 \\ \text{subject to} \quad & A_j u_j + v_j \circ (\mathcal{M}_{\mathrm{d}}\sigma) = s_j, \quad \forall\, j \\ & v_j - u_j = 0 \\ & \sigma \in \mathcal{C}. \end{aligned} \tag{7.20}$$

We can apply ADMM to obtain Algorithm 7.3.4. This algorithm is similar to Algorithm 7.3.2, except an additional update step is added for the surrogate field term, v. The purpose of this reformulation is to save on computational cost associated with factoring perturbations of the matrix A. To solve the update of line 3, we build and compute the LU factorization of the matrix $(\mathcal{M}_{\mathrm{r}}^T \mathcal{M}_{\mathrm{r}} + \rho A_j^T A_j + \xi I)$ in the first iteration and solve the system in subsequent iterations by reusing the pre-computed factorization. Updates to u can be found in each iteration by solving the system for different right hand sides which can be done in $O(n^2)$ time.

Updates for v (line 4) require factoring a matrix at every iteration since the value of σ changes at every iteration, but the matrix is diagonal and can therefore be solved in $O(n)$ time. For this algorithm to work, we have to introduce a new dual variable θ for our enforcement equation $v - u = 0$.

In this algorithm, the internal scaling parameters, ρ, ξ are not the same for the two augmented Lagrangian terms. Selection of the parameters ξ and ρ is discussed

Algorithm 7.3.4 Split Field

Require: $\rho, \hat{u}$

Ensure: u, σ

$u \leftarrow 0,\ v \leftarrow 0,\ \sigma \leftarrow 0,\ \lambda \leftarrow 0,\ \theta \leftarrow 0$

for $m = 1 \dots M$ **do**

$$u_j^{m+1} \leftarrow \underset{u}{\operatorname{argmin}} \quad \|\mathcal{M}_{\mathrm{r}} u - \hat{u}_j\|_2^2 + \frac{\rho}{2}\|v_j^m - u + \lambda_j^m\|_2^2 + \frac{\xi}{2}\|A_j u + v^m \circ (\mathcal{M}_{\mathrm{d}}^T \sigma^m) - s_j + \lambda_j^m\|_2^2$$

$$v_j^{m+1} \leftarrow \underset{v}{\operatorname{argmin}} \quad \frac{\rho}{2}\|v - u_j^{m+1} + \theta^m\|_2^2 + \frac{\xi}{2}\|A_j u_j^{m+1} + v \circ (\mathcal{M}_{\mathrm{d}}^T \sigma^m) - s_j + \lambda_j^m\|_2^2$$

$$\sigma^{m+1} \leftarrow \underset{\sigma}{\operatorname{argmin}}\, I_{\mathcal{C}}(\sigma) + \frac{\rho}{2} \sum_j \|A_j u_j^{m+1} + v^{m+1} \circ (\mathcal{M}_{\mathrm{d}}^T \sigma) - s_j + \lambda_j^m\|_2^2$$

$$\lambda_j^{m+1} = \lambda_j^m + A_j u_j^{m+1} + v_j^{m+1} \circ (\mathcal{M}_{\mathrm{d}}^T \sigma^{m+1}) - s_j$$

$$\theta_j^{m+1} = \theta_j^m + v_j^{m+1} - u_j^{m+1}$$

end for

in Chapter 8.

Phase Split

The phase split method takes advantage of the property that if the sign or phase of the electromagnetics is known, then the optimization problem is convex. Since generally, the phase is unknown, the convex problem is solved iteratively and the estimate of the phase is updated from iteration to iteration. At a single frequency, the phase split method can be developed so that only a single matrix factorization is necessary at the beginning of the algorithm. Unfortunately, in the problems of interest where we wish to incorporate information from multiple models, a matrix factorization for each model must be calculated at each iteration.

For simplicity, we consider the reduced form of the reals-only imaging problem, (7.8), where we have the real-only term, $x \in \mathbb{R}, x = u \circ \sigma$. The true solution can be found in the well-determined case if the sign of u is known in addition to the sign of

σ.

$$\begin{aligned} \text{minimize} \quad & \|u - \hat{u}\|_2^2 \\ \text{subject to} \quad & Au + x = s \\ & x = u \circ \sigma \\ & 0 \leq \sigma \leq \sigma_{\max} \end{aligned}$$

If we know the sign of u, we can then eliminate σ from the equations and obtain a set of equations with linear inequality bounds. For example, if $u \leq 0$, then:

$$\begin{aligned} \text{minimize} \quad & \|u - \hat{u}\|_2^2 \\ \text{subject to} \quad & Au + x = s \\ & 0 \geq x \geq u\sigma_{\max} \end{aligned}$$

which is a convex program. The solution is not unique as described in Section 7.2.3. If we can accurately guess the sign of u, then the problem can be solved efficiently. The solution is unique if the number of measurements is equal to the number of unknowns.

In the complex case, this can be generalized to guessing the proper phase angle of u. We use the variable $r \in \mathbb{R}$ defined in Equation 7.9. Furthermore, we let ϕ be the phase angle of the electromagnetic field, u, $\phi = \arg(u)$. The complex form of the reduced electromagnetic subsurface imaging problem can be expressed in a form that highlights the phase dependence:

$$\begin{aligned} \text{minimize} \quad & \|u - \hat{u}\|_2^2 \\ \text{subject to} \quad & Au + r \circ e^{i\phi} = b \\ & 0 \geq r \geq \Re(e^{-i\phi} \circ u)\sigma_{\max} \\ & \phi = \arg(u). \end{aligned} \tag{7.21}$$

If there is a good way to estimate ϕ, this problem can be solved efficiently in a small number of iterations. The solution is unique provided the number of measurements exceeds the number of model parameters. If the phase is not known, we can derive a method that updates the phase in between iterations. The convergence of this method cannot be proven, but it exhibits suitable performance in practice.

The phase split method does not generalize gracefully to multiple illuminations.

Since the multiple models achieve consensus on a single conductivity parameter, σ, we must reintroduce this variable to the problem in order to coordinate multiple models. Once a consensus variable across models, σ, is introduced, reapplying the same factorization of the projection matrix at every iteration is no longer possible.

Indefinite Projection

Alternatively, we can view the problem as one in which we are trying to find a collection of variables that lie at the intersection of two sets. We analyze the real-only case first and take the collection of variables, $x, u, \sigma \in \mathbb{R}$, as described in Section 7.2.2. The first constraint set is that the variables fit Maxwell's equations and are in our regularized set, the standard constraints for the problem:

$$Au + x = s, \quad 0 \leq \sigma \leq \sigma_{\max} \tag{7.22}$$

Second, we want the three variables to belong to the set defined by the indefinite quadratic constraint,

$$x = \sigma \circ u. \tag{7.23}$$

We use ADMM to find a set of variables that lies at the intersection of these two sets:

$$\begin{array}{rl} \text{minimize} & \|\mathcal{M}u - \hat{u}\|_2^2 + \sum_k I_{x=u\sigma}(\tilde{x}_k, \tilde{u}_k, \tilde{\sigma}_k) \\ \text{subject to} & Au + x = b \\ & x = \tilde{x},\ u = \tilde{u},\ \sigma = \tilde{\sigma} \\ & 0 \leq \sigma \leq \sigma_{\max}. \end{array} \tag{7.24}$$

Finding u, x, σ that belong to set described in 7.22 has been discussed in the context of other methods. It consists of linear projection and application of the thresholding function. Finding $\tilde{x}, \tilde{\sigma}, \tilde{u}$ that belong to the set defined in 7.23 can also be done exactly using a method from control theory. As discussed in Section 7.2.4 we can rewrite the constraint, $u \circ \sigma = x$ in quadratic form as given in Equation 7.10. The solution to a quadratically constrained quadratic program can be solved even if

the matrices involved, A_i in (7.25), are not positive semidefinite.

$$\begin{array}{rl} \text{minimize} & z^T A_0 z + 2b_0^T z + c_0 \\ \text{subject to} & z^T A_1 z + 2b_1^T z + c_1 \leq 0 \end{array} \tag{7.25}$$

The S-Lemma guarantees that there exists a $z \in \mathbb{R}^n$ that satisfies:

$$z^T A_0 z + 2b_0^T z + c_0 < 0, \quad z^T A_1 z + 2b_1^T z + c_1 \leq 0$$

if and only if there exists no $\lambda \geq 0$ such that:

$$\begin{bmatrix} A_0 & b_0 \\ b_0^T & c_0 \end{bmatrix} + \lambda \begin{bmatrix} A_1 & b_1 \\ b_1^T & c_1 \end{bmatrix} \succeq 0.$$

These alternatives guarantee that if a solution exists, it is unique and strong duality holds between the dual problem and the primal problem. Once we have found the optimal dual variables, the optimal primal variables can be found with a simple linear equation as a consequence of the first-order KKT optimality conditions:

$$(A_0 + \lambda A_1)z = -(2b_0 + \lambda 2b_1)$$

Solving this equation for z is guaranteed to return the optimal solution for z. A simple algorithm to find the solution can be constructed using bisection. We search for a $\lambda : 0 \leq \lambda \leq 2$ by bisection and verifying the constraint equation.

Using this projection method, we can describe an ADMM algorithm that involves projection over the indefinite cone. This algorithm is described in Algorithm 7.3.5. In this algorithm, there are two key features. First, all of the steps can be solved and evaluated exactly. The first step, the update for (u, x) is a linearly constrained quadratic program which can be solved efficiently with a single matrix equation. Second, only the right-hand side of the equation, s, changes from iteration to iteration, therefore a cached copy of the factorization or a recycled Krylov subspace technique could be used to efficiently solve these equations with minimal computational overhead.

Algorithm 7.3.5 Indefinite Projection

1: $\tilde{u} \leftarrow 0, \tilde{w} \leftarrow 0, \tilde{\sigma} \leftarrow 0, \lambda_u \leftarrow 0, \lambda_w \leftarrow 0, \lambda_\sigma \leftarrow 0$

2: **for** $m = 1, 2, \ldots$ **do**

3: $$\begin{aligned}(u^{m+1}, w^{m+1}) \leftarrow \quad & \underset{u,w}{\operatorname{argmin}} \, \|\mathcal{M}_{\mathrm{r}} u - \hat{u}\|_2^2 + \tfrac{\rho}{2}\left(\|u - \tilde{u} + \lambda_u\|_2^2 + \|w - \tilde{w} + \lambda_x\|_2^2\right) \\ & \text{subject to} \quad Au + w = b\end{aligned}$$

4: $$\begin{aligned}(\sigma^{m+1}) \leftarrow \quad & \underset{\sigma}{\operatorname{argmin}} \, \|\sigma - \tilde{\sigma} + \lambda_\sigma\|_2^2 \\ & \text{subject to} \quad 0 \le \sigma \le \sigma_{\max}\end{aligned}$$

5: $$\begin{aligned}(\tilde{u}_k^{m+1}, \tilde{w}_k^{m+1}, \tilde{\sigma}_k^{m+1}) \leftarrow \quad & \underset{\tilde{u},\tilde{w},\tilde{\sigma}}{\operatorname{argmin}} \quad \|u_k^{m+1} - \tilde{u}_k + \lambda_{u_k}^m\|_2^2 + \\ & \qquad \|w_k^{m+1} - \tilde{w}_k + \lambda_{w_k}^m\|_2^2 + \\ & \qquad \|\sigma_k^{m+1} - \tilde{\sigma}_k + \lambda_{\sigma_k}^m\|_2^2 \\ & \text{subject to} \quad \tilde{w}_k = \begin{bmatrix} u_k \\ \tilde{\sigma} \end{bmatrix}^T \begin{bmatrix} 0 & 1/2 \\ 1/2 & 0 \end{bmatrix} \begin{bmatrix} \tilde{u}_k \\ \tilde{\sigma}_k \end{bmatrix}\end{aligned}$$

6: $\lambda_u^{m+1} \leftarrow \lambda_u^m + u^{m+1} - \tilde{u}^{m+1}$

7: $\lambda_w^{m+1} \leftarrow \lambda_w^m + w^{m+1} - \tilde{w}^{m+1}$

8: $\lambda_\sigma^{m+1} \leftarrow \lambda_\sigma + \sigma^{m+1} - \tilde{\sigma}^{m+1}$

9: **end for**

Complex Indefinite Projection

The above strategy for projection into the set $u\sigma = x$ is guaranteed for $x, \sigma, u \in \mathbb{R}$. In subsurface imaging, we are concerned with the case where $w, u \in \mathbb{C}$, $\sigma \in \mathbb{R}$. To incorporate this constraint, we optimize over $\mathbb{R}^5$ and have two indefinite quadratic constraints.

We form the variable $z = \begin{bmatrix} \Re(u) & \Im(u) & \Re(w) & \Im(w) & \sigma \end{bmatrix}^T \in \mathbb{R}^5$. We can then express the optimization problem as:

$$\begin{aligned} \text{minimize} \quad & z^T A_0 z + 2b_0^T z + c_0 \\ \text{subject to} \quad & z^T A_\Re z + 2b_\Re^T z + c_\Re \le 0 \\ & z^T A_\Im z + 2b_\Im z + c_\Im \le 0. \end{aligned}$$

In these equations, A_0 is the identity matrix and the matrices $A_\Re$ and $A_\Im$ are given

as:

$$A_{\Re} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 \end{bmatrix} \quad A_{\Im} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 \end{bmatrix}$$

and $b_0 = -z + \lambda_z$. We use the fixed versions of u, w, σ to form z and the appropriate dual variables for λ_z. The vectors $b_{\Re}$ and $b_{\Im}$ are fixed as $b_{\Re} = \begin{bmatrix} 0 & 0 & \frac{1}{2} & 0 & 0 \end{bmatrix}$, $b_{\Im} = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{2} & 0 \end{bmatrix}$. Lastly, $c_{\Re} = c_{\Im} = 0$ and $c_0 = b_0^T b_0$.

Solution of this problem can be found efficiently using a cutting -plane ellipsoid method or an interior point method. There are several circumstances in which this particular projection fails to find the correct answer. This method is often slower due to the number of projection sub-problems that must be solved.

7.3.3 Semidefinite Relaxation

Developing electromagnetic subsurface imaging as a quadratically constrained quadratic program allows us to use a semidefinite embedding in order to approximate the solution. To express electromagnetic subsurface imaging as a semidefinite program (SDP), we introduce the variable $z = \begin{bmatrix} u & \sigma \end{bmatrix}^T$ to simplify notation. We can express the quadratic form, $z^T P z$, a scalar quantity, as $\mathbf{Tr}(z^T P z) = \mathbf{Tr}(P z z^T)$. At this point, we substitue a variable $Z = z z^T$ so that the quadratic constraints become $\mathbf{Tr}(P_j Z)$. We can equivalently rewrite the quadratic objective function as

$$\mathbf{Tr}(\begin{bmatrix} \mathcal{M}^T \mathcal{M} & 0 \\ 0 & 0 \end{bmatrix} X) + \begin{bmatrix} 2\hat{u}^T \mathcal{M}^{\mathcal{T}} \\ 0 \end{bmatrix} x.$$

With these variable transformations, we can rewrite our problem in standard SDP form:

$$\begin{aligned} \text{minimize} \quad & \mathbf{Tr}\left(\begin{bmatrix} \mathcal{M}^T\mathcal{M} & 0 \\ 0 & 0 \end{bmatrix} Z\right) + \begin{bmatrix} 2\hat{u}^T\mathcal{M}^T \\ 0 \end{bmatrix} z \\ \text{subject to} \quad & \mathbf{Tr}(P_k Z) + \begin{bmatrix} a_k \\ 0 \end{bmatrix}^T Z - b_k = 0 \quad k = 1, \ldots, K \\ & 0 \leq \sigma \leq \sigma_{\max} \\ & zz^T = Z. \end{aligned} \tag{7.26}$$

The constraint $zz^T = Z$ is not convex, but we can replace this constraint with its semidefinite relaxation in order to obtain a convex problem:

$$\begin{bmatrix} Z & z \\ z^T & 1 \end{bmatrix} \succeq 0.$$

Solution of the semidefinite relaxation is not guaranteed to give the exact solution $\begin{bmatrix} u & \theta \end{bmatrix}^T$. If the matrix, Z, obtained after solving the semidefinite relaxation has rank 1, then the semidefinite relaxation is tight and the optimal solution has been found for the original QCQP. In the case that there is only one quadratic constraint equation, the relaxation is known to be tight as guaranteed by the S-Lemma, but in the case of multiple constraint equations, the relaxation is not guaranteed to be tight. Numerical solutions to these SDP problems were obtained using `CVX` [70, 69] and `SeDuMi` [135].

Chapter 8

Numerical Results

The seven solvers described in Chapter 7 were implemented and tested in both Matlab and Python and in sequential and distributed environments. A sequence of small tests were developed to explore the performance and speed of these algorithms and establish a subsampling principle. The best performing algorithms in these tests were further developed and tested in electromagnetic subsurface imaging problems. We present the subsampling results in Section 8.3. General electromagnetic subsurface imaging results for the four best-performing solvers from Section 8.3 are presented in Section 8.1.

8.1 General Performance

The four solvers developed for electromagnetic subsurface imaging, sequential linear approximation, artificial source, split field, and split-σ, were compared directly using a sequence of 1000 Monte Carlo trials over randomly generated subsurface conductivity profiles.

An example of a true conductivity profile and the conductivity profiles recovered using each of the four algorithms, sequential linear approximation, artificial source, split field, and split-σ, are shown in Figure 8.1. The true conductivities are at the top; a clear block structure can be seen in this randomly-generated profile.

The profile was embedded in a homogeneous Earth model with background conductivity of 0.001(S/m). For each of these reconstructions 16 illumination conditions were used, 4 frequencies (1 kHz, 3kHz, 13 kHz, 50 kHz) and 4 angles of incidence $(75°, 45°, -45°, -75°)$.

In each of these reconstructions, several common features can be seen. Each of the reconstruction methods accurately image the small, shallow conductivity void near the surface around 150 in the lateral direction. Furthermore each of the methods reveal a general sense of lower conductivity in the left one-third of the profile with stronger conductivities in the right two-thirds of the profile. As is typical, the sequential linear approximation gives results that are smoother, *i.e.*, transitions from one material to another are generally smooth. The artificial source method has a blotchy characteristic, a typical feature of these reconstructions. The split-σ gives the reconstruction with the best accuracy for this particular profile. Inspecting the results from a single profile give us a heuristic understanding of how the algorithms perform, but a more rigorous, randomized testing procedure is necessary to directly compare the algorithms.

8.1.1 Random Profile Generation

Random profiles were generated using a simple idea. The material structures of interest for imaging are typically made of discrete values. Construction of the random profiles is a four step process. We generate a square block structured profile by taking the inverse Haar transform of a square domain where the lowest Haar coefficients are given random values. The large block-structure profile is rotated by an arbitrary angle and truncated to the size of the reconstruction domain. An interpolation to match the grid points is then implemented using a simple nearest-neighbor interpolation scheme. The unique conductivity values are then replaced with conductivity values obtained from the uniform distribution $U[0, \sigma_{\max}]$. The histogram of conductivity values across the 1000 profiles from a randomly selected pixel is shown in Figure 8.2. The Kullback-Leibler divergence for each pixel with respect to the uniform distribution is calculated and also shown in Figure 8.2. The plotted Kullback-Leibler divergences are very low

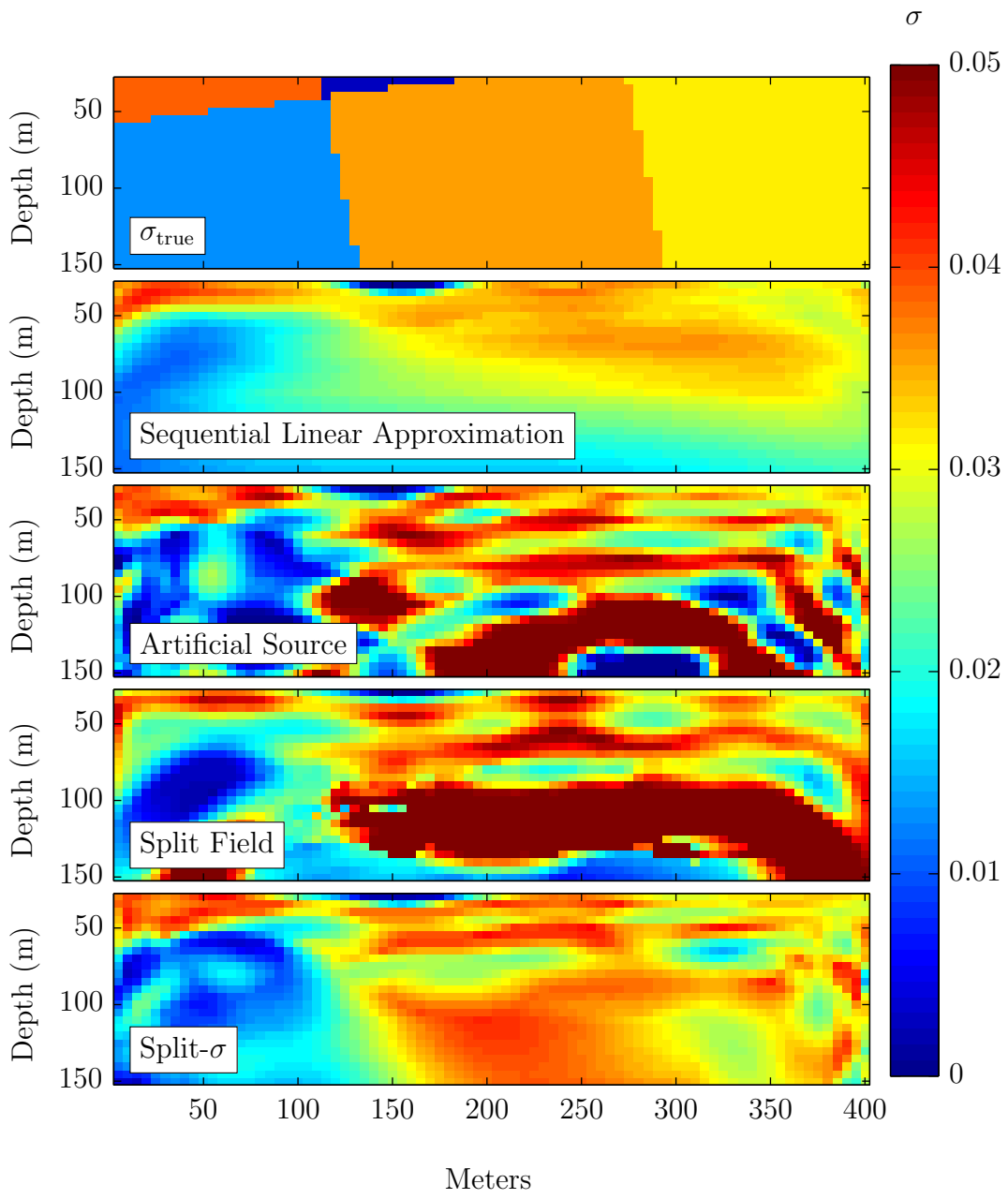

Figure 8.1: Comparison of all four solvers solvers.

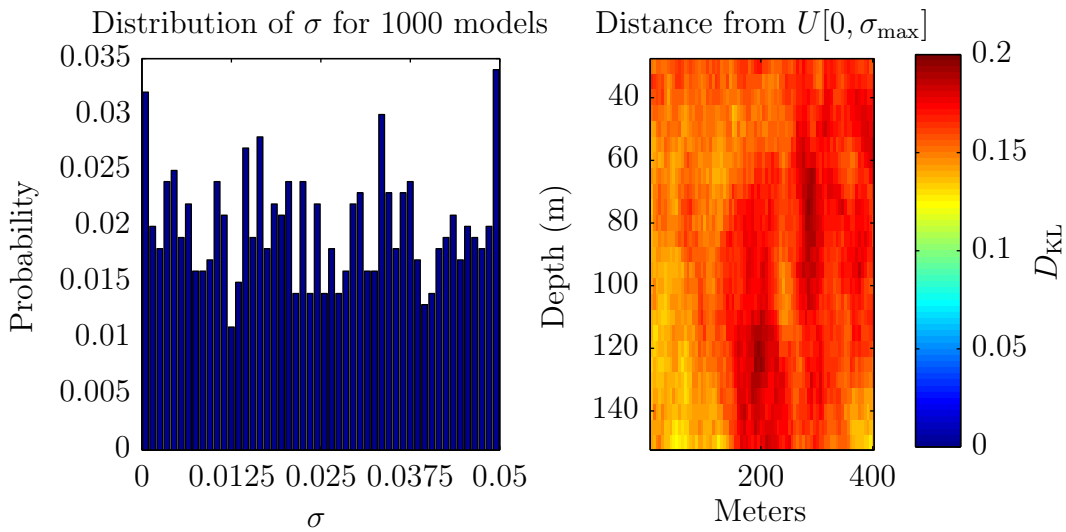

Figure 8.2: Statistics of conductivity profiles. On the left is a histogram showing the probability distribution of the conductivity values for a single pixel in the image domain. On the right is the Kulback-Leibler divergence between the observed distribution of the pixels and the ideal uniform distribution.

indicating that the distribution of conductivity values is close to the ideal uniform distribution.

8.1.2 Relative Entropy Metric

To evaluate the performance of the algorithms, we evaluate the Kullback-Leibler divergence between the distribution of the errors and the distribution of the errors under the worst case scenario. We define the worst case scenario as the case in which our model produces nothing but random numbers. That is, in the worst case,

$$\sigma_{\text{estimated}} \sim \mathcal{U}[0, \sigma_{\max}].$$

Subsequently, the distribution of the error between this worst case and the true conductivity is given by the triangle distribution, which is the difference between two

variables identically drawn from the uniform distribution. That is, we have

$$Q(\sigma_{\text{worst case}} - \sigma_{\text{true}}) \sim \mathcal{T}[-\sigma_{\max}, \sigma_{\max}].$$

Using histogram techniques, we can estimate the probability distribution of the error $P(\sigma_{\text{estimated}} - \sigma_{\text{true}})$. We then evaluate the accuracy of our method by computing the distance between these two probability distributions using the Kullback-Leibler divergence, D_{KL}, where

$$D_{KL}(P||Q) = \sum_i \log\left(\frac{P(i)}{Q(i)}\right) P(i).$$

The error distribution is estimated using 1000 Monte Carlo trials over 1000 different randomly-generated conductivity profiles.

The Kullback-Leibler divergence can be computed for the error distribution over each pixel in the subsurface reconstruction domain. The higher the Kullback-Leibler divergence, the greater the distance between the distributions and the higher the accuracy. A Kullback-Leibler divergence of 0 means the distributions are identical, and the algorithm performs no better than a random number generator.

Numerical results showing the accuracy for each pixel in the reconstruction domain for each of the four methods fully developed for electromagnetic subsurface imaging are shown in Figure 8.3. In this plot, the artificial source method (upper right plot) provides the best accuracy for estimating the subsurface conductivity near the surface, up to a depth of 60 meters. The split-σ method gives the highest accuracy throughout the reconstruction domain, which is clear evidence that rejection filtering of the intermediate conductivities yields more optimal results. The sequential linear approximation and the split field methods exhibit similar performance; both methods allow better estimation of the conductivities closer to the surface than those at depth.

In addition to the detailed pixel-by-pixel comparison, we can compute the average residual $\|\sigma_{\text{estimated}} - \sigma_{\text{true}}\|_2$ across the entire reconstruction domain. In Figure 8.4, the average residual is plotted as a function of execution time for each of the four algorithms. We can compare the runtime and performance of each of the algorithms.

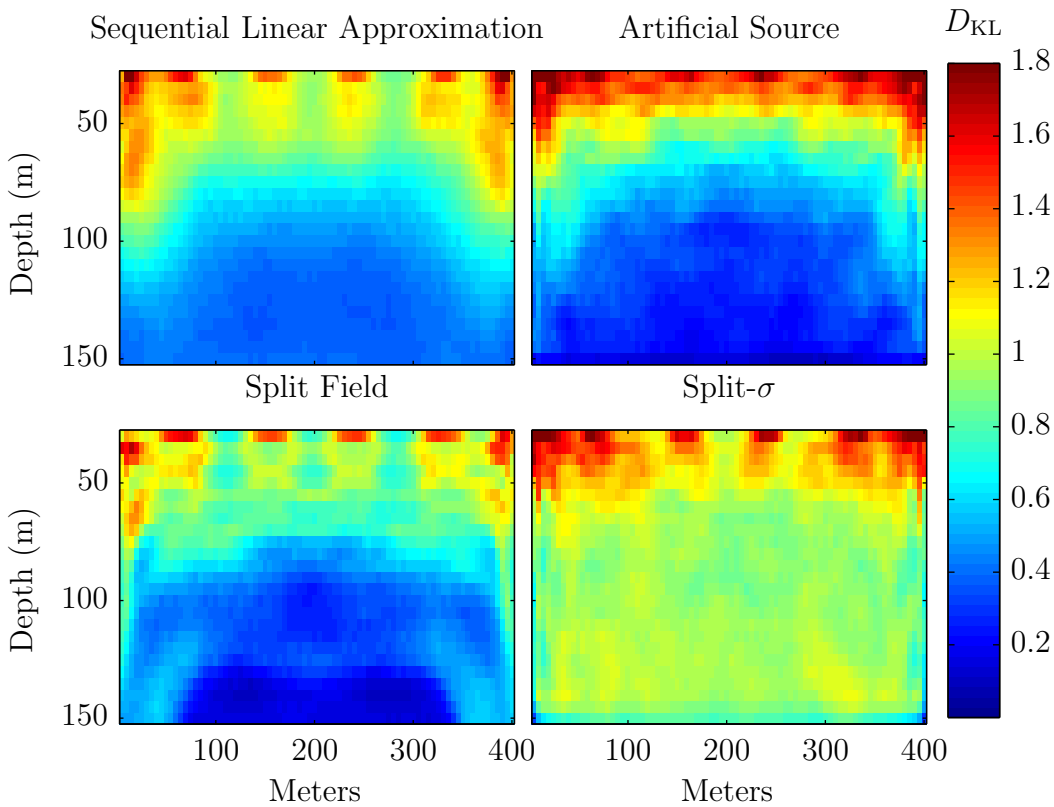

Figure 8.3: Comparison of relative entropies from each of 4 methods. The relative entropy metric indicates the average accuracy for each pixel of the reconstruction domain using each algorithm. Low Kullback-Leibler divergence means the error of the algorithm is distributed according to the worst case, a random number generator. Low Kullback-Leibler divergence implies low accuracy, while high Kullback-Leibler divergence implies high accuracy. Each of the methods is most accurate near the surface. The split-σ method gives the best performance throughout the reconstruction domain.

The artificial source and sequential linear approximation methods provide similar final residuals in the same runtime. The split field method allows 1000 iterations in much less time, about an order of magnitude less, than the other methods. The split-σ method provides the lowest final residual on average, and still runs about twice as fast as either the sequential linear approximation or the artificial source method.

8.2 Model Parameters

In the previous section, we established the accuracy of the algorithms under a single set of conditions. In this section, we explore the accuracy as certain model parameters vary.

8.2.1 Background Conductivity

The assumed background conductivity of the Earth can dramatically affect the performance of the algorithms. As mentioned previously, the skin depth of electromagnetic waves in the Earth is inversely proportional to the square root of the background conductivity, $\sigma_{\text{background}}$. As the conductivity increases, so does the rate of attenuation while the depth of penetration decreases. For high conductivities, it is difficult to get any information about the subsurface conductivity parameters, therefore estimating subsurface conductivity is often inaccurate.

In Figure 8.5 the average accuracy as a function of depth is shown as the background conductivity varies over five orders of magnitude, from $\sigma_{\text{background}} = 0.0001$ S/m to $\sigma_{\text{background}} = 1$ S/m. We see that for low conductivities, *i.e.*, $\sigma_{\text{background}} < 0.1$, our ability to estimate subsurface conductivity remains unchanged. There is an interesting effect between $\sigma = 0.001$ S/m and $\sim \sigma = 0.08$ S/m, namely a weakening of our ability to estimate subsurface conductivities. This effect can be explained by the varying relationships between the skin depths and the exact frequencies used in the inversion and in addition demonstrates the ill-posed nature of the problem; determining the ideal solution is sensitive to the parameter inputs.

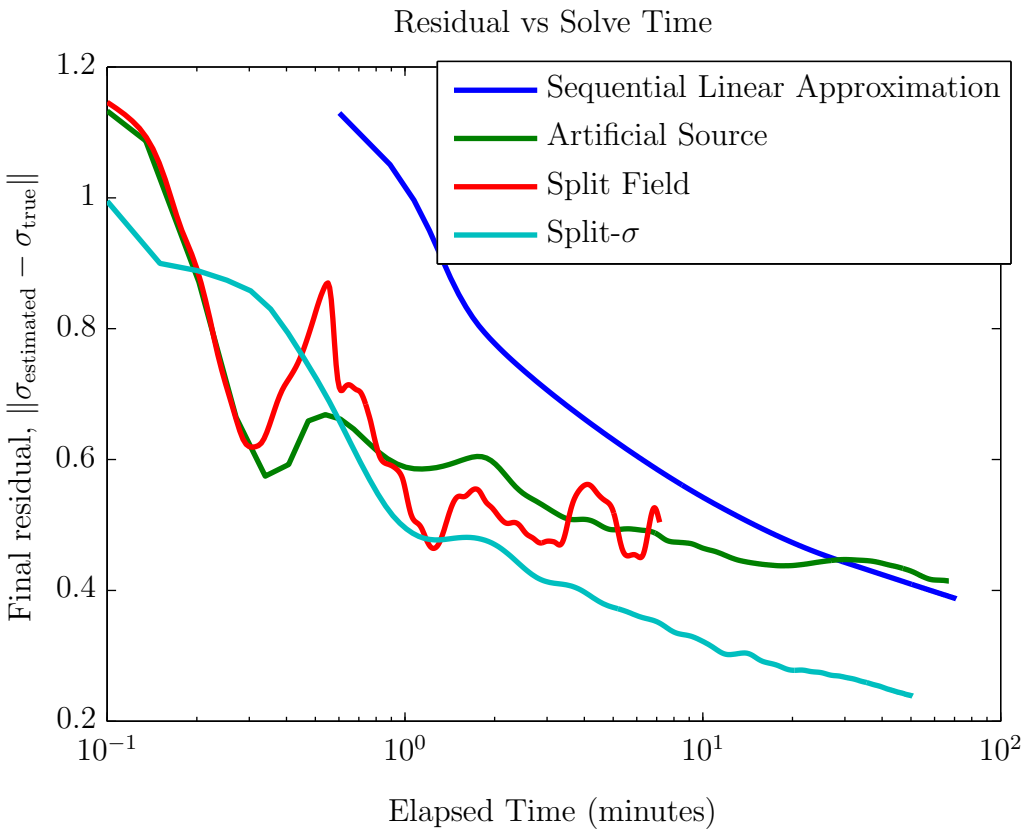

Figure 8.4: Comparison of algorithm run times and final residuals for the four methods developed and implemented for the electromagnetic subsurface imaging problem. The artificial source and sequential linear approximation methods achieve similar errors in similar times. The split field method is an order of magnitude faster, and the split-σ method achieves a lower average residual.

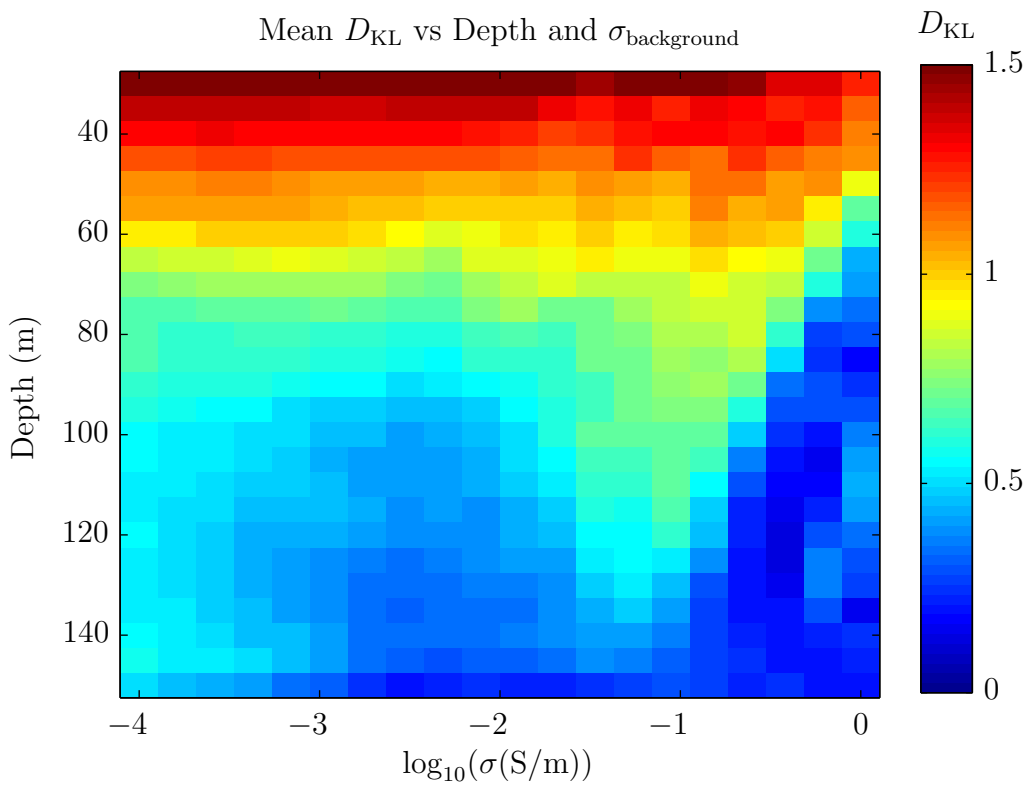

Figure 8.5: Average algorithmic accuracy as a function of depth (y-axis) and background conductivity (x-axis). Color indicates the mean Kullback-Leibler divergence between the observed algorithm error and the worst case algorithm error. Red indicates that the algorithm is accurate for that depth, background conductivity, while blue indicates that the algorithm returns poor results. Results obtained using the artificial source method and 16 illumination conditions.

8.2.2 Noise

Real-world measurements are often corrupted by noise. In this section, we explore the effect of decreasing the signal to noise ratio of the observations used as inputs into the algorithms.

$$\hat{u} = \mathcal{M}u + \nu, \quad \nu \sim \mathcal{N}(0, \sigma_{\text{noise}}).$$

Signal to noise ratio (SNR) is simply defined as the logarithm of the ratio of the norm of the signal to the norm of the noise:

$$\text{SNR} = 20 \log_{10} \left(\frac{\|\mathcal{M}(u)\|_2}{\|\nu\|_2} \right).$$

To control the signal to noise ratio, a set of random numbers is generated and scaled in order to properly control the signal to noise ratio exactly. For each conductivity profile, the true electromagnetic fields are calculated using the full model of Maxwell's equations. Then, a random noise vector is generated and scaled to ensure the proper SNR. The SNR set to be constant across all frequencies and directions of arrival. One set of noisy measurements is produced for each SNR and each conductivity profile.

The noise performance for each of the algorithms is surveyed in Figure 8.6. In each of these subplots, the average accuracy as measured by the Kullback-Leibler divergence is shown as a function of depth and background conductivity. It is clear that in these plots the performance of the algorithms dramatically decreases as the signal to noise ratio drops. Noisy measurements mean worse performance.

For the sequential linear approximation, the performance stays relatively uniform through signal to noise ratios of 50 dB. Below 50 dB, the ability of the algorithms to determine the conductivity parameters underground is significantly compromised. With such high noise levels, the probability that the underground conductivity can be accurately determined using these noisy measurements is unclear. Numerical noise leads to some spurious results and apparent accuracy at lower signal to noise ratios, but those are suspect.

Noise is present in these measurements in two forms. First, noise of the system contributes to noisy measurements of any of the sferics themselves. Each individual

sferic has a signal to noise ratio of around 40 dB. The noise here comes from background atmospheric processes and receiver noise. Averaging over more sferics will improve the signal to noise ratio in this regard. Nominally, the noise decreases with the square root of the number of measurements that are included. Thus if each sferic observation is at 40 dB SNR, then averaging over a million sferics will increase the SNR by 60 dB, giving us a 100 dB SNR.

The second source of noise with respect to measurement is alignment error. When the VLF antennas are set up in the field, we typically align them such that one axis of the antenna is in the north south direction. The sensitivity of the measurements with respect to this dimension is unknown.

8.3 Subsampling

As discussed in Section 7.2.3, the solution to the electromagnetic subsurface imaging problem cannot be uniquely determined if fewer measurements are made than conductivity unknowns to estimate. We wish to establish a subsampling principle that governes the relationship between the numer of sensors and measurements necessary at or near the surface and the number of estimable conductivity parameters underground. The space of possibilities is large for true electromagnetic subsurface imaging problems. For this reason, we develop a set of surrogate problems of smaller dimension, but similar mathematical form in order to test the algorithms developed in Chapter 7. Only the best performing algorithms were later used and applied to the full electromagnetic subsurface imaging problem.

With the insight gained from general problems, we also turn our attention to the feasible solvers and assess the performance as the amount of input information is varied in the problem.

8.3.1 Global Testing Procedure

The key change for these testing procedures is that the total number of unknowns was greatly reduced. Rather than use a 199×199 2D space to represent the grid of

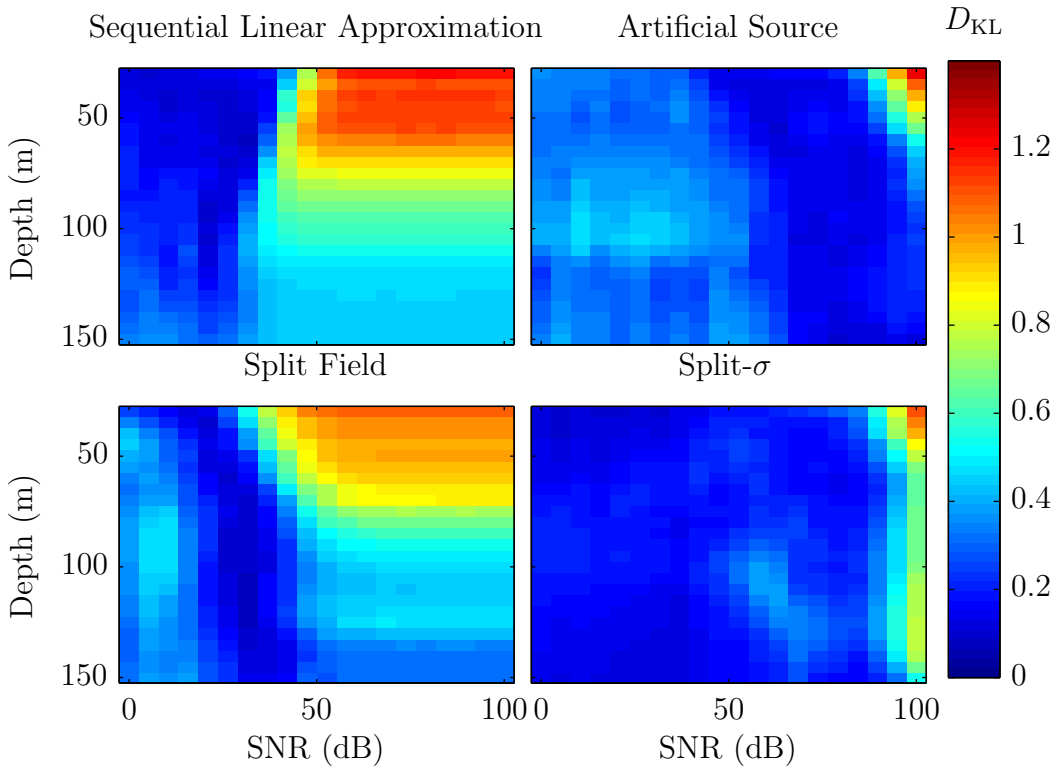

Figure 8.6: Noise performance of four electromagnetic subsurface imaging algorithms. Average accuracy is given as a function of depth and background conductivity, $\sigma_{\text{background}}$, using the Kullback-Leibler divergence metric given in Section 8.1.2. The sequential linear approximation and the split field method show the best performance with respect to noise. Both of these algorithms are able to image accurately down to SNR of 45 dB. The split-σ and artificial source methods are very sensitive to noise and for signal to noise ratios less than 90 dB, these methods perform poorly.

u, we developed models both for $u \in \mathbb{R}^{20}$ and $u \in \mathbb{C}^{20}$. In Chapter 7, algorithms for both the Real-only plane and complex plane were developed. We test in the real-only and complex domains accordingly and chose random, sparse matrices $A \in \mathbb{R}^{20\times 20}$, $A \in \mathbb{C}^{20\times 20}$, respectively. A domain sampling operator $\mathcal{M}_{\mathrm{d}}$

For each algorithm, 200 sparse random matrices for $A \in \mathbb{R}^{n\times n}$, $u^* \in \mathbb{R}^n$, $\sigma^* \in \mathbb{R}^m$ are generated. Random sampling matrices, $\mathcal{M}_r$ and $\mathcal{M}_d$ for $\hat{u}$ and σ respectively, are chosen such that the sets of r, d are random and non-overlapping in the case that $|r| + |d| < n$. In the case that $|r| + |d| > n$ some of the measurement points overlap with the material domain points.

8.3.2 Global Results

To test these algorithms, a sequence of test problems was run in order to establish baseline performance. Random matrices with the proper structure were created, measurements taken, and algorithms applied.

We report the average relative error for the field term, u, $\frac{\|u-u^*\|}{\|u^*\|}$ and the average relative error for the parameter term, σ, $\frac{\|\sigma-\sigma^*\|}{\|\sigma^*\|}$ for the problems over all combinations of number of measurements, $|s|$, and number of parameters $|d|$. Results for problems solved in the real-only regime and complex only regime are presented in Figure 8.7 and Figure 8.8 respectively. Timing results are shown in Figure 8.9 and Figure 8.10 for the real and complex solvers respectively.

These numerical results demonstrate that each of the algorithms is able to obtain the fundamental 1 to 1 subsampling ratio. That is, each of the algorithms can recover the proper conductivity results with very high accuracy if the number of measurements is greater than or equal to the number of unknowns. In both Figures 8.7 and 8.8 a white line indicates the proposed critical subsampling ratio.

In the complex case, a general 1 to 2 subsampling principle is observed. The apparent improvement in subsampling ratio is primarily due to the change in parameter spaces; we are estimating real-only parameters from complex measurements. If we treat the real and imaginary components as independent measurements, then a 1 to 1 subsampling ratio holds.

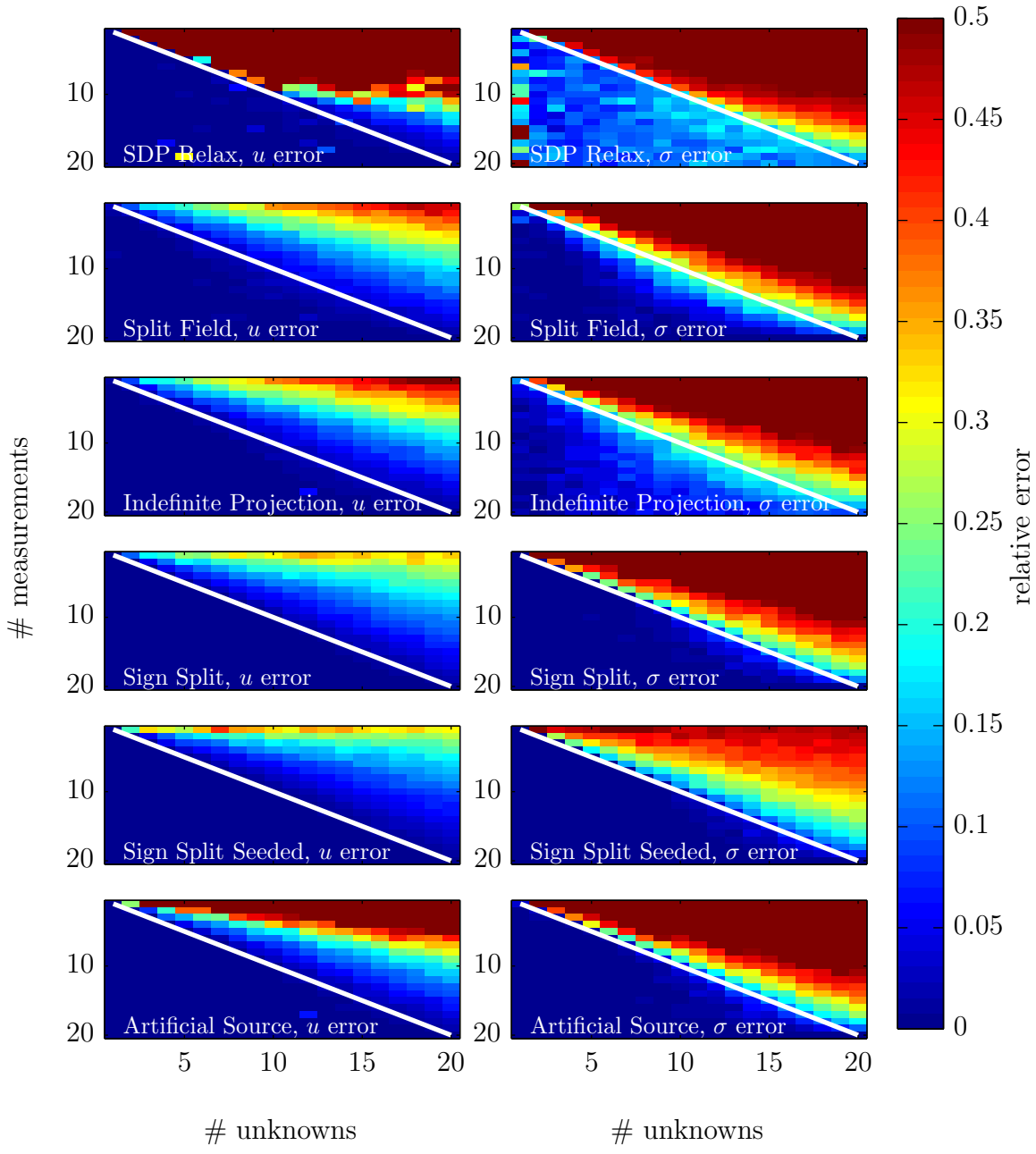

Figure 8.7: Results from 200 Monte Carlo iterations of the algorithms developed for optimization over the real plane. The $1:1$ line is plotted to indicate the transition between over-sampling and under sampling to highlight the change in performance of these algorithms. It is noteworthy that the SDP relaxation method performs poorly when the entire problem is incorporated, but using the elimination method described in Section 7.2.1 SDP relaxation performance improves dramatically.

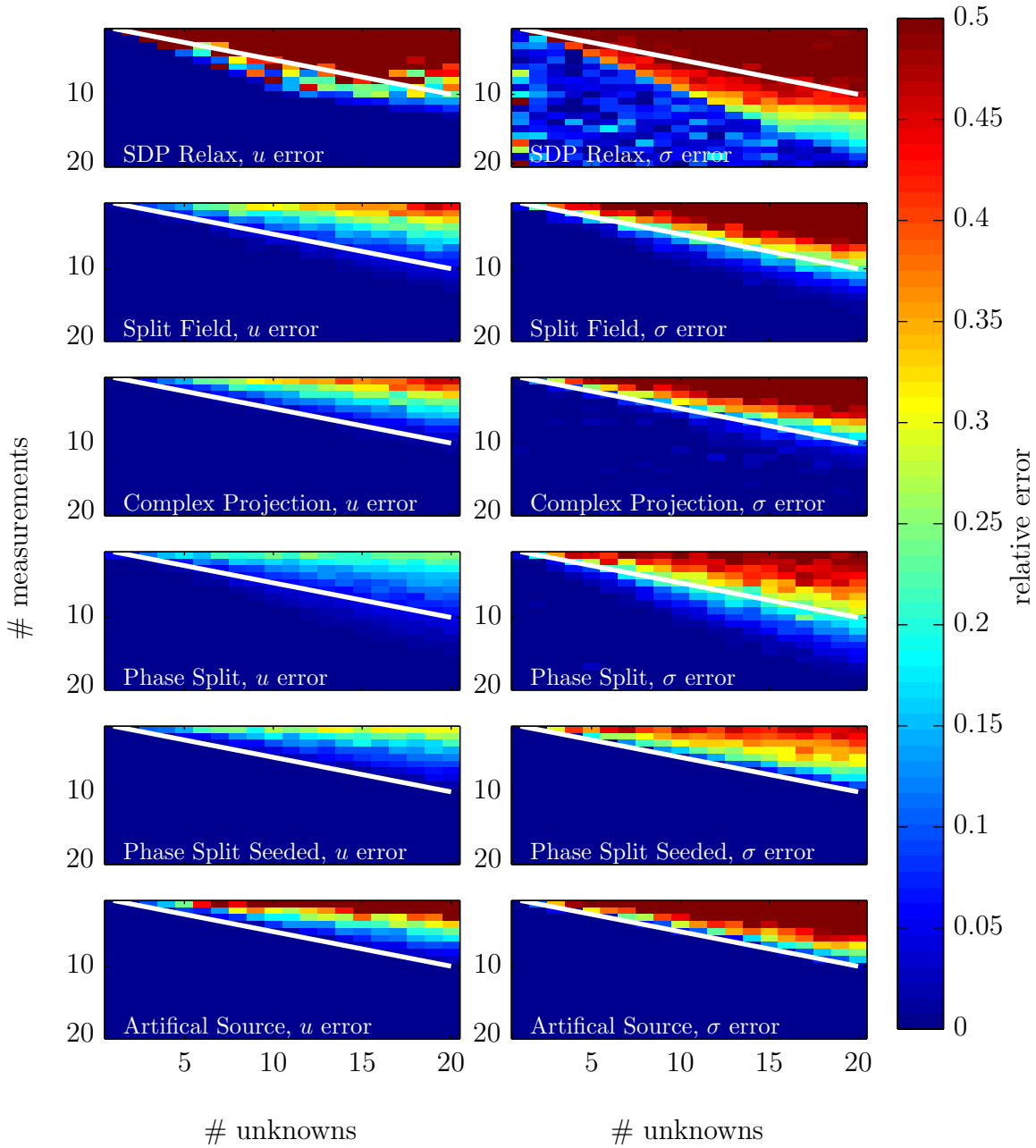

Figure 8.8: Preliminary results from Monte Carlo iterations of the algorithms developed for optimization over the complex plane. The $2:1$ line is plotted to indicate the transition between the regime where perfect reconstruction is expected due to over sampling and where problems are in an undersampling regime.

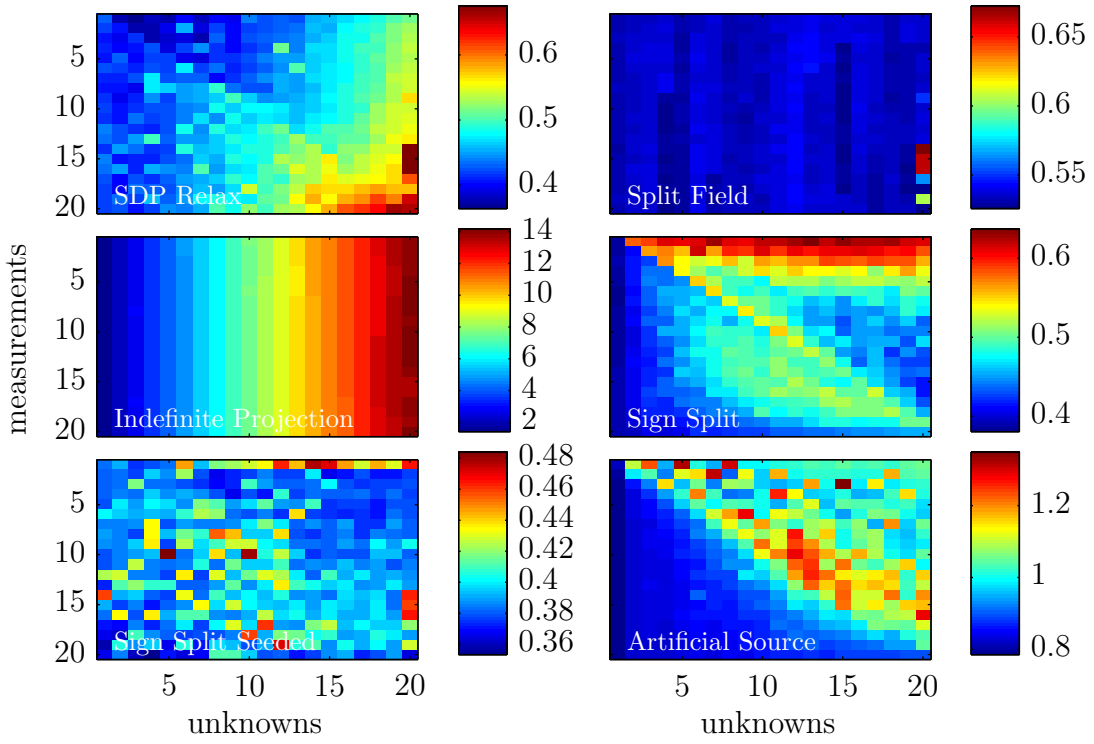

Figure 8.9: Average timing results for each of the algorithms in each of the measurement configurations for the solvers on the real plane. Times are given in cpu seconds.

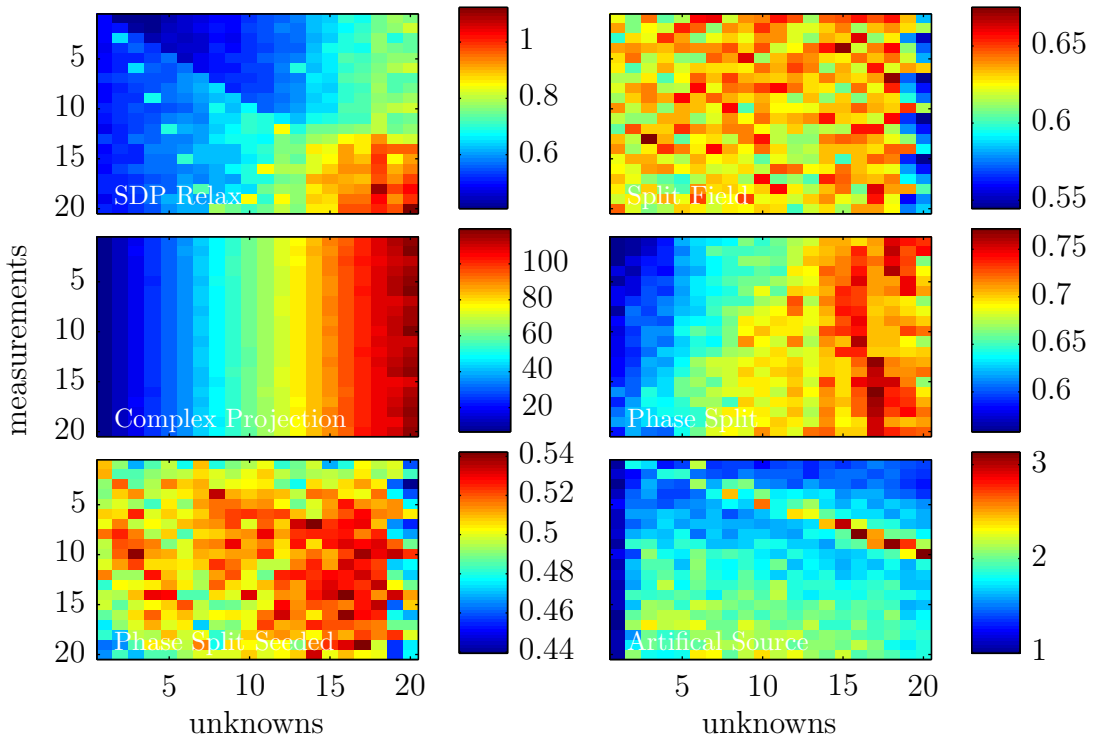

Figure 8.10: Average timing results for each of the algorithms in each of the measurement configurations for the solvers on the complex plane. Times are given in cpu seconds.

Several of the algorithms demonstrate high accuracy even when the subsampling principle is violated. That is, even when there are fewer measurements than unknowns, the algorithm is able to obtain a proper estimate of the unknowns. The seeded methods, both sign split seeded and phase split seeded demonstrate the highest accuracy away from the boundary. This accuracy is largely due to the fact that prior information is given to these solvers, namely the signs and phases respectively of the artificial sources. This additional information directly yields better performance. Other solvers, specifically the ADMM based solvers incorporate a mild self-regularization. The proximal operator provides mild regularization during the initial iterates of the algorithm, which translates to apparent improvements in performance.

The exact nature of these subsampling transitions is still open to further exploration. Results from high dimensional statistics indicate that as the problem size grows, determining the transitions between success and failure of the algorithm, especially subsampling algorithms improves. Extending the results to higher dimensions could lead to interesting results in inverse problem design. Furthermore, the random nature of the creation of these problems ensures that the measurements are truly independent measurements of the state of the PDE. In the electromagnetic imaging problems, the measurements are not truly independent, measurements along the surface are often highly correlated. The effects of these correlations are seen in Section 8.3.3.

Last we comment about the timing of these algorithms. The SDP problems are all solved using `CVX` and SeDuMi. The computational cost of these algorithms largely depends on the square of the problem dimensions and on the conditioning of the systems. Poorly conditioned systems require more internal iterations and thus more time. The projection methods generally perform slowly because projection into the indefinite and complex sets defined in Sections 7.3.2 and 7.3.2 is costly computationally. Projection into the indefinite set can be done using a simple and efficient bisection procedure, but the algorithm must be run for each parameter triplet separately. Projection into the complex indefinite set is not unique and requires setting up an interior point solver to approximate the solution. Running multiple interior point solvers at each iteration is computationally costly. Custom solvers were implemented

in Matlab, but it is possible that further optimization including custom compiled code yield significant improvements. One possibility is to use a tool like CVXGEN [103]. Unfortunately SDP is not currently supported. Due to the timing of these indefinite projection methods, they were not further developed for electromagnetic subsurface imaging. The SDP solvers also present difficulties when trying to scale to the problem sizes reasonable for electromagnetic subsurface imaging.

8.3.3 Electromagnetic Results

To investigate the effect of sampling on the results with respect to electromagnetics, we ran three tests to investigate the improvement as more information is added to the electromagnetic subsurface imaging problem. We investigated adding more information by varying the number of frequencies, the number of incident wave directions, and the number of sensors used near the surface in experiments with transverse electric illumination.

Two series of experiments were run. In the first, we jointly vary the number of incident directions and the number of frequencies. In the second, we vary the number of frequencies and the number of sensors at the surface.

Frequencies and Angles of Incidence

Results are computed and displayed for both the artificial source method (Figure 8.11) and the split field method (Figure 8.12). We take these results to be indicative for all of the solvers.

Two general trends can be observed in these plots. Firstly, as the number of illumination angles increases, the general accuracy of subsurface imaging does not increase. As information from different angles of incidence are added, the accuracy of the algorithm does not improve. We can further verify this trend by computing the covariance between different illumination angles. Waves incident from different directions of arrival are strongly correlated, therefore more observations in this dimension does not add much unique information to the electromagnetic imaging process. A small increase can be seen from one to two angles of incidence, but the improvement

to 6 angles of incidence is negligible. Secondly, as more frequencies are added to the computation, the general accuracy as a function of depth increases. As more frequency information is added to the electromagnetic inverse problem, more information is incorporated and our ability to estimate the subsurface conductivity parameters increases. These results are generally mirrored between the artificial source and split field methods, however the split field method shows a more dramatic improvement in depth imaging as more frequencies are added.

Frequencies and Number of Sensors

Results are compared for the split-σ method (Figure 8.14) and the artificial source method (Figure 8.13). These plots demonstrate that adding more sensors to the set of surface measurements only makes minute improvements after a point. A significant improvement in performance can be observed between 1 and 11 sensor measurements, but the improvement between 11 sensors and 21 sensors is not as pronounced. Adding up to 71 sensors does not dramatically improve the performance. However, adding more frequencies to the optimization problem does make a significant difference to the performance of the algorithm. The average performance of the artificial source method of Figure 8.13 is similar to the performance of Figure 8.11. The performance of the split-σ method shows dramatic improvement with very high accuracy for recovering the subsurface parameters once 6 frequencies have been added to the scope of the optimization problem.

Oversampling in Frequency

Given the trend demonstrating that adding more frequencies to the optimization problem improves the accuracy of the optimization problem more than any other dimension, we ran a test to determine the accuracy for a situation when there are more measurements than unknowns. We achieve this oversampling scenario by adding observations at enough frequencies to over determine the problem. Figure 8.15 shows the relative entropy of the error and shows that even in this over determined case, when there are more observations than unknowns in the problem, we are still not able

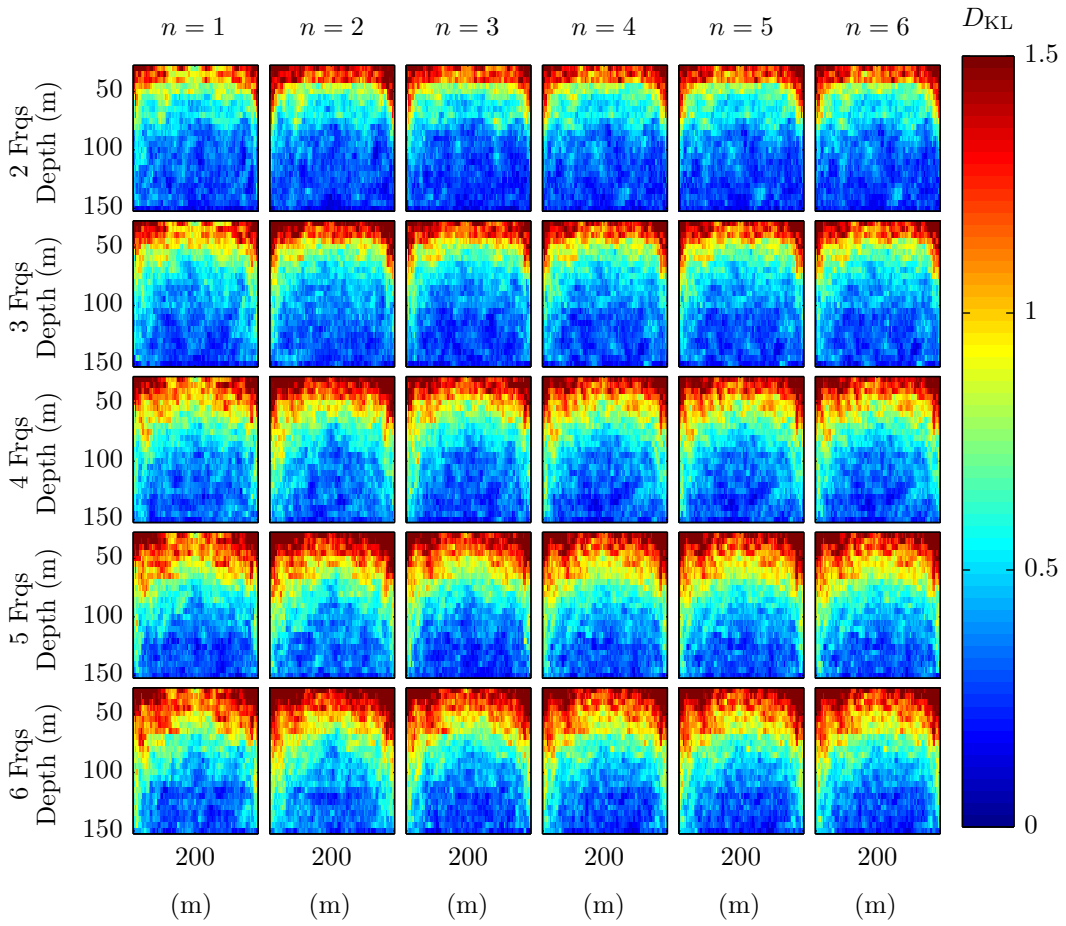

Figure 8.11: Subsampling profiles for artificial source method as a function of number of frequencies and number of incident angles. Greater improvement is seen as more illumination frequencies are added to the problem than when more angles of incidence are added to the problem.

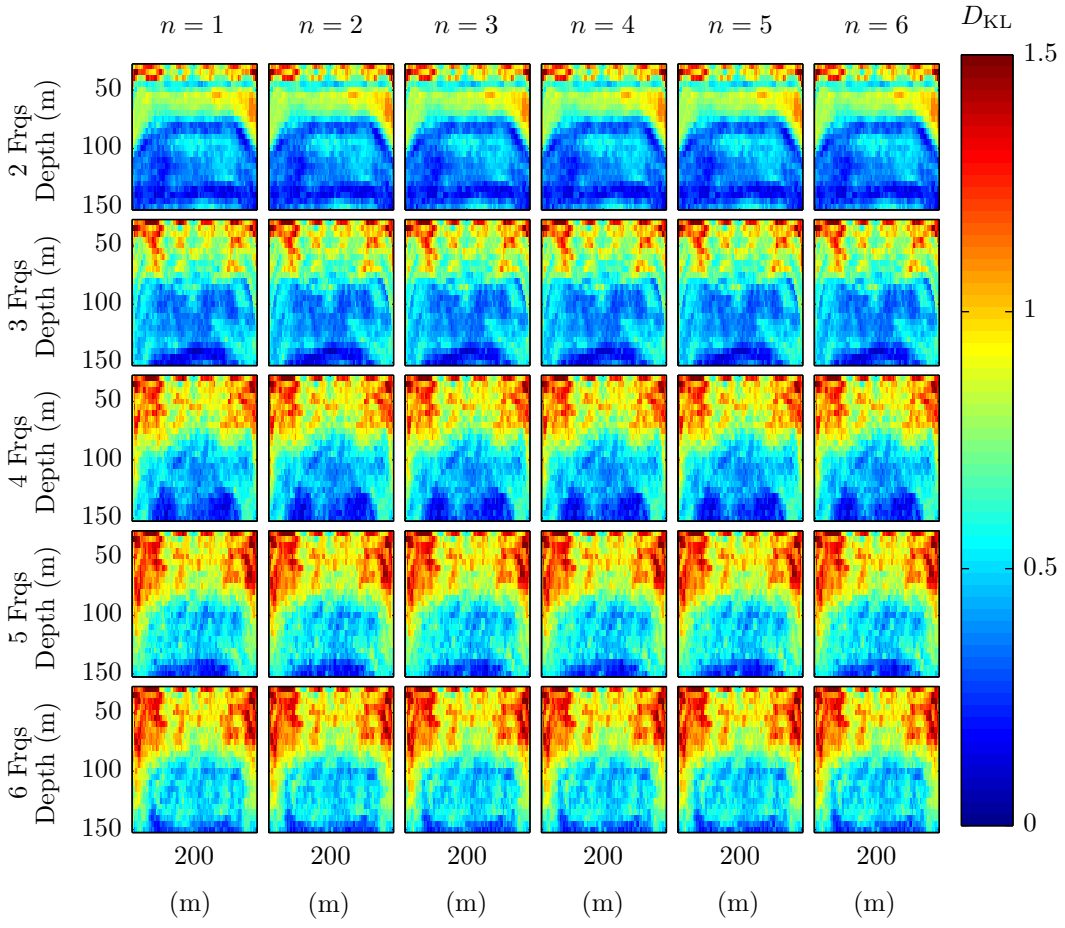

Figure 8.12: Subsampling profiles for split field method as a function of number of frequencies and number of incident angles. Greater improvement is observed as more illumination frequencies are added to the problem than when more angles of incidence are added. The split field method reconstructs with higher accuracy at greater depth than the artificial source method.

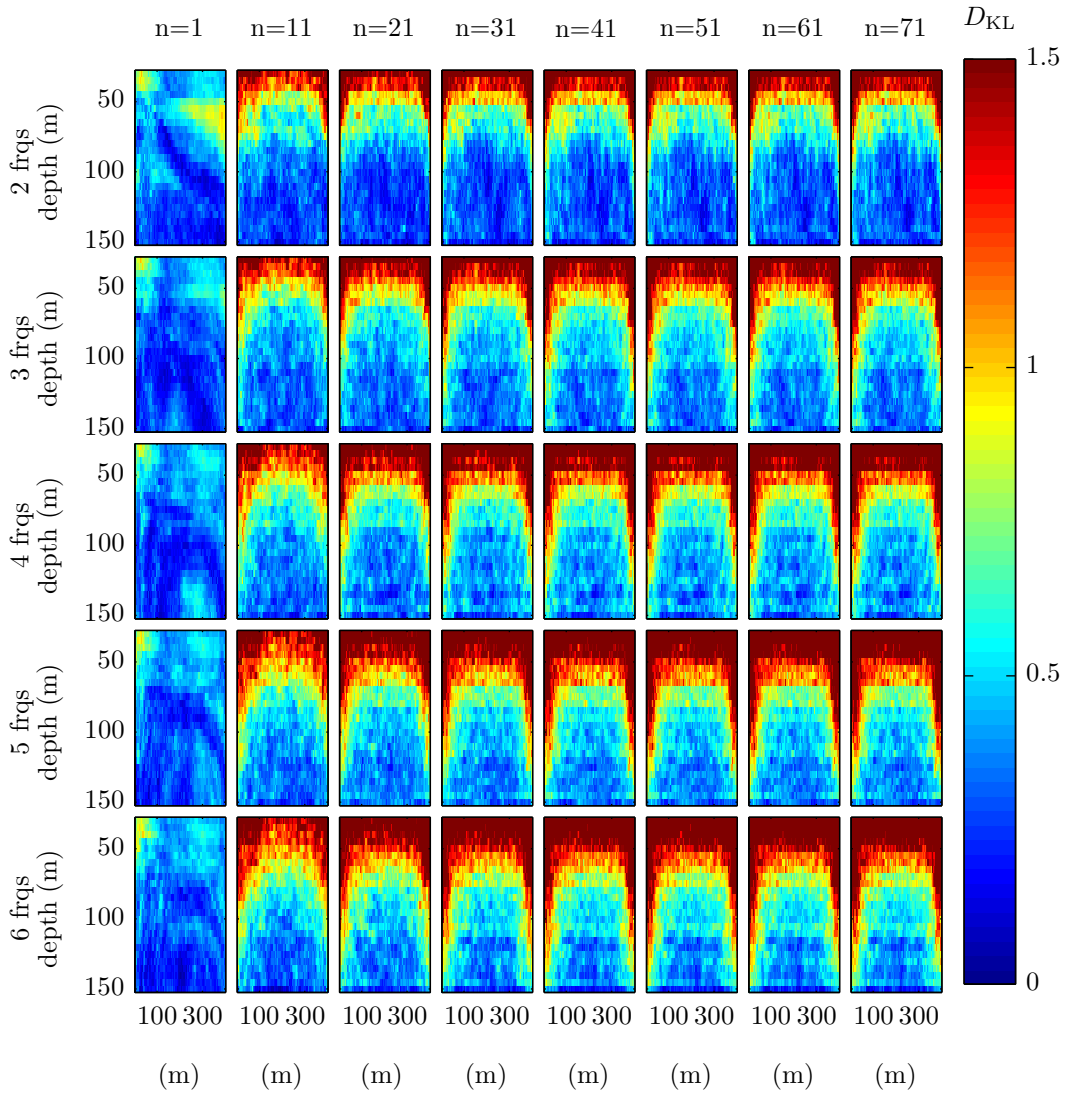

Figure 8.13: Subsampling profiles for artificial source method as a function of the number of frequencies and sampling sites at the surface of the Earth. A clear improvement is made when more than 1 sensor is used for imaging. With 11 or more sensors, adding more frequencies to the imaging problem makes a greater improvement than adding more sensors.

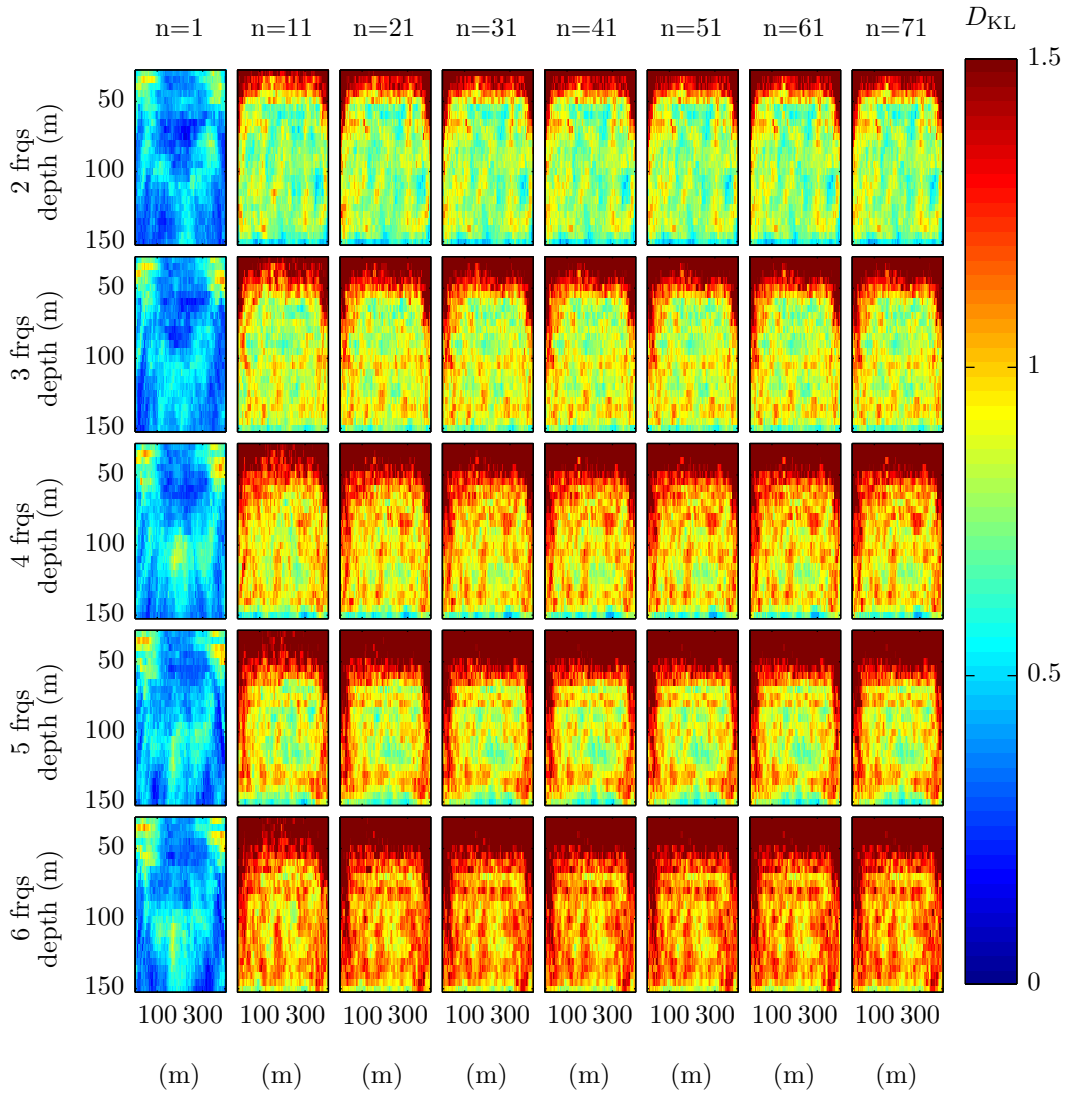

Figure 8.14: Subsampling profiles for split σ method as a function of the number of frequencies and sampling sites at the surface of the model Earth. A clear improvement is visible when more than 1 sensor is used for imaging. With 11 or more sensors, adding more frequencies to the imaging problem makes a greater improvement than adding more sensors. The split-σ method obtains substantially better performance at depth than the artificial source method.

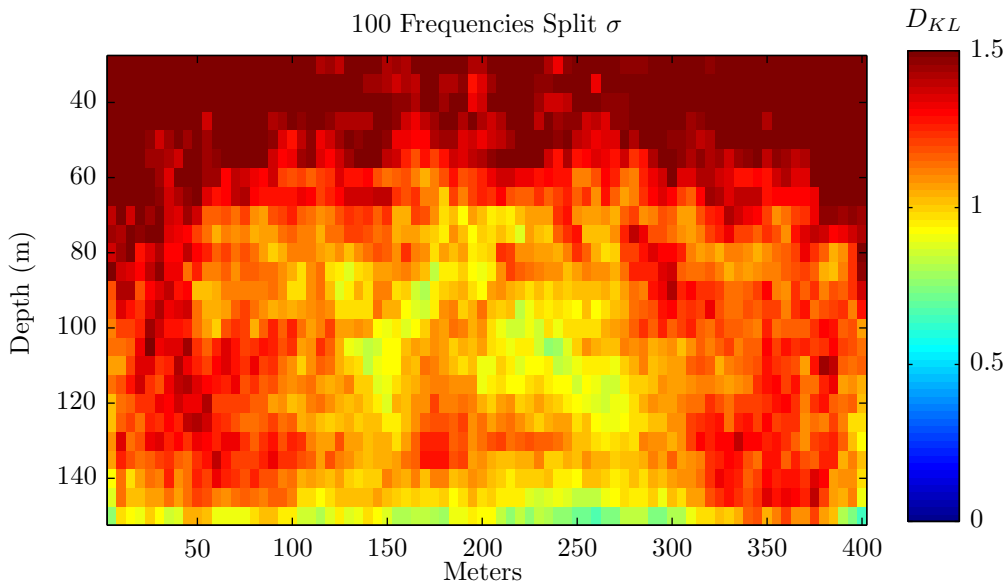

Figure 8.15: Relative entropy across 100 Monte Carlo iterations of the split-σ algorithm where observations at 100 frequencies evenly spaced between 100hz and 50kHz are incorporated. The accuracy of this algorithm is high, but not perfect. Error is still present in the reconstructed results.

to achieve perfect reconstruction in the electromagnetic subsurface imaging problem. Despite over sampling in the frequency space, the samples are not completely independent; adding more frequencies does not continue to add new, unique information to the problem. The ill-posed nature of the problem is once again apparent as it becomes clear that the systems of equations governing different frequencies of propagation are not completely independent. Despite regularization, the problem is poorly conditioned, and unique results cannot be obtained.

Distributed Optimization

In Chapter 7, distributed algorithms for electromagnetic subsurface imaging were described. Timing information about these electromagnetic subsurface imaging algorithms was collected through the course of these Monte Carlo trials. Using these timing results we are able to establish a weak scaling law for the algorithm, which

shows that as we expand the size of the problem and distribute it across more computers, the average execution time does not appreciably change. The weak scaling law gives the average execution time as the work per processor remains constant. For our algorithm, we are able to distribute more incident wave conditions across more computers in roughly constant time as shown in Figure 8.16. Slow down in the algorithms can be observed as more incident wave conditions are added. The algorithm is built such that it has bulk-synchronous structure; all independent processes must wait for completion of all the other independent processes before moving on. Therefore, if one particular node takes more time, the entire algorithm slows down. Since solving Maxwell's equations at different frequencies can take different lengths of time due to numerical precision issues, adding more frequencies to the problem can slow down the algorithm in addition to increased communications overhead.

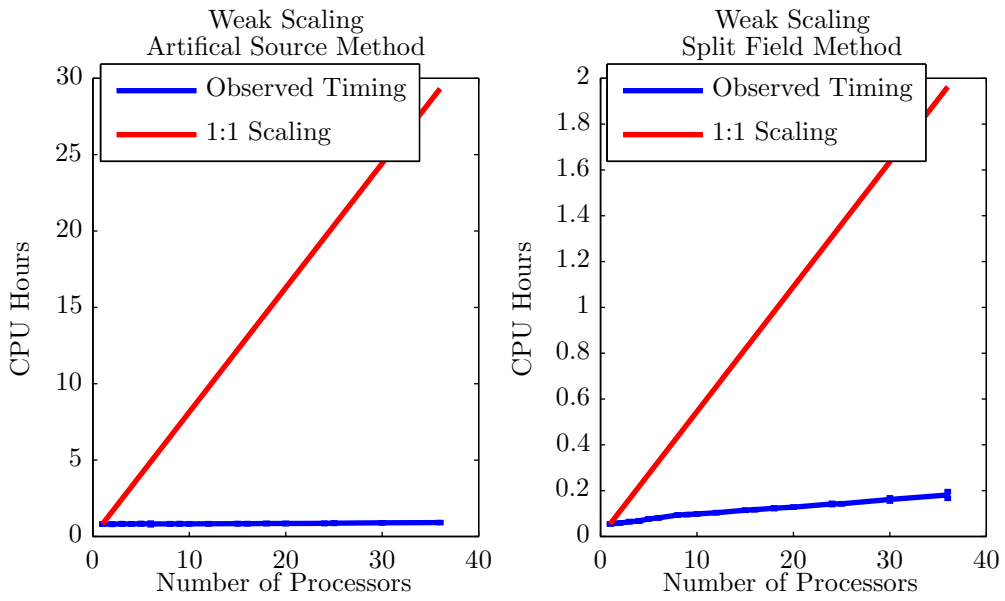

Figure 8.16: Weak scaling laws for the artificial source and split field methods. Shown are the projected time for sequential programs (red line) and the measured time for distributed problems (blue line). The split field method is much faster in general than the artificial source method. A slight slow-down can be observed in the split-field method primarily due to the fact that solving Maxwell's equations at different frequencies can take different amounts of time.

Part IV

Conclusion

Chapter 9

Discussion and Future Work

9.1 Discussion

In this thesis, we have discussed two important aspects of electromagnetic subsurface imaging: data processing and model inversion. In the two parts, we have explored the novel application of optimization algorithms to model and solve the unique problems associated with electromagnetic subsurface imaging at VLF using natural sources.

9.1.1 Sparse Separation

The application of sparse separation to the VLF dataset is a novel application of otherwise well known techniques in signal processing. By exploiting the sparsity in the occurrence of sferics and transmitters in the data record, we have developed an efficient method for separating different content types in our observations. Unlike many of the subsampling results presented in [53], [105], we discover that the wavelet-Fourier dictionary has a constant minimum achievable subsampling ratio. As we change the number of wavelet levels that are included in the sparse separation routine, the probability of recovering the sparsest solution by solving the minimum ℓ_1 problem remains remarkably constant. Direct measurement of these undersampling ratios provides confidence that when we apply this method to real data, we have achieved the sparsest solution. On average, real data separation examples produce

sparse representation vectors on the order of around 10%, which is less than the 20% sparsity boundary between accurate and inaccurate reconstruction in the overcomplete wavelet-Fourier dictionary. This method has already been met with success across multiple recordings in the VLF data library. It shows promise for both understanding the data and providing a means of compression; if the sparse vectors have a sparsity on the order of 10%, then we obtain a 5:1 compression ratio of the data, which could increase the throughput of a distributed network for monitoring VLF signals around the globe. This method for expanding over the Fourier and the wavelet bases is only one potential candidate for representation. Larger representation spaces could be used to improve accuracy and separation as we discuss below.

Development of solvers for sparse separation strongly informed our discussion and development for electromagnetic subsurface imaging methods. The detailed study of linear ill-posed problems made the study of nonlinear ill-posed problems clearer and more rigorous. Furthermore, the detailed involvement in the development of algorithms for sparse separation motivated much of the more detailed work for distributed optimization development for electromagnetic subsurface imaging.

9.1.2 Electromagnetic Subsurface Imaging

Electromagnetic subsurface imaging is an old area of research. The main contributions of this thesis include the development and formulation of electromagnetic subsurface imaging across multiple frequencies and angles of incidence as a PDE-constrained optimization problem, applying and adapting optimization algorithms to fit the specific mathematical structure of the problem, implementing these algorithms, and testing the algorithms across a wide range of models and acquisition conditions. The repeated Monte Carlo evaluation algorithms for electromagnetic inverse problems is another unique feature of the work presented here. To the author's knowledge, no other work in the literature has provided testing results as complete as the numerical results presented here. These Monte Carlo results give us more confidence and more information about the performance of the algorithm than solving any single problem as is typically presented in the literature.

Our Monte Carlo iterations show that the methods developed in this thesis have different accuracy profiles. Under normal conditions, $\sigma_{\text{background}} = 0.001$ S/m and 16 illumination conditions, the sequential linear approximation estimates conductivity with reasonable accuracy to depths of 70 meters. The artificial source method and the split field method image to 50 and 70 meters respectively. The split-σ method achieves the highest accuracy and is able to estimate the conductivity with reasonable accuracy down to 150 meters. In none of these cases is recovery of the true conductivity guaranteed. Furthermore, we can establish that of the algorithms, the split field method is the fastest; it takes about an order of magnitude less time than any of the other methods.

As we change the background parameters, *i.e.*, introduce noise and change the background conductivity, the performance of the algorithms changes. The sequential linear approximation and the split field method prove to be the most robust to noise and enable the estimation of conductivities at depths of 70 meters when the SNR is as low as 45 dB. The artificial source and split-σ method show much higher sensitivity to noise and are tolerant only of the smallest amount of noise 100 dB SNR before subsurface estimates are severely compromised. Increasing the background conductivity also decreases the accuracy of estimates at depth. For background conductivities greater than 0.1 S/m, Monte Carlo tests with the artificial source method show that the accuracy of estimating subsurface perturbations decreases dramatically.

Lastly, we note the general trend that as more information dimensions - more sensor measurements, more frequencies, more directions of arrival - are added to the optimization problem, the accuracy of estimating subsurface conductivity perturbations at depth increases. The most dramatic increase can be seen as a greater number of illumination frequencies are added to the problem. This result makes sense since in our tests, observations across multiple frequencies show less correlation than observations from more directions of arrival. The inclusion of 36 different illumination conditions in the artificial source method enables accurate imaging to 90 meters. Using the split-σ method, 6 frequencies and 71 sensors at the surface enables highly accurate reconstruction throughout the subsurface profile, or down to 150 meters. Interestingly, adding more and more frequencies does not lead to a condition in which

there is enough information unique determination of the subsurface conductivities. If we add observations from 100 different frequencies and 30 sensors at the surface we have a total of 3000 observations in order to determine 2000 subsurface conductivity parameters. Even in such a case, the exact conductivity parameters are not recovered as shown in Figure 8.15.

Incorporation of more illumination conditions comes at very low computational cost. The distributed nature of the algorithm allows us to add more illumination conditions to the problem in parallel. It takes roughly the same amount of time to compute the solution with 36 illumination conditions as it does with 1 illumination condition.

While our results are limited to the 2D electromagnetic models, this limitation is primarily due to computational resource issues. The optimization methods discussed herein can be extended to 3D models with no additional mathematical work.

9.2 Future Work

9.2.1 Real Data

The primary goal of electromagnetic subsurface imaging is to actually find mineral resources underground and map the location and distribution of real conductivity anomalies. Data from real field campaigns has been collected during the course of this thesis work, but have not been used for subsurface electromagnetic imaging for two reasons. Firstly, the results of Section 8.3.3 make it clear that many simultaneous, phase-locked sensors are required for field observations. The majority of the field campaigns conducted during the course of this thesis did not involve such a set of sensors. At most, there were times when three sensors were simultaneously recording data in a synchronous fashion. Due to this lack of synchronous data, and low number of sensors, noise and data quality issues are prohibitive for truly performing electromagnetic subsurface imaging with the data collected.

Secondly electromagnetic subsurface imaging is intrinsically a three-dimensional process. The methods and algorithms developed here can be extended to problems

involving three dimensional geometry. The main difference is the incorporation of a three-dimensional solver for Maxwell's equations. All attempts to do so have run into numerical difficulties. The intrinsic difficulty of computing solutions to Maxwell's equations for three dimensional situations has been acknowledged in the literature [126], [37]. Some approaches have been tried, but most lie beyond the scope of this thesis. Some attempts have been made for small three dimensional problems, but even for these small problems, computational costs are high and the final results are coarse.

9.2.2 Three Dimensional Solvers

Three dimensional electromagnetic models of arbitrary material parameters remain a difficult problem in computational physics. The optimization methods developed in this thesis can be extended to a 3D model without any change in mathematical structure. At this point, a 3D computation of Maxwell's equations with any sufficient resolution is beyond our computational capabilities. Recent work has confirmed this upper bound on the size of 3D electromagnetic problems [134]. Developments have been made for finite difference frequency domain solvers and preconditioners for fully 3D models [126]. Further development of these results is necessary to adapt these methods to low-frequency, sub-wavelength simulations with unstructured media. Alternative methods may be more computationally feasible, adaptable, and accurate, for example finite element methods [37].

9.2.3 Hierarchical Convolutional Methods

Separation and identification of disparate content types is a continuously evolving field in machine learning. Sparse separation over a mathematically defined dictionary consisting of the wavelet transform and the Fourier transform is a first step toward understanding and decomposing the data. Work in this thesis has shown that moving to bigger, larger data sizes can improve the performance of this specific algorithm as is. However, the algorithm reaches a natural upper limit with respect to its computability and tractability. As the size of the data exceeds $2^{23} \simeq 8 \times 10^6$, the run time of the

algorithm slows to hundreds of seconds, that is, it takes more than 6 times realtime processing. This relative processing time can only get worse as larger data sizes are used as the theoretical worst case computational complexity grows as $\mathcal{O}(n^3)$. Parallelization strategies can help the computational burden of these methods and ultimately allow for bigger datasets to be processed, but parallelization will not lower the computational complexity of the algorithm.

However, more powerful data decompositions are available. It is possible to learn a dictionary of the elements in the data themselves using more powerful, more adaptive algorithms. One such promising approach is given by the set of methods commonly know as convolutional nets [91]. This computational model allows for a compact representation of the data; a sequence of convolutional weights w that represent the data in a sparse way are found through iterative application of the ℓ_1 minimization problem and a weight update step. Multiple layers of this algorithm have been wrapped into nonlinear layers and used for powerful image recognition applications. Extensions of these methods to the large scale time series data available at VLF may lead to interesting results and data decomposition principles.

Bibliography

[1] Gassan S Abdoulaev, Kui Ren, and Andreas H Hielscher. Optical tomography as a PDE-constrained optimization problem. *Inverse Problems*, 21(5):1507–1530, October 2005.

[2] A. Abubakar, T. M. Habashy, V. L. Druskin, L. Knizhnerman, and D. Alumbaugh. 2.5D forward and inverse modeling for interpreting low-frequency electromagnetic measurements. *Geophysics*, 73(4):F165, 2008.

[3] A Abubakar, T M Habashy, M Li, and J Liu. Inversion algorithms for large-scale geophysical electromagnetic measurements. *Inverse Problems*, 25(12):123012, December 2009.

[4] A Abubakar, W Hu, P M van den Berg, and T M Habashy. A finite-difference contrast source inversion method. *Inverse Problems*, 24(6):065004, December 2008.

[5] Aria Abubakar and Tarek M Habashy. Application of the MR-CSI Method for Three-Dimensional Imaging of the Triaxial Induction Measurements. *IEEE Transactions on Geoscience and Remote Sensing*, 48(6):2613–2619, June 2010.

[6] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, November 2006.

[7] V Akcelik, G Biros, and O Ghattas. Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation. *Supercomputing, ACM/IEEE 2002 Conference*, pages 1–15, 2002.

[8] R. Barr, D.Llanwyn Jones, and C.J. Rodger. ELF and VLF radio waves. *Journal of Atmospheric and Solar-Terrestrial Physics*, 62(17-18):1689–1718, November 2000.

[9] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183, 2009.

[10] Jean-Pierre Berenger. A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, 114(2):185–200, October 1994.

[11] Peter M Van Den Berg and Ralph E Kleinman. A contrast source inversion method. *Inverse Problems*, 13(6):1607–1620, December 1997.

[12] Michael W. Berry, Murray Browne, Amy N. Langville, V. Paul Pauca, and Robert J. Plemmons. Algorithms and applications for approximate nonnegative matrix factorization. *Computational Statistics and Data Analysis*, 52(1):155–173, September 2007.

[13] D.P. Bertsekas and J.N. Tsitsiklis. *Parallel and distributed computation: numerical methods*. Prentice Hall, 1989.

[14] L.T. Biegler. *Large-Scale Pde-Constrained Optimization*. Lecture Notes in Computational Science and Engineering, 30. Springer Verlag, 2003.

[15] George Biros and Omar Ghattas. Parallel Lagrange–Newton–Krylov–Schur methods for PDE-constrained optimization. Part I: the Krylov–Schur solver. *SIAM Journal on Scientific Computing*, 27(2):687, 2005.

[16] George Biros and Omar Ghattas. Parallel Lagrange–Newton–Krylov–Schur Methods for PDE-Constrained Optimization. Part II: The Lagrange–Newton Solver and Its Application to Optimal Control of Steady Viscous Flows. *SIAM Journal on Scientific Computing*, 27(2):714–739, January 2005.

[17] Jérôme Bobin, J.-L. Starck, J.M. Fadili, Yassir Moudden, and D.L. Donoho. Morphological Component Analysis: An adaptive thresholding strategy. *IEEE Transactions on Image Processing*, 16(11):2675–2681, November 2007.

[18] S.P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[19] Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory*, 52(6):2508–2530, June 2006.

[20] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2010.

[21] R.N. Bracewell. *The Fourier transform and its applications*. McGraw-Hill series in electrical and computer engineering. McGraw Hill, 2000.

[22] Nuno B Brás, J Bioucas-Dias, Raul C Martins, and a C Serra. An alternating direction algorithm for total variation reconstruction of distributed parameters. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 21(6):3004–16, June 2012.

[23] Alfred M. Bruckstein, David L. Donoho, and Michael Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 51(1):34, 2009.

[24] K.G. Budden. *The Propagation of Radio Waves: The Theory of Radio Waves of Low Power in the Ionosphere and Magnetosphere*. Cambridge University Press, 1988.

[25] E.J. Candes, Justin Romberg, and Terence Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, February 2006.

[26] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, August 2006.

[27] Emmanuel J. Candès, Michael B. Wakin, and Stephen P. Boyd. Enhancing sparsity by reweighted $\ell 1$ minimization. *Journal of Fourier Analysis and Applications*, 14(5-6):877–905, October 2008.

[28] Don L. Carpenter. The history of very low frequency (vlf) radio research at stanford. 2012.

[29] Stéphanie Chaillat and George Biros. FaIMS: A fast algorithm for the inverse medium problem with multiple frequencies and multiple sources for the scalar Helmholtz equation. *Journal of Computational Physics*, 231(12):4403–4421, June 2012.

[30] N Champagne. FDFD: A 3D finite-difference frequency-domain code for electromagnetic induction tomography. *Journal of Computational Physics*, 170(2):830–848, July 2001.

[31] Venkat Chandrasekaran and MI Jordan. Computational and statistical tradeoffs via convex relaxation. *Arxiv Preprint*, pages 1–27, 2012.

[32] JT Chen, US Inan, and TF Bell. VLF strip holographic imaging of lightning-associated ionospheric disturbances. *Radio Science*, 31(2):335–348, 1996.

[33] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders. Atomic decomposition by basis pursuit. *SIAM Review*, 43(1):129, June 2001.

[34] M. Chevalier and U. Inan. A Technique for Efficiently Modeling Long-Path Propagation for Use in Both FDFD and FDTD. *Antennas and Wireless Propagation Letters*, 5(1):525–528, December 2006.

[35] W C Chew and Y M Wang. Reconstruction of two-dimensional permittivity distribution using the distorted Born iterative method. *IEEE transactions on medical imaging*, 9(2):218–25, January 1990.

[36] W.C. Chew. *Waves and fields in inhomogeneous media*. IEEE Press series on electromagnetic waves. IEEE Press, 1999.

[37] Weng Cho Chew and Li Jun Jiang. Overview of Large-Scale Computing: The Past, the Present, and the Future. *Proceedings of the IEEE*, 101(2):1–15, 2012.

[38] Weng Cho Chew and William H. Weedon. A 3D perfectly matched medium from modified maxwell's equations with stretched coordinates. *Microwave and Optical Technology Letters*, 7(13):599–604, September 1994.

[39] Hansen Per Christian. *Rank-Deficient and Discrete Ill-Posed Problems*. Society for Industrial and Applied Mathematics, 1998.

[40] Hugh J. Christian, Richard Blakeslee, Dennis Boccippio, William Boeck, Dennis Buechler, Kevin Driscoll, Steven Goodman, John Hall, William Koshak, Douglas Mach, and Michael Stewart. Global frequency and distribution of lightning as observed from space by the Optical Transient Detector. *Journal of Geophysical Research*, 108(D1):4005, 2003.

[41] MB Cohen and RK Said. Mitigation of 50 60 Hz power line interference in geophysical data. *Radio Science*, 45(July):1–12, 2010.

[42] Morris B. Cohen, Umran S. Inan, and Evans W. Paschal. Sensitive broadband ELF/VLF radio reception with the AWESOME instrument. *IEEE Transactions on Geoscience and Remote Sensing*, 48(1):3–17, January 2010.

[43] Steven Constable and Chester J. Weiss. Mapping thin resistors and hydrocarbons with marine EM methods: Insights from 1D modeling. *Geophysics*, 71(2):G43–G51, March 2006.

[44] Benjamin R. T. Cotts and Umran S. Inan. VLF observation of long ionospheric recovery events. *Geophysical Research Letters*, 34(14):L14809, July 2007.

[45] Tie Jun Cui, A.A. Aydiner, Weng Cho Chew, D.L. Wright, and D.V. Smith. Three-Dimensional Imaging of Buried Objects in Very Lossy Earth by Inversion of VETEM Data. *IEEE Transactions on Geoscience and Remote Sensing*, 41(10):2197–2210, October 2003.

[46] T.J. Cui, W.C. Chew, A.A. Aydiner, and S. Chen. Inverse scattering of two-dimensional dielectric objects buried in a lossy earth using the distorted Born iterative method. *IEEE Transactions on Geoscience and Remote Sensing*, 39(2):339–346, 2001.

[47] T.J. Cui, W.C. Chew, X.X. Yin, and W. Hong. Study of Resolution and Super Resolution in Electromagnetic Imaging for Half-Space Problems. *IEEE Transactions on Antennas and Propagation*, 52(6):1398–1411, June 2004.

[48] I Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory*, 36(5):961–1005, 1990.

[49] David Donoho and Jared Tanner. Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1906):4273–93, November 2009.

[50] David L. Donoho and Iain M. Johnstone. Minimax estimation via wavelet shrinkage. *The Annals of Statistics*, 26(3):879–921, June 1998.

[51] David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences of the United States of America*, 106(45):18914–9, November 2009.

[52] David L Donoho, Arian Maleki, and Andrea Montanari. The noise-sensitivity phase transition in compressed sensing. *IEEE Transactions on Information Theory*, 57(10):6920–6941, October 2011.

[53] David L Donoho and Jared Tanner. Precise undersampling theorems. *Proceedings of the IEEE*, 98(6):913–924, June 2010.

[54] D.L. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, May 1995.

[55] D.L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, April 2006.

[56] D.L. Donoho, M. Elad, and V.N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory*, 52(1):6–18, January 2006.

[57] D.L. Donoho and Xiaoming Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 47(7):2845–2862, 2001.

[58] O Dorn, H Bertete-Aguirre, and JG Berryman. A nonlinear inversion method for 3D electromagnetic imaging using adjoint fields. *Inverse*, 1523, 1999.

[59] Richard Duda, Peter Hart, and David Stork. *Pattern Classification*. John Wiley and Sons, 2001.

[60] S.L. Durden, J.D. Klein, and H.A. Zebker. Polarimetric radar measurements of a forested area near Mt. Shasta. *IEEE Transactions on Geoscience and Remote Sensing*, 29(3):444–450, May 1991.

[61] Gary D. Egbert and Anna Kelbert. Computational recipes for electromagnetic inverse problems. *Geophysical Journal International*, pages no–no, January 2012.

[62] J Esser, M Moller, S Osher, G Sapiro, and J Xin. A convex model for nonnegative matrix factorization and dimensionality reduction on physical space. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, pages 1–15, March 2012.

[63] M Jalal Fadili, Jean-Luc Starck, Jerome Bobin, and Yassir Moudden. Image decomposition and separation using sparse representations: an overview. *Proceedings of the IEEE*, 98(6):983–994, June 2010.

[64] F. R. Foust, M. Spasojevic, T. F. Bell, and U. S. Inan. Modeling scattering from lightning-induced ionospheric disturbances with the discontinuous Galerkin method. *Journal of Geophysical Research*, 116(A12):1–13, December 2011.

[65] J.J. Fuchs. Recovery of exact sparse representations in the presence of bounded noise. *IEEE Transactions on Information Theory*, 51(10):3601–3608, October 2005.

[66] Vidya Ganapati, Owen D. Miller, and Eli Yablonovitch. Inverse electromagnetic design for subwavelength light trapping in solar cells. *IEEE Photonics Conference 2012*, 2(c):191–192, September 2012.

[67] D. I. Golden, M. Spasojevic, and U. S. Inan. Determination of solar cycle variations of midlatitude ELF/VLF chorus and hiss via automated signal detection. *Journal of Geophysical Research*, 116(A3):1–15, March 2011.

[68] G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 1996.

[69] M. Grant and S. Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008.

[70] M Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 1.21. `http://cvxr.com/cvx`, April 2011.

[71] Julien Guillemoteau, Pascal Sailhac, and Mickael Behaegel. Fast approximate 2D inversion of airborne TEM data: Born approximation and empirical approach. *Geophysics*, 77(4):WB89, 2012.

[72] T.M. Habashy, R. W. Groom, and B.R. Spies. Beyond the Born and Rytov Approximations: A Nonlinear Approach to Electromagnetic Scattering. *Journal of Geophysical Research*, 98(B2):1759–1775, 1992.

[73] E. Haber and U.M. Ascher. Preconditioned all-at-once methods for large , sparse parameter estimation problems. *Inverse Problems*, 17(6):1847, 2001.

[74] Eldad Haber, Uri M. Ascher, and Douglas W. Oldenburg. Inversion of 3D electromagnetic data in frequency and time domain using an inexact all-at-once approach. *Geophysics*, 69(5):1216, 2004.

[75] Eldad Haber, Douglas W. Oldenburg, and R. Shekhtman. Inversion of time domain three-dimensional electromagnetic data. *Geophysical Journal International*, 171(2):550–564, November 2007.

[76] Aapo Hyvrinen, Juha Karhunen, and Erkki Oja. *Independent Component Analysis*. J. Wiley, New York, 2001.

[77] U. S. Inan, S. a. Cummer, and R. a. Marshall. A survey of ELF and VLF research on lightning-ionosphere interactions and causative discharges. *Journal of Geophysical Research*, 115:A00E36, June 2010.

[78] U.S. Inan and A.S. Inan. *Electromagnetic Waves*. Prentice Hall, 2000.

[79] U.S. Inan and R.A. Marshall. *Numerical Electromagnetics: The FDTD Method*. Numerical Electromagnetics: The FDTD Method. Cambridge University Press, 2011.

[80] Kazufumi Ito and Karl Kunisch. The augmented Lagrangian method for parameter estimation in elliptic systems. *SIAM Journal of Control and Optimization*, 28(1):113–136, 1990.

[81] Bangti Jin and Peter Maass. Sparsity regularization for parameter identification problems. *Inverse Problems*, 28(12):123001, December 2012.

[82] MP Johnson, US Inan, and DS Lauben. Subionospheric VLF Signatures of Oblique (Nonducted) Whistler-Induced Precipitation. *Geophysical research letters*, 26(23):3569–3572, 1999.

[83] H.M. Jol. *Ground Penetrating Radar Theory and Applications*. Elsevier Science, 2008.

[84] C.T. Kelley. *Iterative Methods for Linear and Nonlinear Equations*. Frontiers in Applied Mathematics. CAMBRIDGE - USA, 1995.

[85] Kerry Key, Steven Constable, Lijun Liu, and Anne Pommier. Electrical image of passive mantle upwelling beneath the northern East Pacific Rise. *Nature*, 495(7442):499–502, March 2013.

[86] Dongmin Kim, Suvrit Sra, and Inderjit S Dhillon. Tackling Box-Constrained Optimization via a New Projected Quasi-Newton Approach. *SIAM Journal on Scientific Computing*, 32(6):3548, 2010.

[87] Hyun Keol Kim and Andreas H Hielscher. A PDE-constrained SQP algorithm for optical tomography based on the frequency-domain equation of radiative transfer. *Inverse Problems*, 25(1):015010, January 2009.

[88] Seung-jean Kim, K Koh, M Lustig, Stephen Boyd, and Dimitry Gorinevsky. An interior-point method for large-scale l_1-regularized least squares. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):606–617, December 2007.

[89] D. S. Lauben, U. S. Inan, and T. F. Bell. Precipitation of radiation belt electrons induced by obliquely propagating lightning-generated whistlers. *Journal of Geophysical Research*, 106(A12):29745, 2001.

[90] D D Lee and H S Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–91, October 1999.

[91] Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y. Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proceedings of the 26th Annual International Conference on Machine Learning - ICML '09*, pages 1–8, New York, New York, USA, 2009. ACM Press.

[92] Jason D Lee, Yuekai Sun, and Michael A Saunders. Proximal Newton-type methods for convex optimization. pages 1–25, June 2012.

[93] Nicolas Lee, Sigrid Close, Ashish Goel, David Lauben, Ivan Linscott, Theresa Johnson, David Strauss, Sebastian Bugiel, Anna Mocker, and Ralf Srama. Theory and experiments characterizing hypervelocity impact plasmas on biased spacecraft materials. *Physics of Plasmas*, 20(3):032901, 2013.

[94] N. G. Lehtinen and U. S. Inan. Radiation of ELF/VLF waves by harmonically varying currents into a stratified ionosphere with application to radiation by a modulated electrojet. *Journal of Geophysical Research*, 113(A6):A06301, June 2008.

[95] Nikolai G. Lehtinen and Umran S. Inan. Full-wave modeling of transionospheric propagation of VLF waves. *Geophysical Research Letters*, 36(3):L03104, February 2009.

[96] L. O. Lø seth and B. Ursin. Electromagnetic fields in planarly layered anisotropic media. *Geophysical Journal International*, 170(1):44–80, July 2007.

[97] Jesse Lu and Jelena Vucković. Inverse design of nanophotonic structures using complementary convex optimization. *Optics express*, 18(4):3793–804, February 2010.

[98] Jesse Lu and Jelena Vučković. Objective-first design of high-efficiency, small-footprint couplers between arbitrary nanophotonic waveguide modes. *Optics Express*, 20(7):7221, March 2012.

[99] C. Mackay and a. C. Fraser-Smith. Lightning location using the slow tails of sferics. *Radio Science*, 45(5):n/a–n/a, October 2010.

[100] A. MacOvski. *Medical Imaging Systems*. Prentice Hall Information and System Sci Series. Pearson Education, Limited, 1983.

[101] Stephane Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 2009.

[102] R. a. Marshall and U. S. Inan. Two-dimensional frequency domain modeling of lightning EMP-induced perturbations to VLF transmitter signals. *Journal of Geophysical Research*, 115:A00E29, June 2010.

[103] Jacob Mattingley and Stephen Boyd. CVXGEN: a code generator for embedded convex optimization. *Optimization and Engineering*, 13(1):1–27, November 2011.

[104] Hagit Messer, Artem Zinevich, and Pinhas Alpert. Environmental monitoring by wireless communication networks. *Science (New York, N.Y.)*, 312(5774):713, May 2006.

[105] Hatef Monajemi, Sina Jafarpour, Matan Gavish, and David L Donoho. Deterministic matrices matching the compressed sensing phase transitions of Gaussian random matrices. *Proceedings of the National Academy of Sciences of the United States of America*, 110(4):1181–6, January 2013.

[106] M.N. Nabighian. *Electromagnetic Methods in Applied Geophysics: Application/Parts A and B*. Investigations in Geophysics Series. Soc of Exploration Geophysicists, 1988.

[107] A. Nedic, Asuman Ozdaglar, and P.A. Parrilo. Constrained Consensus and Optimization in Multi-Agent Networks. *IEEE Transactions on Automatic Control*, 55(4):922–938, April 2010.

[108] Angelia Nedic and Asuman Ozdaglar. Distributed Subgradient Methods for Multi-Agent Optimization. *IEEE Transactions on Automatic Control*, 54(1):48–61, January 2009.

[109] G. a. Newman and D. L. Alumbaugh. Three-dimensional massively parallel electromagnetic inversion-I. Theory. *Geophysical Journal International*, 128(2):345–354, February 1997.

[110] Gregory A Newman and David L Alumbaugh. Three-dimensional magnetotelluric inversion using non-linear conjugate gradients. *Geophysical Journal International*, 140(2):410–424, February 2000.

[111] J. Nocedal and S.J. Wright. *Numerical optimization*. Springer series in operations research. Springer, 2006.

[112] B O'Donoghue, Giorgos Stathopoulos, and Stephen Boyd. A splitting method for optimal control. pages 1–20, 2012.

[113] Ardavan Oskooi and Steven G. Johnson. Distinguishing correct from incorrect PML proposals and a corrected unsplit PML for anisotropic, dispersive media. *Journal of Computational Physics*, 230(7):2369–2377, April 2011.

[114] Shadi Oveisgharan and Howard a. Zebker. Estimating Snow Accumulation From InSAR Correlation Observations. *IEEE Transactions on Geoscience and Remote Sensing*, 45(1):10–20, January 2007.

[115] DP Palomar and YC Eldar. *Convex optimization in signal processing and communications*. Cambridge University Press, 2010.

[116] Neal Parikh and S Boyd. Graph Projection Block Splitting for Distributed Optimization. 2012.

[117] W. B. Peter and U. S. Inan. A quantitative comparison of lightning-induced electron precipitation and VLF signal perturbations. *Journal of Geophysical Research*, 112(A12):1–20, December 2007.

[118] Andreas a. Pfaffhuber, Stefan Hendricks, and Yme a. Kvistedal. Progressing from 1D to 2D and 3D near-surface airborne electromagnetic mapping with a multisensor, airborne sea-ice explorer. *Geophysics*, 77(4):WB109, 2012.

[119] V.A. Rakov and M.A. Uman. *Lightning: Physics and Effects*. Cambridge University Press, 2007.

[120] William Rodi and Randall L. Mackie. Nonlinear conjugate gradients algorithm for 2-D magnetotelluric inversion. *Geophysics*, 66(1):174, 2001.

[121] PS Routh and DW Oldenburg. Inversion of controlled-source audio-frequency magnetotelluric data for a horizontally-layered earth. *1996 SEG Annual Meeting*, 64(6):1689–1697, 1996.

[122] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, November 1992.

[123] R. K. Said, U. S. Inan, and K. L. Cummins. Long-range lightning geolocation using a VLF radio atmospheric waveform bank. *Journal of Geophysical Research*, 115(D23):1–19, December 2010.

[124] Piyush Shanker, Francesco Casu, Howard a. Zebker, and Riccardo Lanari. Comparison of Persistent Scatterers and Small Baseline Time-Series InSAR Results: A Case Study of the San Francisco Bay Area. *IEEE Geoscience and Remote Sensing Letters*, 8(4):592–596, July 2011.

[125] Jacob D. Shea, Panagiotis Kosmas, Susan C. Hagness, and Barry D. Van Veen. Three-dimensional microwave imaging of realistic numerical breast phantoms via a multiple-frequency inverse scattering technique. *Medical Physics*, 37(8):4210, 2010.

[126] Wonseok Shin and Shanhui Fan. Choice of the perfectly matched layer boundary condition for frequency-domain Maxwells equations solvers. *Journal of Computational Physics*, 231(8):3406–3431, April 2012.

[127] Weerachai Siripunvaraporn, Gary Egbert, and Makoto Uyeshima. Interpretation of two-dimensional magnetotelluric profile data with three-dimensional inversion: synthetic examples. *Geophysical Journal International*, 160(3):804–814, March 2005.

[128] Weerachai Siripunvaraporn and Weerachai Sarakorn. An efficient data space conjugate gradient Occam's method for three-dimensional magnetotelluric inversion. *Geophysical Journal International*, pages no–no, June 2011.

[129] J. M. Song and W. C. Chew. Multilevel fast-multipole algorithm for solving combined field integral equations of electromagnetic scattering. *Microwave and Optical Technology Letters*, 10(1):14–19, September 1995.

[130] Jean-Luc Starck, Emmanuel J Candès, and David L Donoho. The curvelet transform for image denoising. *IEEE Transactions on Image Processing*, 11(6):670–84, January 2002.

[131] Jean-Luc Starck, Michael Elad, and David L Donoho. Image decomposition via the combination of sparse representations and a variational approach. *IEEE Transactions on Image Processing*, 14(10):1570–82, October 2005.

[132] Jean-Luc Starck, Fionn Murtagh, and Jalal Fadili. *Sparse Image and Signal Processing*. Cambridge University Press, 2010.

[133] M. J. Starks, R. a. Quinn, G. P. Ginet, J. M. Albert, G. S. Sales, B. W. Reinisch, and P. Song. Illumination of the plasmasphere by terrestrial very low frequency transmitters: Model validation. *Journal of Geophysical Research*, 113(A9):A09320, September 2008.

[134] Rita Streich. 3D finite-difference frequency-domain modeling of controlled-source electromagnetic data: direct solution and optimization for high accuracy. *Geophysics*, 74(5):F95, 2009.

[135] Jos F. Sturm. Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones, 1998.

[136] A. Taflove and S.C. Hagness. *Computational Electrodynamics: The Finite-Difference Time-Domain Method*. Artech House Antennas and Propagation Library. Artech House, 2005.

[137] Joel A Tropp and Anna C Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 53(12):4655–4666, December 2007.

[138] Yaakov Tsaig and David L. Donoho. Breakdown of equivalence between the minimal -norm solution and the sparsest solution. *Signal Processing*, 86(3):533–548, March 2006.

[139] G. L. Tyler, I. R. Linscott, M. K. Bird, D. P. Hinson, D. F. Strobel, M. Pätzold, M. E. Summers, and K. Sivaramakrishnan. The New Horizons Radio Science Experiment (REX). *Space Science Reviews*, 140(1-4):217–259, February 2008.

[140] Ashley E. Van Beusekom, Robert L. Parker, Randolph E. Bank, Philip E. Gill, and Steven Constable. The 2-D magnetotelluric inverse problem solved with optimization. *Geophysical Journal International*, 184(2):639–650, February 2011.

[141] Ewout van den Berg and Michael P Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 31(2):890, 2009.

[142] Ewout van den Berg and Michael P Friedlander. Sparse optimization with least-squares constraints. *SIAM Journal on Optimization*, 21(4):1201, 2011.

[143] Bo Wahlberg, Stephen Boyd, Mariette Annergren, and Y Wang. An ADMM algorithm for a class of total variation regularized estimation problems. *arXiv preprint arXiv: ...*, 2012.

[144] Martin J Wainwright. Sharp Thresholds for High-Dimensional and Noisy Sparsity Recovery Using ℓ_1-Constrained Quadratic Programming (Lasso). *IEEE Transactions on Information Theory*, 55(5):2183–2202, May 2009.

[145] PE Wannamaker, JA Stodt, and LE Rijo. Two-dimensional topographic responses in magnetotellurics modeled using finite elements. *Geophysics*, 51(November):2131–2144, 1986.

[146] D.W. Winters, B.D. Van Veen, and S.C. Hagness. A sparsity regularization approach to the electromagnetic inverse scattering problem. *IEEE Transactions on Antennas and Propagation*, 58(1):145–154, January 2010.

[147] Jin Xu, Wei Wang, Jinghuai Gao, and Wenchao Chen. Monochromatic Noise Removal via Sparsity-Enabled Signal Decomposition Method. *IEEE Geoscience and Remote Sensing Letters*, 10(3):533–537, May 2013.

[148] Amer Zakaria, Colin Gilmore, and Joe LoVetri. Finite-element contrast source inversion method for microwave imaging. *Inverse Problems*, 26(11):115010, November 2010.

[149] Amer Zakaria and Joe LoVetri. Application of multiplicative regularization to the finite-element contrast source inversion method. *IEEE Transactions on Antennas and Propagation*, 59(9):3495–3498, September 2011.

[150] Peng Zhao and Bin Yu. On model selection consistency of Lasso. *Journal of Machine Learning Research*, 7:2541–2563, 2007.

[151] M.S. Zhdanov. *Geophysical Inverse Theory and Regularization Problems*. Methods in Geochemistry and Geophysics. Elsevier Science, 2002.

[152] MS Zhdanov, Sheng Fang, and Gabor Hursan. Electromagnetic inversion using quasi-linear approximation. *Geophysics*, 65(5):1501–1513, 2000.