Improving potential-field processing and inversion through the use of gradient measurements and data-adaptive models

Kristofer Davis
Improving potential-field processing and inversion through the use of gradient measurements and data-adaptive models

Kristofer ‘The G.U.’ Davis

Defended: September 01, 2010

Advisor: Prof. Yaoguo Li (GP)
Committee Chair: Prof. Donald Macalady (CHGN/LAIS)
Minor: Prof. Luis Tenorio (MATH)
Committee Members: Prof. Misac Nabighian (GP)
Prof. Dave Hale (GP)

Department of Geophysics
Colorado School of Mines
Golden, CO 80401

http://www.geophysics.mines.edu/cgem
IMPROVING POTENTIAL-FIELD PROCESSING AND INVERSION THROUGH THE USE OF GRADIENT MEASUREMENTS AND DATA-ADAPTIVE MODELS

by

Kristofer Davis
A thesis submitted to the Faculty and the Board of Trustees of the Colorado School of Mines in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Geophysics).

Golden, Colorado
Date _____________________

Signed: _______________________________  On original
Kristofer Davis

Golden, Colorado
Date _____________________

Signed: _______________________________  On original
Dr. Yaoguo Li
Thesis Advisor

Golden, Colorado
Date _____________________

Signed: _______________________________  On original
Dr. Terence K. Young
Professor and Head
Department of Geophysics
ABSTRACT

This thesis focuses on two distinct, yet complementary topics. First, I examine the use of observed gradient and total-field data acquired from aeromagnetic surveys as independent data sets for quantitative interpretation. Second, I develop a more efficient means of processing and inverting these large data sets.

The first topic stems from the recent use of multiple magnetometers, which allows for measurements of magnetic gradients and minimizes diurnal variation and other common-mode noise. I develop equivalent source and 3D inversion algorithms for joint processing and inversion of total-field and gradient data, respectively. These approaches treat the observed gradients as an additional and independent data set rather than just as supplemental information. The direct result of joint processing is a set of enhanced data with increased information content and a higher signal-to-noise ratio. The method diminishes high-frequency noise and accentuates mid-frequency signals and produces a data set with more resolving power than total-field data alone. Likewise, the contribution of gradients in inversion leads to a better resolved 3D susceptibility distribution.

The second topic addresses the fact that equivalent source processing and 3D inversion in potential-field methods involve large and dense coefficient matrices that often exceed the limit of physical memory in commonly available computers. Repeated matrix-vector multiplications during processing or inversion require an immense amount of CPU power. This poses a significant challenge when solving large-scale problems and can render many problems intractable. To overcome these limitations, I have developed a new computational approach for this class of problems by combining an adaptive quadtree or octree model discretization and wavelet transforms on a re-ordered parameter set. To maintain resolution, the adaptive mesh discretizes the model region by starting with large cells and splitting the region into smaller cells
for localized anomalies. Then, Hilbert space-filling curves and similar ordering of the reduced parameter set dramatically improves the compression of the coefficient matrix using wavelet transforms. This combination can reduce the storage requirement, and likewise increase solution speed, by up to three orders of magnitude. As a result, problems that were previously computationally prohibitive can now be solved.
# TABLE OF CONTENTS

ABSTRACT ........................................................................................................ iii

LIST OF FIGURES ............................................................................................... vii

LIST OF TABLES ................................................................................................ xiv

ACKNOWLEDGMENTS ....................................................................................... xv

CHAPTER 1 INTRODUCTION ............................................................................. 1

1.1 Use of Gradients in Processing and Inversion ............................... 2
1.2 Efficiency of Inversion and Inversion-based Processing ............ 4
1.3 Overview of Thesis ................................................................................. 5

CHAPTER 2 INVERSION METHODOLOGY .................................................. 7

2.1 Linear Problem .................................................................................... 7
2.2 Inversion Algorithm ......................................................................... 8
   2.2.1 Global Objective Function ......................................................... 8
   2.2.2 Data Misfit Function ................................................................. 9
   2.2.3 Model Objective Function ....................................................... 9
   2.2.4 Choosing the Tikhonov Parameter .......................................... 11
2.3 Practical Aspects for 3D Inversion .................................................... 14
   2.3.1 Positivity Constraint ................................................................. 14
   2.3.2 Depth Weighting .................................................................... 15
2.4 Summary ............................................................................................. 15

CHAPTER 3 USE OF GRADIENTS IN PROCESSING ............................... 16

3.1 Equivalent Source Technique ............................................................ 17
3.2 A Common Equivalent Source for Total-field and Gradient Data ... 17
3.3 The Translation Matrix ...................................................................... 21
3.4 Synthetic Example ............................................................................. 23
3.5 Field Example .................................................................................... 26
3.6 Summary ............................................................................................. 32

CHAPTER 4 USE OF GRADIENTS IN 3D INVERSION ............................... 35

4.1 A Common 3D Susceptibility Model ............................................... 35
4.2 Synthetic Examples ........................................................................... 38
4.3 Field Example .................................................................................... 42
LIST OF FIGURES

2.1 An example of an L-curve, created by plotting the data misfit versus the model objective function. For solutions with too large a Tikhonov parameter, the model objective function is small because of an overly-smooth model. Conversely, insufficient regularization causes data to over fit the observed data. This results in a geologically unreasonable model. The optimal regularization parameter is found on the ‘elbow’ of the curve where the maximum curvature exists. . . . . . . . . . . . 12

2.2 Recovered models based on three regularization parameters. The top left is the recovered model using too large a regularization parameter. The top right is the recovered model if the parameter is too small. The model with the parameter chosen on the elbow of the L-curve is shown in the bottom left. It has a better geologic structure than the other two, yet still fits the data. The true model is shown in the lower right for comparison. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13

3.1 A schematic of two cells with different susceptibilities and a thin sheet with a width of $\Delta x$. This schematic illustrates the foundation for the use of gradients and total-field data in a common susceptibility model. 20

3.2 The gradient response (upper left panel) of a semi-infinite horizontal dike (lower left panel). The dike has a susceptibility of 0.5 SI surrounded with a background susceptibility of zero. The total-field anomaly (upper right panel) is generated by a thin sheet (lower right panel) with a width of 1 m and susceptibility of -0.5. The thin sheet is located where the semi-infinite dike ended and represents the change between the two susceptibilities ($\kappa_1 - \kappa_2$). The total-field anomaly of the thin sheet is equal to the gradient produced by the semi-infinite dike. 21

3.3 Synthetic "observed" data with noise (1 nT standard deviation). The observation locations are shown by the white dots. This data set is used to compare the joint and total-field equivalent source methodologies. . 24
3.4 The evenly gridded data set calculated by a layer of equivalent sources from total-field data is shown in the upper left. The joint method data set is shown in the upper right. The corresponding differences from the true gridded response and responses from the equivalent source technique (lower right) and equivalent source with gradients (lower left) are shown. The difference maps are on the same scale. The joint methodology had half the errors and less striping than total-field equivalent sources.

3.5 This field example is a subset of data taken from the Central Abitibi Destor-Porcupine-Pipestone Faults area, near Matheson, Ontario. Nominal line spacing is 75 m and along-line observations were taken approximately every 15 m. The data locations are shown as white dots.

3.6 The power spectrum of the total-field and gradient of the observed field data is shown. There is a noticeable shift in power and the gradient has less noise than the total-field as expected. The difference in power will give the ratio of noise level between the two datasets in order to assign a consistent error level between the gradient and total-field data.

3.7 The calculated data sets that were produced by the joint (top) and total-field (bottom) equivalent source techniques at even intervals of 20 m.

3.8 The equivalent source layers produced with (top) and without (bottom) gradients. Both methods used the same model objective function. The model calculated without gradients has a wider and more connected anomaly than the joint technique.

3.9 The differences between predicted data from the joint equivalent source and the normal equivalent source (top), and the minimum curvature gridding (middle). The differences for most of the data maps are within 5 nT, but the edges of the anomaly differ by up to 50 nT for both the minimum curvature and total-field techniques. Also shown is the difference between the normal equivalent source and minimum curvature (bottom). Striping in the normal equivalent source technique is apparent in the two difference maps when compared to the joint methodology and minimum curvature.

3.10 The power spectra of the observed data (dots), data created by total-field equivalent sources (black line), and data modeled by the joint total-field and gradient equivalent sources (crosses). The use of gradients increases power of the signal in mid-frequencies. Additionally, high frequency noise is suppressed greater than in the total-field equivalent source method or minimum curvature gridded data.
3.11 The difference (dB) in power spectra from Figure 3.10 with respect to the total-field and gradient equivalent source data sets. The abscissa is wavelength in meters. The joint gradient and total-field method suppresses high frequency noise in the data by more than 40 dB compared to the minimum curvature and 15 dB with total-field equivalent sources. Additionally, the method contains more mid-frequency signal by 15 dB, illustrating the increase in resolution obtained. These wavelengths include the main anomaly of the data set.

4.1 Synthetic data of total-field (upper left), vertical gradient (upper right), \( x \)-gradient (lower left), and \( y \)-gradient (lower right). Five percent Gaussian noise was added to each dataset with a minimum of 3 nT for the total-field data and 0.002 nT/m for the gradient components. The Cartesian coordinate system has its origin at the Earth’s surface, \( x \)-axis points towards grid north, \( y \)-axis points towards grid east, and \( z \)-axis points vertically downward.

4.2 The true (left column) and recovered models using total-field data (middle column) and total-field with three gradients (right column). The shapes of the recovered models are similar, but the joint method recovers higher susceptibility values that are closer to the true model.

4.3 A cross-section half way through (600 m Easting) the true model that was used to generate data. The susceptibility is 0.05 SI units.

4.4 Synthetic data of total-field (upper left), vertical gradient (upper right), \( x \)-gradient (lower left), and \( y \)-gradient (lower right) for a dipping slab. Five percent Gaussian noise was added to each dataset. The inducing field strength is 50,000 nT with an inclination of 65° and declination of 25°.

4.5 A cross-section half way through the recovered model using total-field only (top). A cross-section through the recovered model created by the joint inversion using the total-field and three orthogonal derivatives (bottom). The cross-section shows the same model region as Figure 4.3.

4.6 Observed total-field data (upper left), northing gradient (lower left), and easting gradient (lower right) from the mining industry used for 3D inversion. The inclination is 79° and in the declination is 12°. An inversion of only the total-field is performed for comparison. The two horizontal gradients will aid in constraining the model laterally via joint inversion.

4.7 Recovered model using only total-field data with susceptibilities higher than 0.028 SI units shown. Due to smoothing parameters, it could be interpreted as a single body. Conversely, the pinch in the middle may also be a fault zone; the main anomaly is composed of two main source bodies.
4.8 Recovered model using the total-field and two horizontal gradients. Susceptibilities higher than 0.028 SI units are shown. The middle of the main anomaly is disconnected where the proposed fault zone is. The northern portion of the anomaly shows multiple compact bodies exemplifying the influence of the gradients. The main anomaly has more structure and is higher in susceptibility than that of the recovered model from the total-field data.

4.9 A slice in depth of the recovered models from total-field only (left) and from total-field and two horizontal gradients (right). The recovered anomalies from the joint inversion have higher susceptibility and are more compact with limited effect of smoothing. The north anomaly shows that there may be three smaller sources close together rather than one elongated feature.

5.1 An example of the curvature of amplitude normalized by amplitude (left) and the corresponding quadtree discretization (right). Small prismatic cells located at the local anomaly allow for increased resolution. The result is the re-production of data from equivalent sources based on a smaller number of model parameters.

5.2 A continuous way of ordering cells from east to west for a normal mesh (left panel), and then moving north versus moving radially outward so each cell in re-ordering is adjacent to the previous (right panel). By creating adjacent indexing, the discontinuities in each row of the sensitivity matrix are minimized allowing for better compression through the 1D wavelet transform.

5.3 The Hilbert space-filling curve is used for ordering model parameters with the 1D wavelet. The first order curve (left) for a $2^1$ square. The second order curve (middle) is for $2^2$ square. Also, the indexing scheme for the same mesh (right) as shown in Figure 5.2 using the third order. The Hilbert curve allows for model cells to be spatially continuous during the adaptive mesh parameter ordering.

5.4 An example of spatial mapping of the $8 \times 8$ mesh example after quadtree discretization. The continuous coordinate mapping is shown in the left panel. The middle panel shows the result of the continuous radius mapping. The Hilbert curve (right panel) outperforms both by ordering every spatially adjacent cell. The upside-down $U$ shape exemplifies the regional movement of the first-order Hilbert curve (left panel of Figure 5.3). The arrows show spatial discontinuities.

5.5 The quadtree mesh used for the field example and comparing ordering schemes. The mesh is based on a rectangle of $244 \times 174$ model parameters and consists of approximately 11,000 prisms.
5.6 A randomly chosen row of $G$ for the square-based quadtree mesh discretization with 16,000 parameters. The continuous indexing (top) and radially continuous (middle) are outperformed by the Hilbert space-filling curve (bottom) because of the scattered high amplitude features. The Hilbert curve consists of only one area of high amplitude. The remaining coefficients in the row can be reconstructed through wavelets with very few significant coefficients.

5.7 Inset of the cluster of high amplitude features from the Hilbert curve as shown in Figure 5.6. The sparseness of the parameters follows from the spatial clustering and continuity produced by the space-filling curve. This feature allows the reconstruction of the row using only a few significant wavelet coefficients, thereby increases the compression ratio by threefold compared to the other two methods.

5.8 A random row of $G$ from a rectangular-based quadtree mesh with 11,000 parameters. The continuous (top) and radially continuous (middle) yield similar results. Like the square-based mesh, the Hilbert curve (bottom) outperforms the two other methods and only has two main groups of peaks.

5.9 The observed data for equivalent source processing from the Central Abitibi Destor-Porcupine-Pipestone Faults area (Ontario Geologic Survey, 2004). The indexing schemes were compared with the 6,145 data. With the quadtree mesh, Hilbert curve, and 1D wavelet transform, the coefficient matrix was compressed 1227 times and had similar data and model results after processing.

5.10 The recovered susceptibility model using a normal (top) and quadtree mesh (bottom).

6.1 An example of a 2.5D quadtree mesh shown in plan view (Top). A slice of the same mesh cut through the main anomalous zone in depth (middle). The cells retain the horizontal size assigned from the top view. The bottom panel is a slice when the depth weighting function is incorporated into the discretization to further reduce the number of cells vertically.

6.2 An example of mesh discretization through the octree process. The upper left panel shows all of the cells that are of minimum cell size (i.e. the original cell widths). The majority of these cells are located in the top of the inverted anomalies to the surface of the mesh. The upper right panel shows locations of all cells grouped into $2 \times 2 \times 2$ the smallest cell size. The bottom left and right panels show groups of $4 \times 4 \times 4$ and $8^3$ number of smallest cells, respectively. The top-south cube of the grouped cell locations are shown. The resulting mesh is 29,499 cells rather than 262,144 cells.
6.3 The Hilbert space-filling curve is used to order model parameters for the 1D wavelet. The first order curve (top) for a $2^1$ cube. The second order curve (bottom) is for $2^2$ cube. The Hilbert curve allows for model cells to be spatially continuous during the adaptive mesh parameter ordering. The ordering starts in the south-west top corner and concludes in the north-west top corner.

6.4 Example of two different model ordering schemes for the same row of a sensitivity matrix. The row has already been adaptively discretized to 68,286 parameters. The top panel is the model response after normal indexing. The inset shows the scattering of the high-amplitude clusters over 2,000 parameters. The bottom panel is the ordering after 3D Hilbert space-filling curves. Two main clusters are present with the inset showing the first cluster. The row was compressed by the 1D wavelet transform and resulted in compression ratios of 76 and 118 for the continuous and Hilbert curve ordering, respectively.

6.5 Example data set of magnetic total-field anomaly acquired for mineral exploration. The inclination is $79^\circ$ and in the declination is $12^\circ$.

6.6 The recovered model through inversion using a regular mesh. Two main elongated features are associated with the large anomaly and three smaller sources are also present. The southern small anomaly is shown to be connected with the southern larger anomaly, though it could be an artifact of model smoothing.

6.7 The curvature of the amplitude normalized by the magnetic amplitude as used with the quadtree is shown in the upper left. Due to the similarities of the amplitudes of the features, a new criterion must be selected. The curvature of the total-field is shown in the upper right. The total-horizontal gradient (THG) of the total-field is also considered (lower left). The final choice of the curvature of the amplitude is used (lower right) without normalization.

6.8 The recovered model through inversion using an octree-based mesh. Two main elongated features are associated with the large anomaly and three smaller sources are also present. The southern small anomaly is shown not to be connected to the southern larger anomaly, unlike the smoother recovered model from the regular mesh solution. The octree mesh design solves for 29,499 parameters compared to the regular mesh with 262,144 parameters.
6.9 A slice in depth of the recovered model through inversion using a normal (left) and octree-based mesh (right). Two main elongated features are associated with the large anomaly and three smaller sources are also present. The southern small anomaly is shown not to be connected to the southern larger anomaly, unlike the smoother recovered model from the regular mesh solution. The susceptibilities are similar using both methods.
### LIST OF TABLES

3.1 The L-2 and L-infinite norms of the data sets created by the joint and total-field equivalent source models as compared to the true forward model show the joint algorithm has 65% less error than using total-field data alone. .................................................. 26

4.1 A comparison of the recovered models with respect to the true models for the joint and total-field alone inversions for the two synthetic examples. The joint inversion recovered more of the true model than that of the total-field inversion in each case. ............................ 43

5.1 Compression based on model indexing and the 1D wavelet transform. The smallest wavelet coefficients are discarded to obtain a 95% accuracy. 62

6.1 A comparison of practical aspects of storage and computational efficiency for the field example. The octree-based mesh compresses the sensitivity matrix by three fold, requiring less storage and solving the inverse problem in under a quarter of the time. The sensitivity matrix for the full problem without compression would consist of 278,914,122 entries. Though the example is a smaller problem, it exemplifies the utility of the method. The difference between these two techniques will increase as the size of the problem increases. ....................... 80
ACKNOWLEDGMENTS

I would like to thank my committee for their guidance, support, and instruction. They are: Donald Macalady, Luis Tenorio, Dave Hale, Misac Nabighian, and Yaoguo Li. I thank Dave Hale for excellent and insightful discussions on the Hilbert curve. I would like to acknowledge the members of CGEM for all the advice, commentary, and criticism, and of course those afternoons listening to live music with beers at the Buffalo Rose. I am specifically grateful to Barry Kirkendall for the no bueno stamp and teaching me the art of the check mark and Rich Krahenbuhl for all his comments, editing, and New Mexican adventures. I owe a deep debt of gratitude to my advisor, Yaoguo Li. The hours of discussion spurred from a single question have been highly beneficial to not only my thesis, but my understanding of the science. He has significantly helped me throughout the entire process.

I owe a great deal to my family and friends. For the last three years, my caring mother has been there asking me when I will finish and if I have added “bubbles” to my thesis. My father has always been right there behind her telling me that no matter what, there will always be Yuengling when I get home. I also thank my former roommates Andy Kass and Matty Wisniewski for some of the most productive after-work discussions and my wife, Jessica, who exhibited infinite patience with us. She has given me an amazing amount of support and has been absolutely wonderful. I would like to acknowledge the hard work of Karen DeShong; without her I would have never been organized enough to get where I am today. I thank the my co-workers with the Colorado Rockies: Brady O’Neill and the entire promotion staff. Walking into a Major League Baseball ballpark and getting paid for it after working on research all day is exhilarating. Furthermore, assisting a large, purple triceratops and having to deal with angry, sometimes profanity-speaking, parents (next to their kids no less), who did not get in line in time to see said dinosaur always places a lighter perspective
on life. Seriously people, chill out.

Finally, I would like to thank my primary sources of financial support over the last several years. My primary support has been through various corporate and government sponsors of the Center for Gravity, Electrical, and Magnetic Studies (CGEM). Partial support came from the Department of Geophysics at the Colorado School of Mines for assistantships as well as the Society of Exploration Geophysicists (SEG). After four years of out-of-state tuition in the undergraduate program, these organizations were instrumental in keeping me indebted to Nelnet for only a couple more decades rather than the rest of my life.
CHAPTER 1
INTRODUCTION

Magnetic surveys are used in a variety of exploration settings. For mineral exploration, deposits associated with magnetite (such as Kimerlite pipes) can be directly detected (e.g. Koulomzine and Brossard, 1947). Surveys are used to map the basement structure (Vacquier et al., 1951) in oil exploration, volcanology (Byers et al., 1946), or even intrasedimentary magnetic deposits associated with the upward migration of hydrocarbons. The magnetic method is also used in archaeological (Becker, 1995; Kvamme, 2005) and hydrothermal applications (Clark, 1999; Nabighian et al., 2005).

There are three types of magnetic survey design: ground, airborne, and marine. It is most common to acquire total-field and gradient data in the first two. To acquire gradient data in airborne surveys, total-field sensors are placed on each wing tip and/or on the front and back of the aircraft. The change in the magnetic field between these sensors is then observed. Aeromagnetic surveys are advantageous since they can acquire accurate data over larger areas (up to thousands of square kilometers) and can be performed using either fixed-wing or rotary aircraft. In general, fixed-wing aircraft are less expensive, faster, and produce higher-quality data (Dobrin and Savit, 1988).

Use of magnetic data requires the removal of time variations associated with diurnal variations in the Earth’s magnetic field during acquisition. One technique is to record the Earth’s magnetic field with a stationary ground magnetometer and simply subtract these changes from the measured magnetic data. Although this practice has been used for decades and is still being used in ground surveys, in airborne surveys the true diurnal variation at the ground base station will differ from the airplane sensors if the conductivity of the earth is non-constant throughout the survey area (Reynolds,
A method that has become more common in order to remove these effects is called leveling. Traditionally, leveling is performed by flying extra lines (tie lines) and observing and removing the difference in the observed magnetic field at points in space between survey and tie lines. Difficulties of this approach is that errors in leveling are introduced by failing to occupy the same locations (Minty, 1991), as well as by poorly approximating the magnetic response of the airplane (Nelson, 1994). In addition, these tie lines increase survey cost.

1.1 Use of Gradients in Processing and Inversion

Hood (1965) first introduced the use of gradients specifically for magnetic exploration because diurnal variation is common among multiple sensors. Currently, the use of 2 to 4 magnetometers in airborne magnetic surveys is becoming more frequent, allowing observation of the total field as well as its horizontal gradients (Reford, 2006). The results is a vast reduction in the need for tie lines and thus more economical data acquisition (Hardwick, 1999). Gradients observed in ground surveys can also be used in this fashion.

The processing of aeromagnetic data has developed into the use of gradients for enhanced gridding. Processing techniques such as applying Hilbert transforms to observed horizontal gradients (Hansen, 1984; Nabighian, 1984) may then be applied. These transforms calculate the vertical gradient that then enables the reproduction of total-field data. Expanding upon Hilbert transforms, the Fourier transform is used in order to derive the total field from the leveled and gridded gradients (Nelson, 1994). Hardwick and Boustead (2007) created pseudo-lines parallel to observed total-field data. The gradient can accurately calculate the total-field at a distance away from, and parallel to, the flight-line on both sides. This method traditionally uses minimum curvature gridding, and it has the advantage of utilizing three lines of data versus one to increase the accuracy of the gridded results. A similar approach using the gradient to enhance gridding has also been performed by Reford (2006). Rather than
calculating three pseudo-lines of data, he uses the gradients to constrain how the reproduced field changes with position.

Though little has been done in the realm of joint gradient and total-field magnetic inversion, gradients of potential fields have been popular in quantitative interpretation (a subject of which includes inversion). Calculated gradients have primarily been used to assist in the problem of ambiguity in potential field theory and application. Papers originally emphasized the advantage of gravity gradients for structural geology interpretation (e.g. Evjen, 1936; Heiland, 1943). Skeels (1947) used gradients to estimate the maximum depth to a gravity or magnetic target. Gradients are used directly with 2D inversion (Butler, 1995), automated processing (Nabighian, 1972; Butler, 1984), and automated interpretation such as Euler (Mushayandebvu et al., 2001; Nabighian and Hansen, 2001) or Werner (Hartman et al., 1971) deconvolution.

Rather than incorporating the gradients as simply a mathematical tool for or supplement to gridding, I take advantage this extra information on its own as a data set which can be interpreted in conjunction with total-field data. Rather than using the gradients as a supplement to gridding, I take advantage of this extra information as a stand alone dataset which can be interpreted in conjunction with total-field data. To do this, the equivalent source processing technique (Dampney, 1969) is applied to jointly process both datasets. This provides a high quality representation of the observed magnetic anomaly by providing de-noised data that contain enhanced higher frequency signals. The same methodology is extended to 3D inversion. For inversion, the focus is the recovery of a geologically reasonable model that fits the observed data without reproducing the noise. The use of total-field and gradient data as two distinct data sets for the response of a single model provides a high quality quantitative interpretation.
1.2 Efficiency of Inversion and Inversion-based Processing

Solving a system (such as with equivalent source processing or 3D inversion) with a dense matrix of size $m \times m$ requires at least $O(m^3)$ operations. For large data sets, this can become computationally too expensive. Many techniques have been introduced to improve efficiency in order to handle these large data sets. Li and Oldenburg (2010, 2003) use a multi-dimensional wavelet transform to represent this dense matrix sparsely. The discrete Fourier transform (DFT) has also been incorporated in inversion (Mareschal, 1985; Cordell, 1992; Pilkington, 1997) in order to speed up the process. The DFT approach is quite fast, but has its own drawbacks such as requiring even gridding of data on a horizontally flat plane. Both the wavelet and Fourier transforms focus on using basis functions in order to store only the coefficients of those functions to reproduce the dense matrix. Portniaguine and Zhdanov (2002) dissected the dense matrix into multiple smaller sparse matrices that can be multiplied to reproduce the original matrix. In practice only a portion of the data set is inverted, or one uses larger model cells than should be used. This type of user-defined compression has its obvious drawbacks. Compression is defined as the ratio of the number coefficients in the full matrix to the number of coefficients of the sparse matrix.

To improve the efficiency of processing and inversion, I use an adaptive mesh design (Frey and Marechal, 1998) where cell size and distribution within the layer or 3D model is data adaptive and model parameters are automatically made coarse where no significant signal is present and fine in areas of interest prior to inversion. The net result is a dramatic decrease in the total number of cells within the model that need to be calculated and therefore a decrease in overall cost of inversion. In addition, I order the model cells by using the Hilbert space-filling curve algorithm (Hilbert, 1891) to form a smooth one-dimensional sequence of rows of the sensitivity matrix and apply the 1D wavelet transform. Through the combination of the 1D wavelet transform and adaptive mesh architecture, I reduce the overall size of the
system of equations by at least three orders of magnitude.

1.3 Overview of Thesis

The thesis contains seven chapters. The first chapter is an overview of the challenges faced with aeromagnetics and the current methodologies used to combat them. I introduce the concept of gradients as independent data to develop a new way of looking at gradients. I then discuss the use of data-adaptive models for both processing and inversion of aeromagnetic data using quadtree and octree discretization, respectively.

I develop the theoretical aspects of equivalent source processing and 3D inversion in Chapter 2, starting with the description of the linear problem for magnetics. I then discuss the solution of the inverse problem to obtain a geologically reasonable model for which the response fits the observed data within the noise. This chapter provides the background and foundation for solving the type of inverse problem discussed in the remaining chapters.

Chapter 3 presents joint processing of gradient and total-field data. I introduce a way to solve for a common equivalent source layer for both types of data. I make use of a well-known but rarely used relationship between the derivatives of the magnetic field and the derivative of its source to relate both data sets to a common equivalent source distribution. I show an increase of resolution in the data with a synthetic and a field example.

I then build upon the processing methodology and extend it to 3D inversion in Chapter 4. If the vertical gradient is acquired, it can be used to constrain the three-dimensional model in depth. This component adds to the set of equations that were developed in Chapter 3. A synthetic and a field example exemplifies the utility of the algorithm.

The amount of data in processing can become quite large. I discuss ways of alleviating this problem in Chapter 5. I introduce using an adaptive quadtree dis-
cretization on a layer of prisms for equivalent source processing, fining the cells in regions of large data change and coarsening where no significant change is present. I also discuss using wavelet transforms on an irregularly gridded mesh and how to efficiently order the model parameters using space filling curves (e.g. Peano, 1890; Hilbert, 1891).

In Chapter 6, I expand upon the idea of a data-adaptive model for use in 3D inversion. I introduce an octree mesh discretization approach. Methodologies similar to the efficient processing techniques are presented. I examine an optimal ordering scheme of the model parameters in three dimensions to increase the effectiveness of the wavelet compression.

Finally, I conclude with a summary of the advantages and disadvantages to the research presented in this thesis. I also give recommendations for potential future research directions, particularly in the advancement of speed and storage algorithms.
CHAPTER 2
INVERSION METHODOLOGY

Geophysical data give insights to the subsurface. They contain information about changes in physical properties, such as magnetic susceptibility, caused by mineralization, faulting, or human activities. Inversion is a quantitative interpretation tool to model the sources that generate the observed data. To have a successful interpretation, the recovered model should be geologically reasonable and the data created by the model should fit the observed data within the error. I utilize Tikhonov regularization to enable such criteria to be achieved.

In this chapter, I present the theoretical base of solving for magnetic source distributions. The framework is similar for both equivalent source processing and 3D inversion. I start with the general linear problem, which will be expanded in the following chapters to include gradients or a data-adaptive mesh. I then present the formulation of the inverse problem and the solution strategy.

2.1 Linear Problem

All aspects of the proposed research involve linear inversion techniques, whether solving for a 2D layer of sources for processing or a 3D distribution for interpretation. The linear problem is defined by

\[
G \vec{\kappa} = \vec{d},
\]  

(2.1)

where \( \vec{\kappa} = (\kappa_1, \kappa_2, \ldots, \kappa_m)^T \) and are the magnetic susceptibilities of the model, \( \vec{d} = (d_1, d_2, \ldots, d_n)^T \) are the data, and \( G \) is a dense coefficient matrix, referred to as the sensitivity matrix. It describes the geometry and physics between the sources (prismatic cells) and the points of observation (Li and Oldenburg, 1996; Blakely,
1996). The sensitivity for the $i^{th}$ datum given the $j^{th}$ model susceptibility is

\[
G_{i,j} = \frac{1}{4\pi} \int_{V_j} \hat{B} \cdot \nabla \nabla \frac{1}{|\vec{r}_i - \vec{r}|} \cdot \vec{B}_o \, dV'.
\]

(2.2)

where $\hat{B}$ is direction of the inducing field, $V$ is the source volume located at location $\vec{r}'$. The observation location is denoted by $\vec{r}_i$. This dense matrix is a function of spatial variables. The elements along the columns correspond to different observation locations ($x, y, z$), while the elements along the rows correspond to different source locations ($x', y', z'$). Thus the dense sensitivity matrix is $m$ model parameters by $n$ data and holds coefficients of the kernel function for each datum and respective model parameter. Once the linear system of equations is set, one can solve for susceptibility.

### 2.2 Inversion Algorithm

With the forward model, it is now possible to formulate the linear inverse problem to construct magnetic sources. I construct the magnetic susceptibility distribution through Tikhonov formalism (Tikhonov and Arsenin, 1977). This is accomplished by minimizing a global objective function subject to the data constraints in Equation 2.1 (Menke, 1989; Parker, 1994). The method consists of three main components: the global objective function, the data misfit function, and the model objective function. In the next three sections, I will describe each component.

#### 2.2.1 Global Objective Function

The global objective function is a product the model objective and data misfit functions (Tikhonov and Arsenin, 1977). The optimal model solution is found by minimizing the global objective function, $\Phi$, such that

\[
\min \Phi = \Phi_d + \beta \Phi_m
\]

subject to \( \Phi_d = \Phi_d^* \),

(2.3)
where $\beta$ is the regularization parameter, $\Phi_d$ is the data misfit function, and $\Phi_m$ is the model objective function. The global objective function is minimized subject to the data misfit equaling the target data misfit, $\Phi^*_d$ which depends upon the noise in the data. The data misfit function defines the fit of the predicted data to the observed data. The model objective function measures complexity of the model. The global objective function is minimized to ensure a smooth model that fits data using one of a number of numerical optimization techniques such as conjugate gradients (Hestenes and Stiefel, 1952; Nocedal and Wright, 1999).

### 2.2.2 Data Misfit Function

The data misfit function defines how well the predicted data reproduces the observed data within each datum’s error. It is defined as

$$\Phi_d = \left\| W_d (G\vec{\kappa} - \vec{d}) \right\|^2_2,$$

where $W_d$ is a diagonal data weighting matrix normalizing the data by its respective standard deviation, $\sigma_i$. It is common to use an L-2 norm measure of data fit. The expected data misfit used is the number of data. This assumes the noise follows a random Gaussian distribution. If the data misfit is smaller than the expected value, the model has been constructed to fit noise. If the data misfit larger than the expected value, not enough of the signal has been reproduced for a desirable result.

### 2.2.3 Model Objective Function

The model objective function quantifies the structural complexity of the model through volume-based derivatives of the model. For a 3D discretized magnetic susceptibility distribution, the model weighting matrix consists of a summation of the following components: a smallest model component, $x-$, $y-$, and $z-$ directional weights of the model, and a depth weighting function. The use of the derivatives is shown...
when the model objective function is in integral form as

\[ \Phi_m = \alpha_s \int_V (w(\vec{r})(\kappa - \kappa_o))^2 dV + \alpha_x \int_V \left( \frac{\partial w(\vec{r})(\kappa - \kappa_o)}{\partial x} \right)^2 dV + \]
\[ \alpha_y \int_V \left( \frac{\partial w(\vec{r})(\kappa - \kappa_o)}{\partial y} \right)^2 dV + \alpha_z \int_V \left( \frac{\partial w(\vec{r})(\kappa - \kappa_o)}{\partial z} \right)^2 dV, \] (2.5)

where \( w(\vec{r}) \) is a depth weighting function which will be discussed further in section 2.3.2 and \( \kappa_o \) is a reference model. Changing \( \alpha \) for any direction places more or less weight of the directional skewness of the model. By changing \( \alpha \) and incorporating a reference model, the function is a vehicle to incorporate prior information about the model region from other geophysical surveys or geology. The smallest model component drives the model towards the reference model. The matrix-vector form of Equation 2.5 is:

\[ \Phi_m = (\vec{\kappa} - \vec{\kappa}_o)^T(W_s^T W_s + W_x^T W_x + W_y^T W_y + W_z^T W_z)(\vec{\kappa} - \vec{\kappa}_o). \] (2.6)

For equivalent sources, \( \Phi_m \) has the same formulation, but uses only the \( x- \) and \( y- \) directional weighting functions (e.g. \( W_x \) and \( W_y \)). It is advantageous to have a zero reference model and no smallest model component. Depth weighting is also unnecessary because there is a single layer of prisms for processing. Incorporating the flattest model minimizes change between survey lines. The model objective function for inversion-based processing then becomes

\[ \Phi_m = \alpha_x \int_V \left( \frac{\partial(\kappa - \kappa_o)}{\partial x} \right)^2 dV + \alpha_y \int_V \left( \frac{\partial(\kappa - \kappa_o)}{\partial y} \right)^2 dV, \] (2.7)

and in short form it can be expressed by

\[ \Phi_m = \| W_m(\vec{\kappa} - \vec{\kappa}_o) \|^2, \] (2.8)
with the reference model, $\vec{r}_o$ and the model weighting matrix, $W_m$. In practice, $W_m$ is formed prior to the minimization process and stored sparsely. This allows for a sparse matrix-vector multiplication and summation of the resulting vector to calculate the value of $\Phi_m$. Now that there are the data misfit and model objective functions, Equation 2.3 can be easily formed. It is important to choose a proper Tikhonov parameter to trade-off between the model complexity and the data fit. This is the topic of the next section.

### 2.2.4 Choosing the Tikhonov Parameter

The goal of the formulation is to obtain a data misfit close to a target data misfit, $\Phi_d^*$. It is possible to define the target misfit when the statistics of error in data are known. The minimization process is completed when the value of $\beta$ is found that yields $\Phi_d^*$. More often, however, data errors are unknown and other approaches such as the L-curve criterion (Hansen, 2000) or generalized cross validation (GCV) (Golub et al., 1979) are used to find the optimal $\beta$ value to achieve balance between model complexity and the fit of the data within the assumed errors.

#### L-curve Criterion

I will first examine L-curve criterion to perform a search for an appropriate Tikhonov parameter. Inversions are performed using a range of values of $\beta$. The result of plotting the data misfit (Equation 2.4) versus the model objective function (Equation 2.6) is a curve that is shaped like an ‘L’ (Figure 2.1). For solutions with large Tikhonov parameters, the model objective function is small because of the smooth model created though minimization has more emphasis on $\Phi_m$; but the data misfit is quite large because of the lack of data fit. Conversely, for small Tikhonov regularization parameters the generated data fits the observations too well, also incorporating the noise, and the model may be geologically unreasonable. Between these two extremes, there is a transition at which an appropriate regularization parameter
can be chosen. This area is the corner, often referred to as the ‘elbow’ and is defined as the maximum curvature of the L-curve. This point is considered to be the optimal trade-off between data misfit and model complexity. To illustrate these extrema, Fig-

Figure 2.1. An example of an L-curve, created by plotting the data misfit versus the model objective function. For solutions with too large a Tikhonov parameter, the model objective function is small because of an overly-smooth model. Conversely, insufficient regularization causes data to over fit the observed data. This results in a geologically unreasonable model. The optimal regularization parameter is found on the ‘elbow’ of the curve where the maximum curvature exists.

ure 2.2 shows recovered models of a block with too large a regularization parameter (upper left), too small a parameter (upper right), and the optimum parameter (lower left). For reference, the true model is shown in the lower right. It is noticeable that with the large regularization parameter the model is smooth and does not fit the data. The recovered model with the small regularization parameter has many low magnetic susceptibility blocks on the surface to be able to over-fit the noise. The optimum Tikhonov parameter recovers a model which is smooth, but is also able to fit the data within the wavelengths of the signal and is geologically reasonable.
Figure 2.2. Recovered models based on three regularization parameters. The top left is the recovered model using too large a regularization parameter. The top right is the recovered model if the parameter is too small. The model with the parameter chosen on the elbow of the L-curve is shown in the bottom left. It has a better geologic structure than the other two, yet still fits the data. The true model is shown in the lower right for comparison.

**Generalized Cross-validation**

Another way to search for an appropriate regularization parameters is generalized cross-validation (GCV) (Golub et al., 1979; Wahba, 1990). The advantage of the method is that it does not require standard deviation. It first follows the leave-one-out principal, which is the basis for cross-validation or ordinary cross-validation (OCV). The idea is that a good recovered model will predict a datum even if it is not used in the inversion. Conceptually, the inversion is performed iteratively without a random observation location, which can be computationally intensive. In practice, more efficient algorithms exist (e.g. Golub et al., 1979; Wahba, 1990). The parameter is minimized with a given regularization parameter through the conjugate gradient method. Like the L-curve, the system of equations is minimized with a range of regularization parameters and the values of the GCV parameter are recorded. The
regularization parameter that corresponds to the smallest GCV parameter is then used to find the final solution.

2.3 Practical Aspects for 3D Inversion

For 3D inversion, there are two main additional and practical aspects that must be included to recover a geologically acceptable model. The first is enforcing positivity of 3D susceptibility distribution. The second is to weight the model in depth to force the recovered anomalous susceptibility to distribute below the model’s surface. This is to deal with the rapidly decaying kernels of magnetics. In this section, I briefly describe both methods.

2.3.1 Positivity Constraint

Unlike gravity, magnetic source anomalies are dipolar in nature and thus have a negative and positive signal. Therefore, it is desirable to mathematically solve for a positive recovered model. Equation 2.3 is expanded to:

\[
\min \Phi(\lambda) = \Phi_d + \beta \Phi_m - \lambda \sum_{j=1}^{M} \ln(\kappa_j),
\]

(2.9)

where \( \lambda \sum_{j=1}^{M} \ln(\kappa_j) \) is a log-barrier function to impose positivity (Wright, 1997). The problem is now non-linear. For each Tikhonov parameter, a cooling schedule of the log-barrier term, \( \lambda \), is imposed based on the step-length of the minimization process and \( \lambda > 0 \). Conjugate gradients are used to minimize \( \Phi(\lambda) \) at each iteration for a single log-barrier and regularization parameter. The process continues until stopping criteria for the global objective function are met. The above equation is used for the joint 3D inversion of gradient and total-field magnetic data and octree-based 3D inversion.
2.3.2 Depth Weighting

Magnetic data has no inherent depth resolution due to the decay of the kernels in the sensitivity matrix. In magnetics, the field decays at a rate of \( 1/r^3 \) where \( r \) is the distance between the source and observation locations. This causes cells at depth to have less influence on the data misfit than cells at the surface. The depth weighting function is based on the decay of the sensitivity matrix kernels to create an “equal” influence of the prisms regardless of depth during the inversion process. The depth weighting function has a general form of

\[
    w(r_j) = \left( \sum_{i=1}^{N} G_{ij}^2 \right)^{p/2},
\]

where \( p = 3 \) is chosen to match the decay of the magnetic kernels away from the source (Li and Oldenburg, 2000). For magnetic gradient data, \( p \) becomes 4. This function is incorporated into the model objective function as shown in Equation 2.8. In the case of an adaptive mesh, the depth weighting function is a vector of \( m \) length and incorporates the geometry and thickness of each prism within the mesh.

2.4 Summary

In this thesis, I present new methods for inversion-based processing and 3D inversion. The problems are formulated with Tikhonov regularization, described in this chapter, and are composed of two functions: the data misfit function and model objective functions. These serve as the fundamental elements to construct a model that fits the data within the survey errors as well as has structure. For 3D inversion, the model objective includes a depth weighting function, reference model, and smallest model component. In addition, a positivity constraint is added. For processing, the horizontal derivatives of the model are utilized to obtain a flatter model.
CHAPTER 3
USE OF GRADIENTS IN PROCESSING

Recently, the processing of aeromagnetic data to account for leveling has developed into the use of gradients. Utilizing multiple magnetometers allows for the measurement of magnetic gradients and minimizes the diurnal variation and other common-mode noise. In this chapter, I develop an equivalent source technique for jointly processing total-field and gradient data. I make use of a well-known but rarely used relationship between the derivatives of the magnetic field and the derivative of its source to relate both data sets to a common equivalent source distribution. This approach treats the observed gradients as an additional and independent data set instead of supplemental information. The direct result of joint processing is a set of enhanced data with higher information content and a signal-to-noise ratio. I present the methodology of the joint equivalent source processing technique and demonstrate it with a synthetic and field example. I show that the method diminishes higher frequency noise, accentuates mid-frequency signals, and has increased resolution over that of total-field data alone.

Airborne magnetic gradiometry data are becoming common in large scale exploration surveys (Reford, 2006). In most applications, multiple total-field sensors are utilized and the gradient is obtained by finite differences in the along-line and cross-line directions, respectively. There are advantages to these data types because of increased and complementary information content. The gradient itself provides advantages over the total-field anomaly alone, such as enhanced signal-to-noise ratio obtained from common-mode noise rejection. Gradients in airborne magnetometry have been used in the past to analyze two-dimensional bodies (Nelson, 1988) as well as to enhance the conventional gridding of total-field data by creating pseudo-lines (Reford, 2006), to improve the minimum curvature operator (O’Connell and Owers,
2008), or to correct for diurnal-related leveling errors (e.g. Nelson, 1994; Hardwick, 1999) in both spatial and wavenumber domains. Yet, much remains to be done beyond simple enhancement of the total-field anomaly to take full advantage of these gradients in data processing.

3.1 Equivalent Source Technique

Equivalent sources (Dampney, 1969) are fictitious source distributions constructed on a single layer below the observation surface to reproduce an observed set of potential-field data. One of the advantages of the method is that it has the ability to reproduce data, rather than minimum curvature gridding (e.g. Briggs, 1974; Swain, 1976; Webring, 1981). For magnetic datasets, a convenient type of equivalent source is magnetic susceptibility. In my application, the method solves for a common susceptibility layer that reproduces both total-field and gradient data. Forward modeling the response of the equivalent source layer can efficiently generate data on a uniform grid based on physics. With additional constraints on the equivalent source, one could also use it for a variety of processing such as upward or downward continuation and reduction-to-pole (RTP) (Silva, 1986). The more accurately gridded and de-noised dataset can naturally be used to perform and enhance a number of processing and interpretation techniques.

3.2 A Common Equivalent Source for Total-field and Gradient Data

The goal of this chapter is to combine magnetic gradient data as an independent data set with the total-field data to derive a high signal-to-noise (SNR) data set. There are two ways to achieve this goal. One way is to combine the total-field and gradients in data domain directly, such as gradient enhanced gridding. The other method is to construct a single equivalent source that reproduces both data sets. This is the approach I develop. The advantage of this approach is the extraction of signal and rejection of noise based on the inherent consistency of signal among
different data sets.

To accomplish this, I start with the mathematical relationship between the total-field anomaly and the equivalent source. For my purpose, I choose to use a layer with variable magnetic susceptibility under a given inducing field as the equivalent source. I then establish the relationship between derivative of the total-field anomaly and the derivative of the equivalent source so as to relate the gradient data and total-field data to the common source. For computational purpose, I discretize the equivalent source layer into a set of contiguous prisms and assume a constant susceptibility in each. The observed total-field is then given by Equation 2.1. For this chapter, I will refer to the total-field data as $\mathbf{d}_t$, so Equation 2.1 would be

$$ G_t \mathbf{\kappa} = \mathbf{d}_t. \quad (3.1) $$

As previously stated, the calculation of a sensitivity matrix, $G_t$, for total-field data assumes that each cell has an induced magnetization in the same direction as the inducing field. This is a valid assumption for the fictitious layer of equivalent sources. Under this assumption, the magnetic field is translation invariant. Therefore, its derivatives are related to the respective derivatives of the source. For example,

$$ \frac{\partial \Delta T[\kappa(x), x]}{\partial x} = \Delta T(-\frac{\partial \kappa}{\partial x}, x). \quad (3.2) $$

That is, the $x$–derivative of the total-field anomaly, $\Delta T$, is equal to the total-field anomaly produced by a source defined by the negative $x$–derivative of the original source (Kogbetliantz, 1944; Nabighian, 1972). It is important to note the negative sign in front of the derivative of the susceptibility, which comes from the fact that the total-field anomaly is given by a convolution of the source and a Green’s function.

I illustrate the utility of this relationship using the simple model in Figure 3.1. Assume there are two semi-infinite sheets with susceptibilities $\kappa_1$ and $\kappa_2$. Given that the gradient in practice is obtained by differencing two total-field measurements,
it is more appropriately expressed as a finite difference for the purpose of forward modeling,

\[
\frac{\partial \Delta T(x)}{\partial x} \approx \frac{\Delta T(x + \frac{\Delta x}{2}) - \Delta T(x - \frac{\Delta x}{2})}{\Delta x} = \frac{1}{\Delta x} \left[ \sum_{i=1}^{2} \Delta T(\kappa_i; x + \frac{\Delta x}{2}) - \sum_{i=1}^{2} \Delta T(\kappa_i; x - \frac{\Delta x}{2}) \right],
\]

where \( \Delta T(\kappa_i; x) \) denotes the total-field at location \( x \) produced by the susceptibility \( \kappa_i \). Using Equation 3.1, it can be shown that the \( x \)-derivative can be equivalently expressed as the total-field anomaly generated by a thin sheet with a width of \( \Delta x \) and susceptibility \( \kappa_1 - \kappa_2 \):

\[
\frac{\partial \Delta T(\kappa_1, \kappa_2, x)}{\partial x} = \frac{\partial \Delta T(\kappa_1 - \kappa_2, x)}{\Delta x},
\]

where the notation \( \kappa_1 - \kappa_2 \) derives from the need for negative derivative of the susceptibility. Thus, given a distributed susceptibility model, the gradient calculated from the total-field anomalies obtained through Equation 3.1 is identical to the total-field anomaly itself obtained from the sequence of thin sheets with widths of \( \Delta x \) and susceptibilities of \( \Delta \kappa \), the difference in susceptibilities, residing on the boundaries between adjacent prisms. This relationship has the general form of:

\[
G_x \Delta \kappa_x = \vec{d}_x
\]

where \( \vec{d}_x \) is the observed gradient in the \( x \)-direction, and \( \Delta \kappa \) are the differences in susceptibilities of two cells adjacent in \( x \)-direction, and \( G_x \) is the sensitivity matrix of \( \vec{d}_x \) with respect to \( \Delta \kappa_x \). The gradient sensitivity matrix, \( G_x \), similar to the total-field sensitivity matrix, is a dense coefficient matrix that is dependent upon the geometry of thin sheets and observation locations. As an illustration, I examine the combined effect of two semi-infinite slabs which can be given by a single thin sheet at the boundary of the two slabs with the difference of susceptibility in the negative
direction along the axis. This is shown by its geometrical relationship in Figure 3.2. The gradient of the total-field generated by a 50-m thick horizontal semi-infinite dike with a susceptibility of 0.5 S.I. units and a background of 0.0 S.I. units is shown in the upper left panel. The total-field anomaly generated by the thin sheet (with a width of 1 m) and difference of susceptibilities (-0.5 S.I. units) is shown in the upper right panel. The two profiles are numerically identical, illustrating the fundamental aspect of using the gradient as an independent data set and relating it to the total-field through a common layer of susceptibility.

The same relationship holds true for the $y-$gradient and is described by:

\[ G_y \Delta \kappa_y = \vec{d}_y, \quad (3.6) \]

where $G_y$ is the sensitivity matrix, $\vec{d}_y$ is the observed gradient, and $\Delta \kappa_y$ is the difference in susceptibilities of the two cells in the $y-$direction as in the $x-$direction. In practical applications, it maybe desirable to only utilize the cross-line gradient if the along-line observation locations are close compared to line spacing. The reason is that, if the line spacing is much greater than the data spacing along lines, too much
Figure 3.2. The gradient response (upper left panel) of a semi-infinite horizontal dike (lower left panel). The dike has a susceptibility of 0.5 SI surrounded with a background susceptibility of zero. The total-field anomaly (upper right panel) is generated by a thin sheet (lower right panel) with a width of 1 m and susceptibility of -0.5. The thin sheet is located where the semi-infinite dike ended and represents the change between the two susceptibilities ($\kappa_1 - \kappa_2$). The total-field anomaly of the thin sheet is equal to the gradient produced by the semi-infinite dike.

emphasize would be placed on the along-line direction of the equivalent source layer and lead to distortions. A direct manifestation of such a distortion is striping effects between lines when a total-field anomaly is computed for a uniform grid. Using the equations for the total-field and horizontal gradients, I show in the next section how these can be used to solve for a common equivalent source.

3.3 The Translation Matrix

I have now a set of linear equations for the gradient and total-field problem. The combination of Equations 3.1, 3.5, and 3.6 gives:

$$
\begin{bmatrix}
G_t & 0 & 0 \\
0 & G_x & 0 \\
0 & 0 & G_y
\end{bmatrix}
\begin{bmatrix}
\vec{\kappa} \\
\Delta \vec{\kappa}_x \\
\Delta \vec{\kappa}_y
\end{bmatrix} =
\begin{bmatrix}
\vec{d}_y \\
\vec{d}_x \\
\vec{d}_y
\end{bmatrix},
$$

(3.7)
which would require solving for three unknown vectors: \( \vec{\kappa}, \Delta \vec{\kappa}_x \) and \( \Delta \vec{\kappa}_y \). However, only \( \vec{\kappa} \) is needed for forward modeling of the equivalent source layer for subsequent processing such as gridding. It is in the best interest to solve for the susceptibility rather than the differences \( (\Delta \vec{\kappa}_x \) and \( \Delta \vec{\kappa}_y \)), but still utilize the gradients as auxiliary data sets to horizontally constrain the change of susceptibilities used to calculate the total-field. The question now arises as how to relate \( \vec{\kappa} \) with \( \mathbf{G}_x \) and \( \mathbf{G}_y \). For this, I introduce the translation matrices \( \mathbf{P}_x \) and \( \mathbf{P}_y \). The matrices represent cell interactions to describe the derivative between the \( i^{th} \) and \( (i+1)^{th} \) susceptibilities and translate each respective \( \Delta \vec{\kappa} \) to \( \vec{\kappa} \) by the relationship of:

\[
\mathbf{P}_x \vec{\kappa} = \Delta \vec{\kappa}_x, \quad \tag{3.8}
\]
\[
\mathbf{P}_y \vec{\kappa} = \Delta \vec{\kappa}_y. \quad \tag{3.9}
\]

These two matrices are only dependent on the number of parameters in the equivalent source layer. They are sparse and have only two values in each row: \(-1\) and \(1\). By substitution, Equations 3.5 and 3.6 now become

\[
\mathbf{G}_x \mathbf{P}_x \vec{\kappa} = \vec{d}_x, \quad \tag{3.10}
\]
\[
\mathbf{G}_y \mathbf{P}_y \vec{\kappa} = \vec{d}_y, \quad \tag{3.11}
\]

yielding the linear set of equations:

\[
\begin{bmatrix}
\mathbf{G}_t \\
\mathbf{G}_x \mathbf{P}_x \\
\mathbf{G}_y \mathbf{P}_y
\end{bmatrix}
\begin{bmatrix}
\vec{\kappa} \\
\vec{d}_t \\
\vec{d}_x \\
\vec{d}_y
\end{bmatrix}.
\]

\[
\begin{bmatrix}
\mathbf{G}_t \\
\mathbf{G}_x \mathbf{P}_x \\
\mathbf{G}_y \mathbf{P}_y
\end{bmatrix}
\begin{bmatrix}
\vec{\kappa} \\
\vec{d}_t \\
\vec{d}_x \\
\vec{d}_y
\end{bmatrix}.
\]

Thus, I have arrived at a set of linear equations that relates the susceptibility, \( \vec{\kappa} \), to a composite data vector containing total-field data as well as gradient data. However, the thin cell configurations \( (\Delta \vec{\kappa}_x \) and \( \Delta \vec{\kappa}_y \)) in Equations 3.10 and 3.11 are still implicitly present. I can denote this relationship with a compact notation as the same
in Equation 2.1:
\[ \mathbf{G}\vec{\kappa} = \vec{d}, \]  
(3.13)

where
\[ \vec{d} = \begin{bmatrix} \vec{d}_t \\ \vec{d}_x \\ \vec{d}_y \end{bmatrix}, \]  
(3.14)
and
\[ \mathbf{G} = \begin{bmatrix} \mathbf{G}_t \\ \mathbf{G}_x \mathbf{P}_x \\ \mathbf{G}_y \mathbf{P}_y \end{bmatrix}. \]  
(3.15)

These equations are linear and I can use them to solve for the susceptibility of the equivalent source using Tikhonov formalism as described in Chapter 2.

### 3.4 Synthetic Example

To demonstrate the algorithm, I use a synthetic data set produced by two dikes. The dikes are horizontal in dip, striking in the north-northeast and east direction, respectively. The inducing field is 52,000 nT in strength at an inclination of 65° and a declination of 25°. North-south flight lines were simulated at a 25-m height and 50-m line spacing. Two magnetometers were flown 6 m apart in the cross-line direction to simulate the total-field and cross-line gradient data. Data are recorded every 10 m along line. White Gaussian noise with a 1-nT standard deviation is added to the total-field data. Five percent standard deviation Gaussian noise was also independently added to the gradients. The minimum curvature gridded noisy line data are shown in Figure 3.3. The white dots indicate data locations. Using this data set, I compare total-field and joint equivalent source methods. First, an equivalent source layer was produced using only the total-field data. The forward modeled data from the equivalent source layer was then calculated at an even grid interval of 20 meters (upper left of Figure 3.4). Next, an equivalent source layer was produced from the
Figure 3.3. Synthetic "observed" data with noise (1 nT standard deviation). The observation locations are shown by the white dots. This data set is used to compare the joint and total-field equivalent source methodologies.

total-field and cross-line gradient data using an identical model objective function, and forward modeled total-field data were calculated on the same even grid (upper right of Figure 3.4). I then compare the difference of these two sets of forward modeled total-field data with the true without noise in lower right and left of Figure 3.4, respectively. The results from both equivalent source techniques have striping effects. These effects stem from the noise added to the line data. The north-northeast (along line direction) striking dike contains the most differences in the model. However, the joint equivalent source layer only had a maximum of approximately 9 nT difference in this area and the total-field equivalent source layer had a maximum of approximately 18 nT difference in the same area. The difference between the true and total-field equivalent source data also shows that considerably more striping is present in the total-field equivalent source data set as compared to the joint equivalent source data set. For a more quantitative comparison, Table 3.1 shows the L-2 and L-infinite
Figure 3.4. The evenly gridded data set calculated by a layer of equivalent sources from total-field data is shown in the upper left. The joint method data set is shown in the upper right. The corresponding differences from the true gridded response and responses from the equivalent source technique (lower right) and equivalent source with gradients (lower left) are shown. The difference maps are on the same scale. The joint methodology had half the errors and less striping than total-field equivalent sources.
Table 3.1. The L-2 and L-infinite norms of the data sets created by the joint and total-field equivalent source models as compared to the true forward model show the joint algorithm has 65% less error than using total-field data alone.

<table>
<thead>
<tr>
<th></th>
<th>Joint algorithm</th>
<th>Total-field only algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-2 norm</td>
<td>52.3</td>
<td>78.8</td>
</tr>
<tr>
<td>L-∞ norm</td>
<td>9.3</td>
<td>18.0</td>
</tr>
</tbody>
</table>

norm of the difference between the forward modeled and true data. In both cases, the forward modeled data from joint processing have approximately 60% less error of the total-field processing alone. The difference in resulting errors means that any quantitative interpretive techniques using the gridded data will be more accurate using the joint equivalent source algorithm than a total-field algorithm.

3.5 Field Example

For illustration, I apply the method to a field example is from the Central Abitibi Destor-Porcupine-Pipestone Faults area, near Matheson, Ontario (Ontario Geologic Survey, 2004). I use a subset of the data to demonstrate the joint-processing algorithm. The nominal station and line spacing is 15 m and 70 m, respectively. The total-field and cross-line gradient are incorporated in this study. To display the observed total-field line data, I first interpolate them to an even grid of 20-m intervals using minimum curvature gridding (Smith and Wessel, 1990). Figure 3.5 displays the gridded data with white dots indicating data locations.

The line data are then processed using both the total-field and joint equivalent source algorithms. The equivalent source covers an area of 9.065 km$^2$ and is divided into rectangular cells with widths of 20 m by 10 m in easting and northing directions, respectively. As discussed in the preceding section, optimal construction of the equivalent source and subsequent processing depends on fitting the observed data to a level consistent with the errors in the data. In the absence of specific knowledge about the errors, I use automated error estimation in the equivalent source construction. To do
so, there must be a reasonable estimate of the relative magnitudes of overall errors in the total-field data and in the gradient data. This is accomplished by comparing the radial power spectra of the gradient and total-field (Figure 3.6). The vertical axis is the logarithm of power of the two datasets. The noticeable shift in this axis is due to the different units of the two datasets. One important feature of the figure is the difference in their shapes. The power spectrum of the total-field levels sooner to a flat noise spectrum than that of the gradient data, indicating a higher signal-to-noise ratio (SNR) for the gradient data. Furthermore, these power spectra can be used to determine a relative error level of the gradient with respect to the total-field data based on the magnitudes of the power, allowing the specification of correct relative error levels for each dataset in the equivalent source construction. Assuming that the high-frequency band is dominated by noise, the noise power in total-field data is approximately 10 times (i.e. 20 dB) of the noise power in the gradient data. Thus, if I assume a 1-nT standard deviation for the total-field data, of the corresponding standard deviation of the noise in gradient data should be 0.32 nT/m (i.e., 1/\sqrt{10}). With this relative data weighting, the inverse problem of equivalent source reconstruction is solved as previously described with generalized cross-validation.

The joint and total-field equivalent source layers are forward modeled to the same even grid of 20 by 20 m (top and bottom panel of Figure 3.7, respectively) as with the
Figure 3.6. The power spectrum of the total-field and gradient of the observed field data is shown. There is a noticeable shift in power and the gradient has less noise than the total-field as expected. The difference in power will give the ratio of noise level between the two datasets in order to assign a consistent error level between the gradient and total-field data.

minimum curvature. As in the synthetic example, the same model objective function was used for both algorithms and both fit the data within the noise tolerance. The constructed equivalent source layer using the joint methodology is shown in the top panel of Figure 3.8 and the equivalent source model from using only the total-field is shown in the bottom panel of Figure 3.8. There are subtle differences in the model, specifically along the main anomaly. The joint recovered model has a thinner anomaly with the west portion of the dike not attached, unlike the total-field model. The data differences between the joint algorithm and the total-field algorithm as well as with the minimum curvature gridding are examined (top, middle, and bottom of Figure 3.9, respectively). The majority of the joint method’s data set is within ±5 nT of the minimum curvature. However, along the edges of the elongated anomaly the difference is as much as ±60 nT. Differences of this magnitudes will create substantially different
Figure 3.7. The calculated data sets that were produced by the joint (top) and total-field (bottom) equivalent source techniques at even intervals of 20 m.

Figure 3.8. The equivalent source layers produced with (top) and without (bottom) gradients. Both methods used the same model objective function. The model calculated without gradients has a wider and more connected anomaly than the joint technique.
results with processing or interpretation techniques that require gridding. It is also observed that there is approximately the same difference of ±5 nT in most of the data set as compared to a total-field equivalent source technique. The differences along the edges of the anomaly are an average of 30 nT in amplitude. There is also a striping effect when examining the difference between the two equivalent source methods. This minor striping occurs when using total-field equivalent sources because of the aspect ratio between the line and station spacing. The cross-line gradient aids in constraining the model to diminish this problem. The striping is very noticeable in the difference map between the equivalent source and minimum curvature methods (bottom of Figure 3.9). The normal equivalent source data set also fits most of the minimum curvature data to within ±5 nT (comparable to the joint methodology), but also has less amplitude of difference compared to that of the minimum curvature data. I also note that the difference on the edges of the anomaly between the joint method and minimum curvature are much more continuous.

I examine the radial power spectra of the re-produced data from the equivalent sources and minimum-curvature gridded dataset (Figure 3.10). The effect of de-noising through equivalent source methods is shown by the significant decrease in high frequency content compared with the minimum curvature data. Furthermore, an increase of power in the mid-frequencies is also noticeable. This increase is based on a combination of recovering information lost through minimum curvature and information gained with the gradients. For a quantitative approach of examining these spectra, the difference between the three methods is presented using the joint total-field and gradient equivalent source dataset as a baseline. Figure 3.11 shows the difference of power spectra in decibels with the abscissa in wavelength (m). At wavelengths of 600 m or longer, three datasets contain the same frequency content. However, power of the total-field equivalent source and minimum curvature data decreases significantly by as much as 15 dB in the band from 80 m to 250 m. This range includes the wavelength of the main anomaly in the dataset. This comparison not only shows an increase in power but also in resolution, because of the
Figure 3.9. The differences between predicted data from the joint equivalent source and the normal equivalent source (top), and the minimum curvature gridding (middle). The differences for most of the data maps are within 5 nT, but the edges of the anomaly differ by up to 50 nT for both the minimum curvature and total-field techniques. Also shown is the difference between the normal equivalent source and minimum curvature (bottom). Striping in the normal equivalent source technique is apparent in the two difference maps when compared to the joint methodology and minimum curvature.
increased high-frequency signal. Furthermore, the minimum curvature power is 40 dB higher where both equivalent source techniques diminish noise. The joint gradient and total-field equivalent source suppresses the noise by as much as 10 dB compared to using only the total-field. Increased resolution provided by the increased power at these wavelengths can potentially lead to much improved results when quantitative interpretation methods are used.

Figure 3.10. The power spectra of the observed data (dots), data created by total-field equivalent sources (black line), and data modeled by the joint total-field and gradient equivalent sources (crosses). The use of gradients increases power of the signal in mid-frequencies. Additionally, high frequency noise is suppressed greater than in the total-field equivalent source method or minimum curvature gridded data.

3.6 Summary

The joint equivalent source processing algorithm directly calculates susceptibilities by solving a linear inverse problem that includes information from the gradients as well as the total-field magnetic data. This technique creates a gridding tool with more resolving power by allowing for physics and gradients to aid in gridding as com-
Figure 3.11. The difference (dB) in power spectra from Figure 3.10 with respect to the total-field and gradient equivalent source data sets. The abscissa is wavelength in meters. The joint gradient and total-field method suppresses high frequency noise in the data by more than 40 dB compared to the minimum curvature and 15 dB with total-field equivalent sources. Additionally, the method contains more mid-frequency signal by 15 dB, illustrating the increase in resolution obtained. These wavelengths include the main anomaly of the data set.

The algorithm that has been developed retains higher frequency signal and de-noises the data more than the normal equivalent source technique or minimum curvature. Using the synthetic example, it was shown that the joint methodology has approximately half the error than that of total-field equivalent sources. The field example also had large (60 nT) differences arise between the minimum curvature and equivalent source techniques of which had comparable results. A comparison of power
spectra between the joint and total-field equivalent source layers revealed the use of gradients retained higher frequency signal and de-noised the data more than that of the total-field only method. A substantial resolution increase was also present and should increase the accuracy of quantitative interpretation techniques which use the generated data set.
CHAPTER 4
USE OF GRADIENTS IN 3D INVERSION

In this chapter, I use the methodology presented in Chapter 3 to create a 3D susceptibility distribution using gradient and total-field magnetic data. The approach treats the observed gradients as additional and independent data sets instead of being just supplemental information. These data are incorporated to spatially constrain the recovered model that will fit the total-field and gradient data. I present an expansion of work in Chapter 3 to a 3D susceptibility inversion. I also add the vertical gradient, if available and needed, and demonstrate the method with a synthetic and field example.

4.1 A Common 3D Susceptibility Model

The goal of this chapter is to combine magnetic gradient data as an independent data set with the total-field data to derive a 3D model of susceptibility distribution. The work in Chapter 3 combined the horizontal gradients and total-field with a common equivalent source model. The two horizontal gradients are described by

\[
G_x \Delta \kappa_x = \vec{d}_x \tag{4.1}
\]

\[
G_y \Delta \kappa_y = \vec{d}_y, \tag{4.2}
\]

where \( G_x \) and \( G_y \) are the sensitivities of thin sheets in the \( x- \) and \( y- \) directions, respectively. Likewise, for a three dimensional distribution, thin sheets can be constructed vertically when the vertical gradient is acquired. This additional gradient can be modeled by

\[
G_z \Delta \kappa_z = \vec{d}_z, \tag{4.3}
\]

where \( G_z \) is the sensitivity for thin sheets between the vertically layered cells within the model. The system of equations, given the total-field and three orthogonal gra-
Solving a single source model with multiple data sets is desirable. To accomplish this, I introduce the mathematical relationship between the gradient and the source distribution through translation matrices. These matrices are slightly different then previously described. They do represent cell interactions to describe the derivative between the \( i^{th} \) and \((i+1)^{th}\) susceptibilities. However, these sparse matrices represent the cell interactions throughout the mesh. This means if there are \( m_x, m_y, \) and \( m_z \) cells in the \( x-, y-, \) and \( z-\), directions of the mesh, respectively, \( G_x \) would have \((m_x - 1) \times (m_y m_z)\) thin sheets. Though these matrices are larger than the equivalent source due to the third dimension, they are still sparse and only contain -1 and 1s. The translation matrices translate each respective \( \Delta \vec{\kappa} \) to \( \vec{\kappa} \) by the relationship of:

\[
\begin{bmatrix}
P_x \vec{\kappa} = \Delta \vec{\kappa}_x, \\
P_y \vec{\kappa} = \Delta \vec{\kappa}_y, \\
P_z \vec{\kappa} = \Delta \vec{\kappa}_z,
\end{bmatrix}
\]

where \( P_x, P_y \) and \( P_z \) are the translation matrices for the three orthogonal directions. It is of interest to solve only for \( \vec{\kappa} \) for interpretation, but still use the gradient to constrain the model. The relation between \( \Delta \vec{\kappa} \) and \( \vec{\kappa} \) is

\[
\begin{bmatrix}
\vec{\kappa} \\
\Delta \vec{\kappa}_x \\
\Delta \vec{\kappa}_y \\
\Delta \vec{\kappa}_z
\end{bmatrix}
= 
\begin{bmatrix}
\vec{d}_t \\
\vec{d}_x \\
\vec{d}_y \\
\vec{d}_z
\end{bmatrix}
\]

(4.4)
The combination of Equations 4.4 and 4.8 yields

\[
\begin{bmatrix}
G_t \\
G_x P_x \\
G_y P_y \\
G_z P_z
\end{bmatrix}
\vec{\kappa} = 
\begin{bmatrix}
\vec{d}_t \\
\vec{d}_x \\
\vec{d}_y \\
\vec{d}_z
\end{bmatrix}.
\]  
(4.9)

Thus, I have arrived at a set of equations that relates the susceptibility, \( \vec{\kappa} \), to a composite data vector containing total-field data and horizontal and vertical gradient data. However, the thin cell configurations (\( \Delta \vec{\kappa}_x \), \( \Delta \vec{\kappa}_y \), and \( \Delta \vec{\kappa}_z \)) in Equations 4.5, 4.6, and 4.7, respectively are still implicitly present. I can denote this relationship with a compact notation as:

\[
G \vec{\kappa} = \vec{d},
\]  
(4.10)

where

\[
\vec{d} = 
\begin{bmatrix}
\vec{d}_t \\
\vec{d}_x \\
\vec{d}_y \\
\vec{d}_z
\end{bmatrix}
\quad \text{and} \quad
G = 
\begin{bmatrix}
G_t \\
G_x P_x \\
G_y P_y \\
G_z P_z
\end{bmatrix}.
\]  
(4.11)

I can use this system of equations to solve for a 3D susceptibility distribution in the same manner as described in Chapter 2.

Positivity constraints are applied to the solution through a log-barrier method. It is important to note that the positivity is applied to \( \vec{\kappa} \), but the change in susceptibility can be negative. The gradient can drive the solution from a larger positive amplitude to a smaller positive amplitude. The dimensionality of the problem requires depth weighing to offset the field decay. In theory, the depth weighting has an exponent of 3 for the total-field portion of the sensitivity matrix (\( G_t \)), and an exponent of 4 for the gradient kernels. However, through numeric experiments I found that using the slower decay rate for all of the sensitivities yielded the best model. This is because I am
explicitly solving for susceptibility and not the change of susceptibility. Furthermore, the total-field anomaly of the thin sheets reproduce the observed gradient. Therefore, it is desirable to incorporate the decay rate of the total-field anomaly rather than gradient. I now turn to synthetic and field examples to illustrate the method.

4.2 Synthetic Examples

I first introduce a vertical prism model as a synthetic example. The prism has a susceptibility of 0.05 SI units, inducing field strength of 52,000 nT, an inclination of 45°, and a declination of 15°. Five percent Gaussian noise was added to each dataset with a minimum of 3 nT for the total-field data and 0.002 nT/m for the gradients. The data are shown in Figure 4.1. The total-field is in the top-left corner. The vertical gradient is presented in the top right. The bottom panels show the $x-$gradient (left) and $y-$gradient (right), respectively. The coordinate system is Cartesian with vertical pointing down, $x-$axis pointing North, and $y-$axis pointing East. Data were inverted first using total-field anomaly and then using the joint algorithm with total-field and gradient data for comparison.

The vertical prism is shown in the left column of Figure 4.2. The top panel is a plan view at 170 m in depth. The bottom panel is a cross-section at 560 m Easting. The total-field recovered model is presented in the middle column. The maximum recovered susceptibility was 0.0125 SI units. The recovered model from total-field and the three orthogonal gradients is displayed in the right column. The results are spatially similar to the model recovered from total-field only. The amplitude of the recovered susceptibility, however, has been better preserved. For a quantitative comparison, the first two rows in Table 4.1 shows the L-2 and L-infinite norms of the recovered models compared to the true vertical prism model. Though the L-infinite norms are practically alike, the use of gradients increases the L-2 norm comparison. This means that the recovered model is much more accurate through the use of gradients, than that of total-field data alone.
Figure 4.1. Synthetic data of total-field (upper left), vertical gradient (upper right), $x-$gradient (lower left), and $y-$gradient (lower right). Five percent Gaussian noise was added to each dataset with a minimum of 3 nT for the total-field data and 0.002 nT/m for the gradient components. The Cartesian coordinate system has its origin at the Earth’s surface, $x-$axis points towards grid north, $y-$axis points towards grid east, and $z-$axis points vertically downward.

Next, I move to a more complicated example. The data were generated by a dipping slab (Figure 4.3) with a susceptibility of 0.05 SI units, inducing field strength of 50,000 nT, an inclination of 65°, and a declination of 25°. Total-field anomaly, $x-$, $y-$, and $z-$derivatives were calculated at an even grid of 50 m. Five percent Gaussian noise with a minimum of 1 nT for the total-field data and 0.002 nT/m for the gradients was added. The four panels in Figure 4.4 show the generated total-field, $x-$gradient, $y-$gradient, and vertical gradient data. I have adopted a Cartesian coordinate system having its origin at the Earth’s surface, $x-$axis pointing towards grid north, $y-$axis pointing towards grid east, and $z-$axis pointing vertically downward.

Again, I have inverted the data first with total-field anomaly only, and jointly
Figure 4.2. The true (left column) and recovered models using total-field data (middle column) and total-field with three gradients (right column). The shapes of the recovered models are similar, but the joint method recovers higher susceptibility values that are closer to the true model.

Figure 4.3. A cross-section half way through (600 m Easting) the true model that was used to generate data. The susceptibility is 0.05 SI units.
Figure 4.4. Synthetic data of total-field (upper left), vertical gradient (upper right), $x$-gradient (lower left), and $y$-gradient (lower right) for a dipping slab. Five percent Gaussian noise was added to each dataset. The inducing field strength is 50,000 nT with an inclination of 65° and declination of 25°.

with total-field and the three gradients. Cross-sections of the recovered models for each method are shown in the top and bottom panels of Figure 4.5, respectively. The improvement in amplitude and rate of change of the recovered susceptibility model is noticeable when gradients are incorporated. The recovered model is sharper and has a greater dip than that of the inverted model from total-field data only. The recovered model based on gradient and total-field also recovers both the depth and lateral extent of the true model. The model reaches 400 m in depth. For a quantitative comparison, Table 4.1 shows the L-2 and L-∞ norms of the recovered models and true model. The joint inversion has less error and thus recovers more of the true model than the total-field alone.
Figure 4.5. A cross-section half way through the recovered model using total-field only (top). A cross-section through the recovered model created by the joint inversion using the total-field and three orthogonal derivatives (bottom). The cross-section shows the same model region as Figure 4.3.

4.3 Field Example

To show the utility of the method, I present a field example of 1,066 total-field data from mineral exploration. The two horizontal gradients were observed for a total of 3,198 data to be inverted jointly. The total-field magnetic data is shown in upper panel of Figure 4.6. The northing \((x-\) and easting \((y-\) gradients are shown in the lower panels from left to right, respectively, on the same color scale. The inclination of the inducing field is 79° and the declination is 12°. Qualitative interpretation suggests there are two elongated features associated with mineralization that are separate from each other with minor anomalies on the side. The goals of this data set are to interpret any faulting between the two main anomalies that strike north-east
Table 4.1. A comparison of the recovered models with respect to the true models for the joint and total-field alone inversions for the two synthetic examples. The joint inversion recovered more of the true model than that of the total-field inversion in each case.

<table>
<thead>
<tr>
<th>Model / Norm</th>
<th>Total-field inversion</th>
<th>Joint inversion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vertical prism / L-∞</td>
<td>0.0452</td>
<td>0.0442</td>
</tr>
<tr>
<td>Vertical prism / L-2</td>
<td>0.9131</td>
<td>0.754</td>
</tr>
<tr>
<td>Dipping slab / L-∞</td>
<td>0.0492</td>
<td>0.0484</td>
</tr>
<tr>
<td>Dipping slab / L-2</td>
<td>0.9015</td>
<td>0.8705</td>
</tr>
</tbody>
</table>

and whether or not the south-east smaller anomaly is connected to the southern large amplitude anomaly. I will first invert the data using only the total-field anomaly. I then invert the data set using the total-field anomaly and two horizontal gradients.

4.3.1 Total-field Inversion Results

I start by inverting the data set using a mesh of $128 \times 64 \times 32$ prisms in the Easting, Northing, and vertical directions, respectively. The prisms are $60 \times 35 \times 25$ m in dimension. A volume rendered image of the recovered model is shown in Figure 4.7. I examine the higher range of susceptibilities which one may be more inclined to define as the target. Therefore, susceptibilities lower than 0.028 SI units are not shown. The two minor anomalies in the north are separate from the main anomaly of interest. It also becomes clear that there are two different sources in the elongated main feature. The northern source of this feature is slightly offset, which indicates there is a fault striking near north-west to south-east between the features. The smoothing of the model does not allow for the determination of whether or not the smaller south-east anomaly is connected with the main feature. There is a bridge between these two sources that could either be geological or due to mathematics from the model smoothness constraint. The depth of the center of the anomalies is at 450 m. This inversion has increased the knowledge of these sources as far as depth extent is concerned, but the second question of whether the smaller anomaly is connected
Figure 4.6. Observed total-field data (upper left), northing gradient (lower left), and easting gradient (lower right) from the mining industry used for 3D inversion. The inclination is 79° and in the declination is 12°. An inversion of only the total-field is performed for comparison. The two horizontal gradients will aid in constraining the model laterally via joint inversion.

to the large is still unknown.

4.3.2 Joint Inversion Results

Next, I invert for the gradient and total-field data jointly using the same mesh. Figure 4.8 is the recovered model with susceptibilities above 0.028 shown. The first observation is that the area of the questionable fault zone shows no high-susceptibility sources. It also is observed that the model is showing multiple smaller, compact anomalies in the north. The compact anomalies in the southeast also have a higher susceptibility than obtained from inversion of the total-field data alone. Moreover, the general shape of the anomaly is thinner using gradients, forcing a higher susceptibility
Figure 4.7. Recovered model using only total-field data with susceptibilities higher than 0.028 SI units shown. Due to smoothing parameters, it could be interpreted as a single body. Conversely, the pinch in the middle may also be a fault zone; the main anomaly is composed of two main source bodies.

than that of the total-field inversion. This exemplifies one of the advantages of the method by recovering more structure of the anomalies than from only total-field data. Even using the same type of measure for model complexity, the gradient data, which defines the edges of the anomaly, forces it to be thinner and less spread out.

The joint inversion result answers both questions posed in the beginning of the section. First, there is faulting between the two main anomalies. Furthermore, the northern anomaly has more structure rather than being simply one elongated source body. Secondly, the smaller anomaly in the southeast of the data set is not connected to the main southern anomaly. Though smoothing does join both features in the recovered model, when examining the higher susceptibilities, they are two distinct features.
Figure 4.8. Recovered model using the total-field and two horizontal gradients. Susceptibilities higher than 0.028 SI units are shown. The middle of the main anomaly is disconnected where the proposed fault zone is. The northern portion of the anomaly shows multiple compact bodies exemplifying the influence of the gradients. The main anomaly has more structure and is higher in susceptibility than that of the recovered model from the total-field data.

4.3.3 Comparison of Methods

For further analysis, I take a slice of both recovered models at a depth of 400 m. Figure 4.9 shows the recovered model from the total-field inversion (left panel), and the recovered model via joint inversion (right panel), respectively. The slices demonstrate the smoothing of the total-field inversion versus the slender, more structural model with the gradients. The recovered model based on the joint inversion is slimmer, and compact anomalies in the north are defined rather than one elongated body. Lower susceptibilities associated with smoothing extend further out with the total-field inversion. The southeast body has higher susceptibility.
4.4 Summary

The 3D joint inversion algorithm directly calculates susceptibilities by solving an inverse problem that includes information from the gradients as well as the total-field magnetic data. The gradients aid directly in the construction of the model. This technique leads to a quantitative interpretation with more resolving power. I have shown with the synthetic example that by using the extra information from gradients, the recovered model is constrained laterally and in depth extent. Quantitative analyses of the synthetic models show that the joint methodology recovers more accurate and compact bodies and higher susceptibilities than the total-field algorithm. The field example further demonstrates the ability of the method. The main feature is shown to be two bodies separated by a fault zone. The northern anomaly has much more structure and is thinner than the results from the total-field inversion. The results have implications for choosing drilling locations and calculating reserves.
Quantitative interpretation of magnetic data, for example Euler deconvolution (Thompson, 1982; Nabighian and Hansen, 2001), require a grid of data. One method of gridding often used is minimum curvature. Alternatively, the equivalent source technique (Dampney, 1969) can be used. The goal of this method is to solve for a layer of fictitious sources which reproduce the data within the noise. A gridded data set can then be calculated by these sources at any elevation. However, this valuable inversion-based processing approach can be computationally prohibitive for large problems.

In this chapter, I first introduce the equivalent source technique. I then use a data-adaptive model discretization in order reduce the number of sources prior to processing. It is also desirable to use the wavelet transform for its compression properties. I discuss how to apply the transform to an irregular mesh by implementing the transform to the kernel functions of the prismatic model elements for each datum. I also use the ordering the model parameters to increase compression ability. For a quantitative comparison of techniques, I will use compression ratio defined as the ratio of total coefficients in the dense matrix over the number of stored coefficients (i.e. after the wavelet transform). The reduction in storage requirement and computation time is approximately directly proportional to the compression ratio. A field example is used to illustrate the different steps of the method.

5.1 Introduction

Equivalent source construction (Dampney, 1969) typifies a large class of geophysical inverse problems. The method requires the solution for a layer of fictitious magnetic source distributions in a layer below the observational surface to reproduce
the observed magnetic data. One of the advantages of equivalent sources is that it has the ability to reproduce data in three dimensions based on physics, rather than minimum curvature gridding (e.g., Briggs, 1974; Swain, 1976; Wehring, 1981), for most magnetic datasets. I use single layer of prismatic cells each with a constant susceptibility to discretize the layer of equivalent sources. With certain constraints on the model, one could also use the equivalent source layer for upward or downward continuation or reduction-to-pole (RTP) processing (Silva, 1986). The more accurately gridded and de-noised dataset can naturally be used to perform and enhance a number of processing and interpretation techniques. To solve for the susceptibilities of the equivalent source layer, I employ Tikhonov regularization as outlined in Chapter 2.

Numerous methods have been developed in order to reduce the processing time or physical memory required. The Fourier transform is used to accelerate problems based on the convolution theorem (Cordell, 1992; Pilkington, 1997). The separable wavelet transforms (Li and Oldenburg, 2010; 2003) are used to compress the coefficient matrix by winnowing small coefficients that do not significantly change the function after the inverse transform. These techniques have been the principal approaches for accelerating this class of inverse problems in geophysics. The main limitations of processing in the Fourier domain are the restrictions placed on the data that they must be over a uniform grid and on a plane. The principal drawback of wavelet approach is the calculation of the dense coefficient matrix prior to applying the transform. However, the wavelet approach can be improved upon by reducing the size of the dense coefficient matrix prior to compression. For example, Ridsdill-Smith (2000) use wavelets along profiles of magnetic data in order to find edges in equivalent sources so that no unnecessary parameters are used.

Adaptive mesh algorithms are an alternative approach for model discretization. These are designed primarily to decrease the number of model parameters in forward and inverse modeling. Quadtree-based approaches have been utilized in 2D DC resistivity inversion (Eso and Oldenburg, 2007). Octree-based methods have been used for
3D inversion of electromagnetic data (Ascher and Haber, 2001). Furthermore, grid refinement (e.g. multi-grid methods) based on hierarchical mesh structures can increase the efficiency of non-linear problems (e.g. Haber et al., 2007). For linear problems, it is desirable to create an adaptive mesh prior to inversion and use it throughout. The actual reduction in model parameters from an adaptive mesh, however, is limited.

5.2 Quadtree Mesh Discretization

Quadtree mesh discretization has been primarily used in geophysics for remote sensing applications (e.g., Gerstner, 1999). The quadtree method is particularly useful in large scale problems by minimizing the number of required parameters. Quadtree design has $2^n$ number of cells in both the easting and northing directions as a result of a series of the model splitting to halves or regrouping in orders of two iteratively based on problem-dependent property values. For some problems such as 2D DC resistivity forward modeling, the quadtree structure is based on the geometry of the signal (Eso and Oldenburg, 2007). In the case of exploration magnetics, the extent of sources are only known laterally to a certain degree and therefore I adopt an algorithm adaptively dividing the mesh based on values that represent the severity of change of the observed field. This allows for a mesh architecture that places larger model cells where no or little signal change is present in the overlying magnetic data.

The values with which to choose the mesh are located in the center of each smallest cell (i.e. a regular mesh), enabling the actual discretization of the mesh to be fast. The observed data is coarsely gridded using a nearest neighbor interpolation of the data to the center location of each of the smallest cell size. The accuracy of the gridded data is not as important as the speed of the algorithm. However, the gridding will allow for an examination of a value at each model cell. It is important to note that this procedure is only used to choose a close-to-optimal mesh for inversion.

The adaptive quadtree discretization in general relies upon a criterion quantity to determine the location dependent resolution or size of the resulting elements. This
quantity is problem specific and may or may not be the quantity to be processed using the adaptive mesh. In remote sensing, for example, this quantity is the image intensity to be represented and stored. In potential-field processing or similar inverse problems in geophysics, however, the final quantity of interest is not available at the outset. I, therefore, must choose a proxy for use as the criterion quantity.

Formation of the quadtree structure of the equivalent source model should be based on a technique that will allow for the detection of the edges of anomalies such as gradients. It is well known that the gradient of potential fields aids in anomaly edge detection (e.g., Nabighian, 1972; Roest et al., 1992; Haney et al., 2003; Mushanyaddebu and Davies, 2006). However, regardless of the property used, the choice of property threshold is key and can highly influence the final result. I discovered that the total- and total-horizontal gradients were too noisy (most likely due to the gridding) and a proper dynamic threshold was difficult to implement. Likewise, the use of an apparent susceptibility in the Fourier domain (e.g., Bambrick et al., 1982; Yunsheng et al., 1985) was also very noisy, due to the implicit downward continuation that was involved. The basic problem of discretization becomes one of “local frequency.” Therefore, I use the curvature of the amplitude normalized by the amplitude of the magnetic field. The amplitude, $A$, of magnetic anomaly is given by

$$A = \sqrt{B_x^2 + B_y^2 + B_z^2},$$

(5.1)

with $B_x$, $B_y$, and $B_z$ the three orthogonal components of the anomalous magnetic field, $B_a$, respectively and are computed in the Fourier domain. The curvature, $c$, is approximated as

$$c = \sqrt{\left(\frac{\partial^2 A}{\partial x^2}\right)^2 + \left(\frac{\partial^2 A}{\partial y^2}\right)^2},$$

(5.2)

and computed by the second derivatives of the amplitude. This step is straightforward because the data values are already gridded and padded for the Fourier transform. The mesh is then split based on the value $q$ given by the ratio of the curvature over...
the amplitude:

\[ q = \frac{c}{A}, \]

such that if a change of \( q \) within a grouping is more than twice the standard deviation of \( q \) for the entire dataset, the cells in region of interest are split. The threshold prevents smaller cells from forming where small deviations in the data are present because of noise from acquisition or gridding. The sensitivity matrix (Equation 2.2) is calculated based on the nodal points of the quadtree cells and the inversion of the data can carried out as it would be for a regular mesh. As an example, Figure 5.1 shows values \( q \) (left panel) at the smallest cell size, and the corresponding discretization (right panel), respectively.

![Figure 5.1](image)

Figure 5.1. An example of the curvature of amplitude normalized by amplitude (left) and the corresponding quadtree discretization (right). Small prismatic cells located at the local anomaly allow for increased resolution. The result is the re-production of data from equivalent sources based on a smaller number of model parameters.

The quadtree mesh design has the ability to reduce the model parameters by one half to one-sixth on average and is highly data dependent. The algorithm is designed to compress the rows of the sensitivity matrix (i.e. model parameters). With the average reduction, the compression ratio for the quadtree mesh discretization alone averages between 3 – 6. Unfortunately, this is not enough compression to be
practical as the desired compression ratio should be a minimum of 50 to advance
current methodology. I next examine how wavelets can be applied to further reduce
computational cost.

5.3 Wavelet Transforms

The separable wavelet transform-based algorithms (Li and Oldenburg, 2003) ap-
plied to regular mesh have produced an average compression ratio of 20 – 50 in 3D
inversions. Using wavelets along profiles to identify edges in the equivalent source
to reduce the required parameters (Ridsdill-Smith, 2000) has led to an average com-
pression ratio of 25 – 30 in magnetic data processing. I now examine the possibility
of combining the wavelet compression with the quadtree discretization to achieve a
much higher compound compression.

The wavelet transform expands a function in the bases formed by the translation
and dilation of a function called the mother wavelet (Mallat, 1989; Daubechies, 1992;
Meyer, 1993). The particular wavelet transform I consider is based on orthonormal,
compactly-supported wavelets. The corresponding wavelet coefficients often have
entries that are small or nearly zero. Winnowing the coefficients whose magnitudes
are below a certain level still allows reconstruction of the original function to a high
degree of accuracy. Thus a function can have a sparse representation in the wavelet
domain.

Wavelets are defined recursively and possess three important properties: (1) the
wavelet is spatially localized; (2) it has a number of vanishing moments,

\[ \int_{-\infty}^{\infty} \psi(x)x^v dx = 0, \quad v = 0, \cdots, V - 1, \quad (5.4) \]

and (3) a set of orthonormal bases is formed from its translation, \( k \), and dilation, \( j \):

\[ \psi_{j,k}(x) = s^{-j/2}\psi(2^{-j}x - k). \quad (5.5) \]
Given the orthonormality, the reconstruction of a function through an inverse transform is easily obtained. The compression property of the wavelet transform arises from the presence of a large number of small coefficients in the transformed function. These small wavelet coefficients are produced because the wavelets are localized and orthogonal to low degree polynomials. Smoothly decaying portions of the matrix can be represented accurately by only a few wavelets at coarse scales. At low threshold levels, the reconstruction of the function is virtually identical to the function itself, yet only requires few significant elements near the location of the peaks. As the threshold level increases, small scale distortions begin to appear but the long wavelength features remain. This is the essential property of the wavelet transform that is utilized to compress functions.

With a regular mesh, each row of the sensitivity matrix can be compressed in the same way by using the separable wavelet transform. Analogous to a 2D Fourier transform, the 2D wavelet transform is a series of consecutive 1D transforms performed in each direction of the 2D mesh. With an irregular mesh, such as a quadtree discretization, the 2D separable wavelet transform is not applicable. Therefore, I apply the 1D wavelet to each row of the sensitivity matrix as if it were simply a 1D function. In general, wavelet compression increases when the function to be transformed has either clusters of high amplitudes or has large smooth segments represented by low-degree polynomials. For this reason, a simple ordering of the parameters may not produce the desired compression because the rows of the sensitivity matrix are neither clustered nor smooth. This occurs because each row will have peaks scattered throughout and many discontinuities will be present. To overcome this difficulty, I re-order the model parameters. The goal is to employ an indexing scheme that will group parameters will small sensitivities together consecutively while placing parameters with strong responses together; both groupings will require fewer coefficients in the wavelet domain. This approach is similar to the one used by Lamarque and Robert (1996) who used space-filling curves and 1D wavelets for simultaneous multi-scale edge detection and compression of an image.
5.4 Model Indexing

To obtain clusters of high-amplitude elements in rows of the sensitivity matrix, the ordering of the elements in an equivalent source layer should be spatially continuous. Minimizing large spatial jumps between consecutively ordered, or indexed, model parameters produces smooth, low amplitude segments and reduce the number of discontinuities. Such discontinuities require more significant wavelet coefficients to accurately reconstruct the row. I examine three different types of indexing and illustrate their performance with an example. For simplicity, I will illustrate these schemes with a regular model.

5.4.1 Continuous Coordinate Indexing

The first method for a continuous ordering is shown in the left panel of Figure 5.2. Starting in the southwest corner and simply moving in a user-axis direction, and then back will ensure continuity for a regular model. This straightforward technique works well for regular meshes. It can be adapted to work with irregular meshes as a search algorithm.

5.4.2 Continuous Radius Indexing

I next examine a consecutive indexing scheme that is similar to the previous but expands in a radial pattern. First, I begin indexing in the south-west corner, and then expand radially outward in a square pattern. An example is shown in the right panel of Figure 5.2. The change in location between model parameters next to each other in ordering is minimal for a full model and will ultimately decrease the number of wavelet coefficients needed to reconstruct a row of the matrix.
5.4.3 Hilbert Space-filling Curve Indexing

I now turn to a general means of mapping an $n$–dimensional space onto a one-dimensional space: space-filling curves (Butz, 1971; Sagan, 1994; Jin and Mellor-Crummey, 2005). Peano (1890) first presented a space-filling curve to efficiently and consecutively map every node for a closed $2^n$ square, appropriately called the Peano space-filling curve. The Peano curve has a general shape of an $N$ for the 1$^{st}$ order curve and is calculated recursively for higher orders. Likewise, Hilbert (1891) presented his own curve, by which the entire class of space-filling curves is now referred to in general (Moon et al., 2001). The Hilbert space-filling curve has a general $U$-type shape ($1^{st}$ order curve). Hilbert curves are currently used in image compression (Lempel and Ziv, 1984), multi-grid methods for up-scaling and down-scaling mesh (Aftosmis et al., 2004; Griebel et al., 1998; Behrens and Zimmermann, 2000), the N-body problem (Salmon et al., 1994), among other applications. Simple examples of Hilbert curves of orders 1, 2, and 3 onto a mesh are shown in Figure 5.3.
5.4.4 Application to Adaptive Mesh

All three schemes of ordering cluster consecutive model parameters spatially for a regular grid. However, in my application of an adaptive mesh, some of these grid points will be re-grouped into large mesh cells. The indexing schemes are turned into a search algorithm with two goals: (1) to ensure that all model parameters are accounted for and (2) to map adjacent model parameters spatially. It is important to note that two adjacent points after mapping should be close in space, but are not necessarily guaranteed to be close after mapping.

I introduce a conceptual quadtree discretization of the $8 \times 8$ regular base mesh
Figure 5.4. An example of spatial mapping of the $8 \times 8$ mesh example after quadtree discretization. The continuous coordinate mapping is shown in the left panel. The middle panel shows the result of the continuous radius mapping. The Hilbert curve (right panel) outperforms both by ordering every spatially adjacent cell. The upside-down $U$ shape exemplifies the regional movement of the first-order Hilbert curve (left panel of Figure 5.3). The arrows show spatial discontinuities.

in Figure 5.4. All mapping schemes start with the $4 \times 4$ cell in the southwest corner. The arrows indicate places of large spatial jumps between model parameters. The continuous coordinate mapping is shown in the left panel. One can observe the two large spatial discontinuities when applying this back-and-forth mapping algorithm. The middle panel is the result of continuous radius mapping and yielded similar results. The right panel of Figure 5.4 is the mapping with the Hilbert space-filling curve. It also starts in the southwest corner, but continues in an upside-down $U$ shape regionally around the mesh (e.g. Hilbert curve of order 1 from left panel of Figure 5.3). The second order $U$ shapes from the Hilbert curve is observed in the southeast corner of the mesh. There are no large spatial jumps with the space-filling curve and all cells are spatially continuous.

The Hilbert space-filling curve and the two continuous lattice curves map in two distinctly different ways. The continuous coordinate and radius map in line segments across the base mesh. The line segments are contiguous with a regular mesh, but can become detached with the grouping of parameters from quadtree or octree discretization. By contrast, the hierarchical Hilbert curve maps the mesh regionally.
It searches clusters of base nodal points so that all sizes of model cells within a region are mapped consecutively. This main feature of space-filling curves is the intuitive reason they should be superior to the continuous coordinate or continuous radius mapping. I now turn to a quantitative comparison.

5.4.5 Comparison of Indexing Methods

I compare the three methods used for ordering model cells to determine which one achieves the highest compression ratio with 1D wavelet transforms. All of the compression ratios given are after the transform has been applied with a reconstruction accuracy of 95%. I examine rectangular versus square, and quadtree versus regular meshes. For reference, data geometry for all examples is consistent with the field example. I perform the 1D wavelet transform on all cases and calculate the required relative threshold level to obtain the final compression ratio. I first examine a regular mesh (i.e., without quadtree discretization) and the three methods of indexing. Figure 5.5 shows the quadtree mesh used for the comparison and for the field example. The Hilbert curve outperformed the continuous indexing schemes by more than a factor of 2 for square meshes and had higher compression with rectangular meshes. A randomly selected row of the sensitivity is used for illustration. It is important to note that the same row is shown in each indexing scheme, and only the ordering of that information has changed.

I first analyze a full square mesh of $256 \times 256$ cells. The upper panel of Figure 5.6 is a row of the sensitivity matrix after simple continuous coordinate ordering. Scattered high-amplitude features will decrease compression with the wavelet transform because more coefficients are needed. The continuous radius indexing (middle panel of Figure 5.6) of the same row is slightly more condensed, but is not as compact as the Hilbert space-filling curve shown in the bottom panel of Figure 5.6. The large gap between two groups of high amplitude features in the row after Hilbert curve indexing only needs a few wavelet coefficients to represent it accurately and thus outperforms
Figure 5.5. The quadtree mesh used for the field example and comparing ordering schemes. The mesh is based on a rectangle of $244 \times 174$ model parameters and consists of approximately 11,000 prisms.

the other two schemes. Figure 5.7 is an inset of the high amplitude region for the Hilbert curve to show sparseness.

I perform the indexing on a regular rectangular mesh of $244 \times 172$ parameters. Figure 5.8 shows one row of $G$ for the three methods of indexing on this mesh. High-amplitude features are scattered for both continuous methods. The results of these two re-ordering approaches are similar to the square mesh case. The Hilbert curve still outperforms both, although not as significantly.

Comparable results with the quadtree discretization were observed. The quadtree algorithm was implemented on both the square and the rectangular meshes. Table 5.1 is a comparison of compression for the three indexing techniques with the full and quadtree mesh designs. It is clear that the Hilbert curve is more advantageous over either continuous lattice type of ordering in general. It is better to have a square base mesh and use the Hilbert space-filling curve because the compression is much higher than that of a rectangular-based mesh. In the case of the quadtree, the mesh was discretized to 16,000 cells for a $256 \times 256$ mesh and 11,000 cells for the $244 \times 172$ geometry. Yet, the overall compression was more than doubled. The Hilbert curve will have the largest relative decrease in compression for a $2^n - 1 \times 2^m - 1$ model. The actual addition of model cells to create a $2^n \times 2^m$ mesh will enable more overall storage than
that with fewer cells. It has been shown that even without quadtree discretization, a space-filling curve will offer more compression than a more conventional continuous system for distance-based kernels.

5.5 Example

I apply the method to field data from the Central Abitibi Destor-Porcupine-Pipestone Faults area, near Matheson, Ontario. The data were acquired for the Ontario Geologic Survey (2004). I use a subset of the data to demonstrate the combined compression capabilities of the quadtree, wavelet, and Hilbert curve algorithms. The nominal station and line spacing are 15 m and 70 m, respectively.
Figure 5.7. Inset of the cluster of high amplitude features from the Hilbert curve as shown in Figure 5.6. The sparseness of the parameters follows from the spatial clustering and continuity produced by the space-filling curve. This feature allows the reconstruction of the row using only a few significant wavelet coefficients, thereby increases the compression ratio by threefold compared to the other two methods.

Table 5.1. Compression based on model indexing and the 1D wavelet transform. The smallest wavelet coefficients are discarded to obtain a 95% accuracy.

<table>
<thead>
<tr>
<th>Mesh type</th>
<th>Cont. coord.</th>
<th>Cont. radial</th>
<th>Hilbert curve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full (rect.)</td>
<td>206.5</td>
<td>320.3</td>
<td>384.3</td>
</tr>
<tr>
<td>Full (square)</td>
<td>211.8</td>
<td>282.3</td>
<td>629.0</td>
</tr>
<tr>
<td>Quadtree (rect.)</td>
<td>405.6</td>
<td>533.0</td>
<td>584.9</td>
</tr>
<tr>
<td>Quadtree (square)</td>
<td>426.6</td>
<td>599.9</td>
<td>1227.0</td>
</tr>
</tbody>
</table>
Figure 5.8. A random row of $G$ from a rectangular-based quadtree mesh with 11,000 parameters. The continuous (top) and radially continuous (middle) yield similar results. Like the square-based mesh, the Hilbert curve (bottom) outperforms the two other methods and only has two main groups of peaks.

I process the line data consisting of 6,145 observations (shown with contours in Figure 5.9). The equivalent source covers an area greater than one million square meters and is discretized into rectangular prisms with widths of 20 m by 10 m in Easting and Northing directions, respectively. The full model consists of 38,305 cells ($244 \times 172$). The quadtree discretization produces a final model of 10,948 cells, approximately one quarter of the full model. The compression ratio with the three-step approach using continuous ordering was 211. The generated quadtree mesh is shown in Figure 5.5. The model indexed using the Hilbert space-filling curve followed by the 1D wavelet transform leads to a compression ratio of 585. I solve the inverse problem of equivalent source reconstruction by the conjugate gradient (CG) method and choose the regularization parameter by the generalized cross-validation (GCV).
Figure 5.9. The observed data for equivalent source processing from the Central Abitibi Destor-Porcupine-Pipestone Faults area (Ontario Geologic Survey, 2004). The indexing schemes were compared with the 6,145 data. With the quadtree mesh, Hilbert curve, and 1D wavelet transform, the coefficient matrix was compressed 1227 times and had similar data and model results after processing.

criterion (Golub et al., 1979; Wahba, 1990) as discussed in Chapter 2.

As expected, results from the adaptive quadtree and regularly gridded models are quite similar (top and bottom of Figure 5.10, respectively). The quadtree model has higher resolution where needed near the main anomaly. The full sensitivity matrix consists of 235,384,225 entries. With the method presented, a comparable model and data misfit were obtained with a dramatic reduction to only 402,435 entries in the sparse representation of the sensitivity matrix.

5.6 Summary

Equivalent source processing in potential-field methods involve large and dense coefficient matrices that often exceed the limitations of physical memory in commonly available computers. The adaptive quadtree approach is useful in reducing the number of model parameters, effectively reducing the $m \times m$ system of equations needed for inversion-based processing. Model ordering via Hilbert space-filling curves and the 1D wavelet transform for the compression of rows in the kernel matrix create a three-step approach towards achieving maximum storage potential. Tradi-
Figure 5.10. The recovered susceptibility model using a normal (top) and quadtree mesh (bottom).

tional methods have achieved compression ratios between 10-50 on average. The 2D separated wavelet transform assumes an regularly gridded mesh and thus is not applicable to quadtree mesh discretization. Instead, I use a 1D wavelet transform on each row of the sensitivity matrix. To optimize the efficiency of the wavelet transform, I re-order the model using the Hilbert space-filling curve creating few smooth high amplitude regions that can be compressed using a minimal number of significant coefficients. With fewer coefficients needed for the wavelet and the prior use of quadtree mesh discretization, I am able achieve much more compression (1227) than traditional methods as shown by the example. The three-step method is capable of achieving 100s to 1000s of compression in processing any given data set.
Numerous methods have been developed in order to reduce 3D inversion time or physical memory required. In general, the most computational efficiency has been achieved through integral transforms. These not only increase the speed of solving the inverse problem, but also enable larger data sets to be inverted on a computer with limited resources. The Fourier transform accelerates problems based on the convolution theorem (Cordell, 1992; Pilkington, 1997). The separable wavelet transform has a compression property by winnowing small coefficients that do not significantly change the function after the inverse transform (Li and Oldenburg, 2010, 2003). These two techniques have been the principal approaches for accelerating this class of inverse problems in geophysics. I presented a quadtree approach in the last chapter to process magnetic data. In this chapter, I expand the research to an octree-based 3D inversion method in order to minimize the number of model parameters in the inverse solution. I incorporate wavelet transforms and 3D Hilbert space-filling curves for the ordering of parameters. I show the effect of the technique through a synthetic and field example.

6.1 Introduction

Solving the system of equations to invert for a model that reproduces the data with a dense matrix of size $m \times m$, requires at least $O(m^3)$ operations. For large data sets, inversion can become computationally prohibitive. One way of increasing the efficiency is minimizing the number of parameters, $m$, that are required. This is the topic of this chapter.

Octree-based methods have been used for 3D inversion of electromagnetic data (Ascher and Haber, 2001). Further more, grid refinement (e.g. multi-grid methods)
based on hierarchical mesh structures can increase the efficiency of non-linear problems (e.g. Haber et al., 2007). I create an octree-based mesh prior to inversion and use it throughout because of the linearity of the problem. For a quantitative comparison of techniques, I will use the previously defined compression ratio. The reduction in storage requirement and computation time is directly proportional to this ratio.

6.2 Octree Mesh Discretization

Octree mesh discretization minimizes the number of model parameters by re-grouping finer cells into coarser ones in regions where high resolution is not needed. The approach keeps finer cells where high resolution is needed to recover features in a model. The method has been primarily used in geophysics for non-linear inversion such as electromagnetics (Ascher and Haber, 2001; Haber and Heldmann, 2007) and large-scale earthquake finite element modeling (Kim et al., 2002; Bielak et al., 2005). The octree method is particularly useful in large scale problems. Octree design has $2^i \times 2^j \times 2^k$ number of prisms in the base mesh as a result of a series of the model splitting to octants iteratively. There are a few different approaches one can use to incorporate an octree mesh in geophysical inversion. Information such as the source geometry, known geology, or a reservoir in time-lapse cases, can be used to pre-define model cells. Using the geometry of a source and receiver configuration, one can direct mesh resolution, such as in the case of electromagnetics. In the case of exploration magnetics, the extent of sources are only known laterally to a certain degree. Therefore there are two main techniques that can be used. First is an algorithm that divides the mesh laterally (top panel of Figure 6.1), then carries a consistent aspect ratio of the prisms in depth (middle panel of Figure 6.1). This would be considered a 2.5D quadtree, however, and would also not achieve the maximum reduction of model parameters. The second method on which I will focus is one that divides the mesh laterally and uses a depth weighting function (e.g. Li and Oldenburg, 1996) in order to further discretize the cells vertically (bottom panel of Figure 6.1) based on field
decay. The depth weighting function will allow the sizes of cells to increase as the sensitivity becomes weaker at depths.

The adaptive octree mesh architecture places larger model cells where little or no signal change is present in the overlying magnetic data. Smaller cells are assigned to regions requiring increased resolution such as edges of source bodies in order to properly fit overlying, rapidly-changing magnetic data. The adaptive octree discretization in general relies upon a criterion quantity to determine the location-dependent resolution or size of the resulting elements. This quantity is problem specific and may or may not be the quantity to be processed using the adaptive mesh. For example the criterion can stem from known geology, reservoirs, or transformed data. In my case, I will assume no knowledge about the subsurface and will choose a proxy of the observed field for use as the criterion quantity similar to the work presented for quadtree mesh discretization in Chapter 5.

For the criterion quantity with which to split the mesh, I begin with the curvature of the amplitude normalized by the amplitude. The top layer of cells in the mesh are assigned values exactly like in the 2D case. Unlike quadtree, however, expanding to three dimensions requires some knowledge in depth for discretization. Although I have assumed no specific knowledge of the subsurface, it is known how the field decays as a function of distance. Therefore, I continue the values in depth, multiplying them by the depth weighting function at each layer based on the geometry of the gridded mesh and observation locations. The depth weighting allows groups of smaller cells to be re-grouped into larger cells as they approach the bottom of the mesh and lose individual sensitivity.

As with the quadtree, I use one standard deviation of the criterion as a threshold of change within large groups for the re-grouping of parameters. The calculation of the standard deviation is done using the top layer of cells only since the values become smaller as depth increases (due to the weighting function). Figure 6.2 shows an example of the mesh discretization. The upper left panel shows locations of the smallest cells, representing the original base cells. The upper right panel shows the cell
Figure 6.1. An example of a 2.5D quadtree mesh shown in plan view (Top). A slice of the same mesh cut through the main anomalous zone in depth (middle). The cells retain the horizontal size assigned from the top view. The bottom panel is a slice when the depth weighting function is incorporated into the discretization to further reduce the number of cells vertically.
Figure 6.2. An example of mesh discretization through the octree process. The upper left panel shows all of the cells that are of minimum cell size (i.e. the original cell widths). The majority of these cells are located in the top of the inverted anomalies to the surface of the mesh. The upper right panel shows locations of all cells grouped into $2 \times 2 \times 2$ the smallest cell size. The bottom left and right panels show groups of $4 \times 4 \times 4$ and $8^3$ number of smallest cells, respectively. The top-south cube of the grouped cell locations are shown. The resulting mesh is 29,499 cells rather than 262,144 cells.

locations that were re-grouped to $2 \times 2 \times 2$ base cells. The bottom left and right panels shows where groups of $4^3$ and $8^3$ cells are also located, respectively. The majority of the cells grouped with $8^3$ small cells lie at the base of the mesh. Conversely, the smallest cells lie in the regions of interest and towards the top of the mesh.

The octree mesh design has the ability to reduce the model parameters by an order of magnitude or more on average and is highly data dependent. Therefore, the compression ratio for the octree mesh discretization is typically larger than 10. Unfortunately, this is not a high enough compression to be practical as the desired compression ratio should be at least 50 to advance current methodology. Similar to the techniques presented for quadtree-based processing, I will use wavelet transforms
and the ordering of parameters to increase compression.

### 6.3 Wavelet Transforms

The wavelet transform expands a function in the bases formed by the translation and dilation of a function called the mother wavelet (Mallat, 1989; Meyer, 1993). The wavelet transform of a function that I use is based on orthonormal, compactly supported wavelets (Daubechies, 1992). Winnowing the coefficients whose magnitudes are below a certain level still allows reconstruction of the original function with a high degree of accuracy. Thus a function can have a sparse representation in the wavelet domain.

With a regular mesh, separable 3D wavelet transforms can be applied to each row of the coefficient matrix \( G \). Analogous to a 3D Fourier transform, the 3D wavelet transform is a series of consecutive 1D transforms performed on each respective direction of the 3D mesh. Since the wavelet transform is applied to the mesh and not the data, it is unnecessary for the data to be regular unlike the Fourier transform that requires both to be gridded. However, the wavelet transform implies that the mesh is regular. With an adaptive mesh such as with an octree discretization, the separable wavelet transform is no longer applicable. Therefore, I apply the 1D wavelet to each row of the sensitivity matrix as if it were simply a 1D function. Similar to Chapter 5, clusters of high-amplitude features are desired to increase the compressibility of the transform. Therefore, I turn to the model ordering in three dimensions.

### 6.4 3D Model Ordering

To increase the compression of the rows, I attempt to cluster high-amplitude features within each row of the sensitivity matrix by re-ordering the model parameters. Indexing the model cells appropriately results in a large number of small model responses grouped consecutively, which require few coefficients in the wavelet domain. The idea is similar to the one used by Lamarque and Robert (1996) where they used
space-filling curves and 1D wavelets for multi-scale edge detection that had a by-
product of high compression for images. The approach of re-ordering the model is to
optimally map the 3D structure of the model mesh onto the 1D row of the sensitivity
matrix for a given datum. Based on my previous work, I utilize the general means of
mapping an \( n \)-dimensional space onto a one dimensional space with the Hilbert curve
(Hilbert, 1891). Three dimensional space-filling curves have been applied in multi-
grid methods for up-scaling and down-scaling mesh (Griebel et al., 1998; Behrens and
Zimmermann, 2000; Aftosmis et al., 2004). The regional trend of the curve allows for
the minimization of large spatial jumps of consecutively ordered cells in the octree
mesh.

The Hilbert curve starts in the top south-west prism of the mesh. The base
mesh is assumed to have \( 2^i \times 2^j \times 2^k \) cells, the same requirement for both octree
discretization and Hilbert curve. If \( i = j = k \) is not true, multiple Hilbert curves
are merged throughout the mesh. The curve searches throughout the base mesh
regionally to consecutively order the neighboring prisms in the octree mesh to ensure
spatial continuity. Simple examples of 3D Hilbert curves of orders 1 \((2^1 \times 2^1 \times 2^1)\) and
2 \((2^2 \times 2^2 \times 2^2)\) onto a base mesh are shown in the top and bottom panels of Figure
6.3, respectively. The Hilbert curve moves regionally throughout the mesh expanding
the two-dimensional \( U \) into three dimensions.

As an illustration, I show the response of a datum for 68,286 prisms after oc-
tree discretization for normal and Hilbert curve ordering. The row created by the
normal ordering is presented in the top panel of Figure 6.4 and is prior to wavelet
compression. The inset exemplifies the scattering of high sensitivities associated with
the normal ordering. The bottom panel shows the same row after the 3D Hilbert
curve. Two main clusters are present. The first cluster is shown in the inset. This
example exhibits the similar results to the indexing schemes previously discussed.
Fewer wavelet coefficients with larger amplitudes are expected with high-sensitivity
clusters. The normal indexing requires 903 non-zeros after the 1D wavelet trans-
form. The same row via Hilbert curves needs 578 wavelet coefficients. The resulting
Figure 6.3. The Hilbert space-filling curve is used to order model parameters for the 1D wavelet. The first order curve (top) for a $2^1$ cube. The second order curve (bottom) is for $2^2$ cube. The Hilbert curve allows for model cells to be spatially continuous during the adaptive mesh parameter ordering. The ordering starts in the south-west top corner and concludes in the north-west top corner.
compression ratios are 76 and 118, respectively.

6.5 Inversion Results

I now have a three-step approach to obtain high compression in the 3D inverse problem. First, the number of model parameters are reduced through octree mesh discretization. The parameters are then ordered through Hilbert space-filling curves to isolate clusters of high-amplitude peaks of sensitivity. Finally, the 1D wavelet transform applied to each row of the sensitivity matrix. The increase of compressibility of the wavelet transform stems from the need of few wavelet coefficients between the high-amplitude clusters.

To further validate the method, I present a small field example of 1,066 data from the mining industry. The magnetic data are shown in Figure 6.5 with the observation locations indicated by white dots. The inclination is 79° and the declination is 12°. Qualitative interpretation suggests there are two elongated features associated with potential zones of mineralization that are offset from each other with minor anomalies on the side. There are two important research questions about this data set: (1) is there any faulting between these two main anomalies that strike north-east and (2) is the south-east smaller anomaly connected to the southern large anomaly? For interpretation of the data set, I invert the data first using the traditional method of a regular mesh and then with octree-based discretization.

6.5.1 Traditional 3D Inversion

I start by inverting the data set using a gridded mesh of $128 \times 64 \times 32$ prisms in the easting, northing, and vertical directions, respectively. The prisms are $60 \times 35 \times 25$ m in dimension. A volume-rendered image of the recovered model is presented in Figure 6.6. Susceptibilities lower than 0.012 SI units are not shown. The two minor anomalies in the north are separated from the main anomaly of interest. The northern source of the elongated feature may be slightly offset, but it is not distinct enough
Figure 6.4. Example of two different model ordering schemes for the same row of a sensitivity matrix. The row has already been adaptively discretized to 68,286 parameters. The top panel is the model response after normal indexing. The inset shows the scattering of the high-amplitude clusters over 2,000 parameters. The bottom panel is the ordering after 3D Hilbert space-filling curves. Two main clusters are present with the inset showing the first cluster. The row was compressed by the 1D wavelet transform and resulted in compression ratios of 76 and 118 for the continuous and Hilbert curve ordering, respectively.
to be interpreted as a fault. The smoothing of the model also makes it unclear whether or not the smaller south-east anomaly is connected with the main feature. The bridge between these two sources could either be geological or mathematical from the model smoothness constraint. The depth of the center of the anomalies are at approximately 450 m. This inversion has only increased the knowledge about the depth of these sources.

### 6.5.2 3D Inversion with Octree Mesh

I next invert the data set discussed in the previous section using the octree-based mesh discretization. I use the same mesh from the traditional inversion as the base mesh. Initially, the curvature of the amplitude of the magnetic data normalized by the amplitude is employed as a horizontal discretization criterion as stated earlier. This is the same scheme that has worked for quadtree-based processing. It did not work, however, because most of the anomalies in the smaller data set are within a
similar amplitude range. The criterion is not as robust in this specific case as it would be with large data sets that include a wide range of frequencies and amplitudes of anomalies. A different approach should be taken specifically when the anomalies are similar in amplitude and close together. I next examine derivatives of the field with which to map the edges of anomalies.

I first present the current criterion of the curvature of the magnetic amplitude normalized by the amplitude (upper left of Figure 6.7). The curvature of the total-field anomaly is shown in the upper right. This criterion fails with the smaller anomalies associated with this data set. Next, I examine the total-horizontal gradient (THG) of the total-field data (lower left). This criterion places broad high values on edges of anomalies. These broad features present a challenge by forcing the grouping of small cells in places where one would aim to keep smaller cells. The algorithm searches for changes in criterion so it is more valuable to have consistent values throughout the anomaly and high rates of change on the edges. The desire of constant values within
anomalies and non-eventful places is the reason the curvature of the amplitude was used in the first place for adaptive-mesh processing in Chapter 5. For the processing of large spatial data sets, variable amplitude features are observed and thus I normalize the curvature by its amplitude. Therefore, when similar amplitudes are present, it is best to use the curvature of the amplitude (lower right of Figure 6.7). The choice then becomes whether or not to normalize by amplitude is data set dependent. This decision can be left up to the interpreter creating the most appropriate octree mesh for inversion of a specific data set.

Based on the above discussion, I choose to use the curvature of the amplitude
as the criterion with which to laterally discretize the mesh for this data set. The values are extended vertically in the base mesh and multiplied by the depth weighting function. The standard deviation of the curvature (e.g. lower right panel of figure 6.7) for the mesh’s top layer is used as a threshold value for the octree discretization. The result is a gradual coarsening of the mesh in depth. I apply this method to the field data set previously discussed.

Of course, the method is not useful unless the resulting model is comparable with the base-mesh inversion result. Figure 6.8 is the recovered model of the octree-based inversion. Susceptibilities under 0.012 SI units are not shown to be consistent with the solution from traditional inversion. The features between the two inversions are similar. The two smaller features in the northwest are present as well as the two main sources in the middle of the mesh. These two main anomalies are slightly offset in the easting direction, but it is still not clear enough to declare a fault offset between them. The approximate depth of these anomalies is 500 m, two cells greater than the smoother inversion. The smaller source located in the southwest is not connected to the south main anomaly. The mesh design place larger cells between the two and insignificant susceptibilities were placed there. This answers the question that the traditional method is connecting the two because of model smoothness rather than true geology.

6.5.3 Comparison of Methods

The octree discretization produces a mesh with 29,499 rather than 262,144 cells. This is a reduction of model parameters by 8.9 times. Adding the Hilbert-curve ordering and 1D wavelet with a reconstruction error of 5%, the total compression amounts to 354 times from the full matrix calculations. The 3D wavelet transform used in the regular-mesh based inversion was able to compress the sensitivity matrix 123 times. Though it may seem that 3 times more compression is insignificant, it is important to note that the larger the data set and more variables in the mesh,
Table 6.1. A comparison of practical aspects of storage and computational efficiency for the field example. The octree-based mesh compresses the sensitivity matrix by three fold, requiring less storage and solving the inverse problem in under a quarter of the time. The sensitivity matrix for the full problem without compression would consist of 278,914,122 entries. Though the example is a smaller problem, it exemplifies the utility of the method. The difference between these two techniques will increase as the size of the problem increases.

<table>
<thead>
<tr>
<th></th>
<th>Regular mesh</th>
<th>Octree mesh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compression ratio</td>
<td>123</td>
<td>354</td>
</tr>
<tr>
<td>Entries in matrix</td>
<td>2,263,584</td>
<td>787,893</td>
</tr>
<tr>
<td>Matrix file size (Mb)</td>
<td>19</td>
<td>11</td>
</tr>
<tr>
<td>Inversion time (sec)</td>
<td>579</td>
<td>121</td>
</tr>
</tbody>
</table>

the higher compression will be achieved. Furthermore, even at this compression it could mean 2 days rather than 6 days to obtain a solution. Table 6.1 compares the practical aspects of the compression including computation time, file size of the sensitivity matrix, and compression ratios between the octree-based and regular-mesh inversion algorithms. Although the reduction in inversion time does not seem that significant, it is important to note the discrepancy will increase with the problem size.

Figure 6.9 shows a horizontal slice of the recovered models at depth with octree and traditional inversion algorithms, respectively. It is observed that the south-east anomaly recovered in octree inversion is not connected to the main anomaly. This is based on an implicit regularization through cell sizes from the large cells between the two regions. The susceptibility for these large cells has to be small, enabling the model to fit the data. The traditional inversion can place slightly higher susceptibilities in the small cells and smooths between sources. The final octree result is blockier, but consistent with a regular mesh model.

6.6 Summary

Inversion of magnetic data involves large and dense coefficient matrices that often exceed the limitations of physical memory in commonly available computers. The adaptive octree approach is useful in decreasing the number of model parameters, ef-
Figure 6.8. The recovered model through inversion using an octree-based mesh. Two main elongated features are associated with the large anomaly and three smaller sources are also present. The southern small anomaly is shown not to be connected to the southern larger anomaly, unlike the smoother recovered model from the regular mesh solution. The octree mesh design solves for 29,499 parameters compared to the regular mesh with 262,144 parameters.

Figure 6.9. A slice in depth of the recovered model through inversion using a normal (left) and octree-based mesh (right). Two main elongated features are associated with the large anomaly and three smaller sources are also present. The southern small anomaly is shown not to be connected to the southern larger anomaly, unlike the smoother recovered model from the regular mesh solution. The susceptibilities are similar using both methods.
fectively reducing the \( m \times m \) system of equations needed for inversion. Model ordering using the 3D Hilbert space-filling curve prior to applying the 1D wavelet transform to compress each row of the sensitivity matrix creates a three-step approach towards achieving maximum compression potential. This combination can consistently reduce the storage requirement by 100s to 1,000s times and, therefore also speeds up the computation of the inversion process by the same amount. The field example shows that the recovered model is consistent with, and superior to, the models from traditional inversion. The octree-based model also required only 30,000 model parameters and compressed the sensitivity matrix 354 times compared to 123 via multi-dimensional wavelet transforms alone applied to the regular mesh.

In addition, the field data set also exemplifies that different criteria should be used in different cases for a robust discretization. Large data sets that have anomalies of vastly varying amplitudes should use the curvature of the amplitude normalized by the amplitude as a criterion for discretization. However, if octree inversion is carried out on a subsection of data or data that have limited variability in amplitude, it is appropriate to use only the curvature of the amplitude. Higher compression will be achieved for a larger or more variable data set than a subset of data.
CHAPTER 7
CONCLUSIONS AND RECOMMENDATIONS

In this chapter, I summarize my research and give future recommendations. I will start with the benefits of gradients in processing and inversion. I will then discuss the general algorithm I have developed to accelerate processing and inversion algorithms. Finally, I will give recommendations towards future research opportunities that have arisen from the work.

7.1 Benefits of Gradients in Processing and Inversion

The processing of aeromagnetic data had developed into the use of gradients for enhanced gridding. Processing techniques such as applying Hilbert transforms to observed horizontal gradients (Hansen, 1984; Nabighian, 1984) may then be applied in order to calculate the more accurate vertical gradient or reproduce high signal to noise ratio (SNR) total-field data. Utilizing Hilbert transforms, the Fourier transform was used in order to derive the total field from the leveled and results in gridded gradients (Nelson, 1994). Hardwick and Boustead (2007) created pseudo-lines parallel to observed total-field data. The gradient allows for the calculation of the total-field at a distance away on both sides parallel to the total-field lines. This method traditionally uses minimum curvature gridding, and it has the advantage of utilizing three lines of data versus one to increase the accuracy of the gridded results. A similar approach using the gradient to enhance gridding has also been performed by Reford (2006). Rather than calculating three pseudo-lines of data, he uses the gradients to constrain how the reproduced field changes with position.

Instead of using the gradients as a mathematical tool for, or supplement to gridding, I take advantage of this extra information on its own as an independent data set which can be interpreted in conjunction with total-field data. I apply the
equivalent source processing technique to jointly process both total-field and gradient data sets. This is done through a common model that fits the two kinds of data within the respective error levels, providing a high-quality representation of the observed magnetic anomaly. Furthermore, the data produced from the equivalent sources are de-noised and contain enhanced higher frequency signals. I examined the power spectra of the created data and concluded it has higher resolution bands than just total-field data. The result is more accurate gridded data to perform quantitative interpretation techniques.

The same methodology is extended to 3D inversion. For inversion, the focus is the recovery of a geologically reasonable model that fits the observed data within the noise. The use of total-field and gradient data as two distinct data sets for the response of a single model provides a more compact and sharp model than just the total-field data. Though the L-2 norm is used as a smoothness measure of models, the observed gradients constrain the model laterally and in depth, if the vertical gradient is acquired. As shown by the synthetic dipping slab model, the inverted susceptibility extended to the depth of the true model when the vertical gradient was utilized. The field example is able to distinguish multiple source bodies rather than one and revealed a proposed fault.

7.2 Accelerating Potential-field Inversion Algorithms

Solving a system arising from an inverse formulation (such as with equivalent source processing or 3D inversion) with a dense sensitivity matrix of size \( m \times m \) requires at least \( O(m^3) \) operations. For large data sets, it can become computationally too expensive. Many techniques have been introduced to improve efficiency in order to handle these large data sets. The multi-dimensional wavelet transform has been used to represent this dense matrix sparsely in the wavelet domain. The discrete Fourier transform (DFT) has also been used in inversion in order to speed up the process through the convolution theorem. The DFT approach is quite fast, but has its own
drawbacks such as requiring evenly gridded data on a horizontally flat observation plane and a regular mesh. Both the wavelet and Fourier transforms focus on using basis functions in order to store only the coefficients of those functions to reproduce the dense matrix. Dissecting the dense matrix into multiple smaller sparse matrices that could be multiplied to reproduce the original matrix has also been done. In practice only a portion of the data set is inverted for or one uses larger model cells than should be used. This type of user-defined compression has its obvious drawbacks.

I have developed a three-step approach that can be applied to both inversion-based processing and three-dimensional inversion. The first step is creating a data-adaptive model to reduce the number of parameters prior to the calculation of the sensitivity matrix. The second step is to re-order the model cells using Hilbert space-filling curves. The 1D wavelet transform is then applied to re-ordered sensitivities to achieve compression.

For equivalent source processing, the adaptive quadtree mesh is discretized based on the curvature of the amplitude normalized by the amplitude. This allows for edges of localized anomalies with varying amplitude to have smaller cell sizes. The adaptive quadtree mesh can reduce the number of model parameters by one-third to one-sixth. For 3D inversion, I present a data-adaptive octree mesh discretization. The horizontal discretization follows the same methodology as the quadtree mesh design. For subsets of data that have less variability in amplitude, the curvature of the amplitude is used as the criterion for horizontal discretization. After one of the two criteria is chosen, I then use the depth weighting function to discretize vertically based on the kernel decay of the field. The mesh is ordered after discretization.

The ordering is important to the compression that can be achieved through the 1D wavelet transform. The goal is to have the consecutively indexed model cells spatially continuous, limiting jumps in amplitude of the kernel function into confined clusters of high amplitudes. To efficiently map a 2D or 3D function to one dimension, I utilize space-filling curves, specifically Hilbert curves. The curves originated by mapping an $n$–dimensional space onto 1D and thus are the optimum way to order.
I use the 2D and 3D Hilbert curves for processing and inversion, respectively. The three-dimensional curve starts in the top south-west corner of the mesh and finishes in the top south-east corner. Once the rows of the sensitivity matrix are ordered, the kernels are calculated and the 1D wavelet transform is carried out.

The irregular mesh created by quadtree or octree discretization prohibits the use of multi-dimensional wavelet transforms. However, the 1D wavelet transform can be utilized for each row (i.e. fields of the model cells for one datum) of the sensitivity matrix. The ordering of the model cells create clusters of high amplitude features for each row that ultimately need less wavelet coefficients to keep high reconstruction accuracy. The three-step method compresses the processing and inversion problems by 100s to 1,000s on average.

7.3 Future Research

My recommendations for future research lie in the area of advancing solution speed of potential-field inversion. The research suggestions cover two general topics: improvement of practicalities of the current algorithm and how the approach can be utilized in other applications.

In this thesis, I have outlined a three-step approach for a fast solution of linear inverse problems. The 1D wavelet could be improved upon in two general ways. First is that in practice the threshold is chosen from a single row of the sensitivity matrix, typically in the middle of the mesh as an overly conservative estimate. Rather than one threshold, rapidly computing a threshold for every row would produce large savings while only having to store a vector of \( m \) length for the inverse transform. The other way to improve the compression of the 1D wavelet would be to re-order the data using a Hilbert-curve shape. Mimicking the order of the data with the model would further limit the number of peaks within each row of the sensitivity matrix.

A second research direction that may prove fruitful is focusing on the creation of the sensitivity matrix. For large problems, the calculation of the sensitivity matrix
can use 10s to 100s times the computing time than the actual inversion. The problem lies in the fact that the wavelet transform is applied after the computation of a row and then the smallest coefficients are discarded. In such cases, the formation of the sensitivity matrix takes longer than the actual solution to the inverse problem itself. Research understanding where the coefficients would be small and discarded based on the kernels prior to the calculation of the row could speed up the inversion significantly. This savings is not translated as with the compression ratio, but could be much more substantial.

Finally, adaptive quadtree discretization has additional potential-field applications. Specifically, a quadtree-based mesh could substantially speed up surface-based inversion and forward modeling algorithms. One area of application is terrain corrections for gravity or gravity gradiometry. The threshold for the adaptive mesh discretization would be the assumed density contrast or susceptibility. Kass et al. (2010) have shown that depending on the complexity and relative relief of the terrain model, the adaptive quadtree correction method should reduce the number of model cell calculations by at least a magnitude resulting in significant computational savings.
REFERENCES CITED


Behrens, J., and J. Zimmermann, 2000, Parallelizing an unstructured grid generator with a space-filling curve approach: Presented at the Euro-Par 2000 Parallel Processing, 6th International Euro-Par Conference. (Munich, Germany).


Hansen, R. O., 1984, Two approaches to total field reconstruction from gradiometer data: Expanded Abstracts, 245–247, 54th Annual International Meeting, SEG.


Hardwick, C. D., and G. A. Boustead, 2007, Total field levelling using measured horizontal gradient in place of tie lines: Presented at the High Resolution Geophysics Workshop, SEG.


Kogbetliantz, E. G., 1944, Quantitative interpretation of magnetic and gravitational anomalies: Geophysics, 9, 463–493.


Li, Y., and D. W. Oldenburg, 1996, 3-D inversion of magnetic data: Geophysics, 61, 394–408.

——–, 2000, Joint inversion of surface and three-component borehole magnetic data: Geophysics, 65, 540–552.


Peano, G., 1890, Sur une courbe, qui remplit toute une aire plane: Mathematische Annalen, 36, 157–160.


Ridsdill-Smith, T. A., 2000, The application of the wavelet transform to the processing of aeromagnetic data: PhD thesis, University of Western Australia.


Wright, S. J., 1997, Primal-dual interior-point methods: SIAM.