

**A Cooperative Magnetic Inversion Method With L_p -norm
Regularization**

by

Dominique Fournier

BSc Honours Geophysics, The University of British Columbia, 2012

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL
STUDIES
(Geophysics)

The University of British Columbia
(Vancouver)

October 2015

© Dominique Fournier, 2015

Abstract

In this master's thesis, I implement a Cooperative Magnetic Inversion (CMI) algorithm for the 3-D modeling of magnetic rocks at depth. While in most cases it is assumed that the magnetic response is purely induced, certain rocks have the ability to retain a permanent magnetic moment in any orientations, also known as remanent magnetization. The effect of remanence has long been recognized as an obstacle for the geological interpretation and modeling of magnetic data. My objective is to improve current magnetic inversion methods to recover simpler and better defined magnetization models.

The CMI algorithm brings together three inversion techniques previously introduced in the literature. First, magnetic data are inverted for an equivalent-source layer, which is used to derive magnetic amplitude data. Next, amplitude data are inverted for an effective susceptibility model, providing information about the geometry and distribution of magnetized objects. Finally, the effective susceptibility model is used to constrain the Magnetic Vector Inversion (MVI), recovering the orientation and magnitude of magnetization. All three algorithms are formulated as regularized least-squares problems solved by the Gauss-Newton method.

I further constrain the solution by imposing sparsity constraints on the model and model gradients via an approximated l_p -norm penalty function. I elaborate a Scaled Iterative Re-weighted Least-Squares (S-IRLS) method, allowing for a stable and robust convergence of the algorithm while combining different l_p -norms on the range $0 \leq p \leq 2$. The goal is to reduce the complexity of magnetization models while also imposing geometrical constraints on the solution.

As a final test, I implement the CMI algorithm on an airborne magnetic survey over the Ekaty Property, Northwest Territories. I formulate a tiled inversion scheme

in order to reduce the computational cost and increase the level of parallelization. The final merged magnetization model provides insights into the distribution of dyke swarms and kimberlite pipes. Following the regional inversion, I focus my analysis on sixteen known kimberlite pipes. Magnetization vector inclinations are compared to the expected polarity of rocks inferred from radiometric dating.

Preface

All the work in this thesis is my own. The algorithms and ideas used from other sources are duly cited.

Table of Contents

Abstract	ii
Preface	iv
Table of Contents	v
List of Tables	viii
List of Figures	ix
Acknowledgments	xxi
1 Introduction	1
1.1 Magnetic methods	6
1.2 Thesis outline	8
2 Magnetostatic Problem	10
2.1 Magnetization	14
3 Inverse Problem	16
3.1 Regularized inversion	16
3.1.1 Iterative solver	20
3.1.2 Preconditioner	22
3.1.3 Bound constraints	24
3.2 Magnetic susceptibility inversion	25
3.2.1 Synthetic example	26

3.2.2	Effect of remanent magnetization	31
3.3	Magnetic vector inversion	34
3.4	Magnetic amplitude inversion	38
3.5	Equivalent source method	43
3.5.1	Comment for future research	47
4	Mixed L_p-norm Regularization	49
4.1	Least-norm and regularized least-squares	49
4.2	L_p -norm and iterative re-weighted least squares	53
4.3	IRLS solver	58
4.3.1	Basic IRLS algorithm	58
4.3.2	1-D synthetic example	60
4.3.3	Regularization scaling	65
4.3.4	Threshold parameter ϵ	68
4.4	Scaled-IRLS method (S-IRLS)	74
4.4.1	Cell-based weights (w_r , w_m)	79
4.5	Mixed l_p -norm regularization	82
4.5.1	Localized S-IRLS	82
4.5.2	2-D example	87
4.5.3	3-D example	92
4.6	Case study - Tli Kwi Cho kimberlite complex	95
4.7	Summary	98
5	Cooperative Magnetic Inversion (CMI)	101
5.1	Methodology	102
5.2	CMI with S-IRLS regularization	107
5.3	Conclusion	107
6	Case Study	110
6.1	Ekati Property	111
6.1.1	Geology	111
6.1.2	Airborne magnetic data	113
6.2	Regional scale inversion	114
6.3	Deposit scale inversion	118

6.4	Discussion	122
7	Conclusion	127
7.1	Future work	128
	Bibliography	130

List of Tables

Table 1.1	Magnetic susceptibility for various rock types.	3
Table 3.1	Conjugate Gradient algorithm	22
Table 3.2	Line-search	22
Table 3.3	Inversion parameters.	28
Table 4.1	IRLS Solver 1: Fix parameters	59
Table 4.2	IRLS Solver 2: β -search	65
Table 4.3	IRLS Solver 3: Scaled regularization	68
Table 4.4	IRLS Solver 4: ϵ -Cooling	70
Table 6.1	Dyke swarms of the Lac de Gras region listed in increasing age as published by Buchan et al. (2009)	113
Table 6.2	Regional-scale inversion parameters.	116
Table 6.3	Deposit-scale inversion parameters.	121
Table 6.4	Average magnetization amplitude and direction from the local inversion over 16 known kimberlite pipes. Uncertainties were calculated from a three-cell cube standard deviation around the approximate location of each kimberlite pipe.	121
Table 6.5	Summary of published results over selected kimberlite pipes . .	125
Table 6.6	Radiometric age and inverted magnetization direction for 11 pipes.	126

List of Figures

Figure 1.1	(a) Dipolar magnetic field of the Earth in its normal orientation. (b) Polarity of the field inferred from the geological record over the past 90 Ma (Cande and Kent, 1995).	4
Figure 1.2	Magnetic field data \vec{b} observed on a plane over a magnetized sphere located at the origin. The orientation and magnitude of magnetization of the sphere is function of the magnetic sus- ceptibility κ , the inducing field \vec{B}_0 and remanent components \vec{M}_{NRM} , neglecting self-demagnetization effects.	6
Figure 3.1	(a) Synthetic susceptibility model consisting of a folded anomaly ($\kappa = 0.075$ SI) arching around a discrete block ($\kappa = 0.05$ SI) .	27
Figure 3.2	(a) Data generated from the synthetic susceptibility model sub- ject to a vertical 50,000 nT inducing field. (b) Data are then corrupted with (c) random Gaussian noise, 1 nT standard devi- ation.	27
Figure 3.3	Convergence curve showing the data misfit $\phi_d^{(k)}$ and model norm $\phi_m^{(k)}$ as a function of β iterations. The inversion achieves target misfit after the 7 th iteration, which in this case also corre- sponds to the point of maximum curvature on the misfit curve. Attempting to further lower the data residual comes at the risk of fitting some of the Gaussian noise.	28

Figure 3.4	(a) Iso-surface (0.002 SI) and (b) sections through the recovered susceptibility model for a purely induced response. The model is smooth but recovers the arc and block anomaly at roughly the right depth.	30
Figure 3.5	Comparison between (a) observed and (b) predicted data from the recovered susceptibility model. (c) The normalized data residuals appear to be correlated with the location of the magnetic body.	30
Figure 3.6	Perspective view and sections through the synthetic magnetization model. The arc-shaped anomaly is magnetized at 45° from horizontal and with variable declinations directions between [−45°N ; 45°N].	32
Figure 3.7	(a) Data generated from the synthetic magnetization model. (b) Observed data are corrupted with (c) random Gaussian noise, 1 nT standard deviation.	32
Figure 3.8	(a) Iso-surface (0.002 SI) and (b) sections through the recovered susceptibility model assuming no remanence. The arc-shaped anomaly is poorly recovered and magnetic susceptibilities are pushed at depth and outwards.	33
Figure 3.9	Comparison between (a) observed and (b) predicted data from the recovered susceptibility model assuming a purely induced response. (c) The inversion has a harder time fitting the large negative fields along the arc.	33
Figure 3.10	(a) Iso-surface ($\kappa_e = 0.001$) and (b) sections through the recovered magnetization model from the MVI method. The inversion recovers the true orientation of magnetization inside the block, but the thin arc is poorly resolved.	37
Figure 3.11	Comparison between (a) observed and (b) predicted data from the recovered magnetization model from the MVI method. The model can replicate the data at the same level achieved by the purely induced problem.	37

Figure 3.12 (a) Iso-surface (0.002 SI) and (b) sections through the recovered effective susceptibility model. The arc-shaped and block anomalies are recovered at the right location, but smoothly stretched vertically.	42
Figure 3.13 Comparison between (a) observed and (b) predicted data from the recovered effective susceptibility model. The inversion can predict most of the data within one standard deviation.	42
Figure 3.14 (a) Synthetic model consisting of 200 unit cubes of susceptible material in a non-susceptible background. Data are generated on a plane one unit above the source location, assuming a purely vertical inducing field. Various components of the fields are shown in figure (b) to (f).	45
Figure 3.15 (a) Recovered equivalent source layer from TMI data using a positivity constraint. The residuals between observed and predicted data are shown in figure (b) to (f) for various components of the field. Each component is well recovered within the noise level.	46
Figure 3.16 (a) Recovered equivalent source layer from TMI data after removing a portion of data over the corner of the magnetic anomaly. The residuals between observed and predicted data are shown in figure (b) to (f) for various components of the field. Note the large correlated artifacts recovered on the \mathbf{b}_x , \mathbf{b}_y and $ \mathbf{b} $ components.	48

Figure 4.1	Comparative contour maps for various objective functions over a range of model values $[m_1; m_2]$. (a) The minimum of the misfit function ϕ_d forms a line spanned by $\mathcal{N}(\mathbf{F})$ (red dash). The <i>least – norm</i> solution is marked as a solid dot. (Middle) Regularization functions and gradient directions (blue arrows) for approximated l_p -norm measures of the model for (b) $p = 2$, (d) $p = 1$ and (f) $p = 0$. The gradient directions are shown for two different starting models (black arrows). (bottom) Contour maps of the initial objective functions $\phi(m) = \phi_d + \phi_m$ for the same two starting models: $\mathbf{m}_1^{(0)}$ (solid) and $\mathbf{m}_2^{(0)}$ (dash). (c) In the l_2 -norm case , the function has a global minimum regardless of the starting model, while for non-linear functions for (e) $p = 1$ and (g) $p = 0$, the objective function changes with respect to the starting model.	52
Figure 4.2	Contour maps for various objective functions after convergence of the IRLS algorithm. (a) Final model obtained with the l_2 -norm penalty on the model for two starting models at $\mathbf{m}_1^{(0)} = [0.2; 0.4]$ and $\mathbf{m}_2^{(0)} = [0.6; 0.2]$ for a fixed trade-off parameter ($\beta = 1e - 4$). In both cases, the solution converges to the global minimum, which is also the <i>least – norm</i> solution at $\mathbf{m}_{\text{ln}} = [0.2; 0.4]$. (b) Solution with the l_1 -norm penalty for the same starting models and trade-off parameter, converging to a global minimum at $\mathbf{m}^* = [0; 0.5]$. This solution is sparse and can reproduce the data. (c) The same experiment is repeated for the l_0 -norm penalty, converging to two different solutions depending on the relative magnitude of the starting model parameters. Both solutions are sparse and honor the data.	53

Figure 4.3	(a) Penalty function $\rho(x)$ for different approximate l_p -norm measures, and enlarged view near the region of influence of ε , making the l_p -norm continuous at the origin. (b) IRLS weights $\mathbf{r}(x)$ as a function of model function $\mathbf{x}(m)$ and p -values and for a fix threshold parameter ($\varepsilon = 1e - 2$). (c) Gradients $\mathbf{g}(x)$ of the model penalty function for various p -values. Note that the gradients are on a logarithmic scale due to the rapid increase as $x_i \rightarrow \varepsilon$ for $p < 1$	57
Figure 4.4	(a) Synthetic 1D model made up of a rectangular pulse and a Gaussian function. (b) Kernel functions consisting of exponentially decaying cosin functions of the form $f_j(x) = e^{-jx} \cos(2\pi jx)$. (c) Data generated from $\mathbf{d} = \mathbf{F} \mathbf{m}$, with 5% random Gaussian noise added.	61
Figure 4.5	(a) Recovered models from the smooth l_2 -norm regularization, and (bottom) measure of model misfit and model norm as a function of iterations. Both the rectangular pulse and Gaussian function are recovered at the right location along the x -axis, but the solution is smooth and dispersed over the entire model domain. (b) Solution obtained from the IRLS with sparsity constraints on the model gradients ($q=0$), using the l_2 -norm model (a) to initiate the IRLS steps. The algorithm uses a fixed threshold parameter ($\varepsilon = 1e - 8$) and fixed trade-off parameter β . The final solution is blocky, as expected from the norm chosen, but fails to achieve the target data misfit. Clearly the influence of the regularization function has overtaken the minimization process.	64

Figure 4.6	(Top) Recovered models from two different algorithms used to implement the IRLS method for $q = 0$ and a fix threshold parameter ($\varepsilon = 1e - 8$). (Bottom) Measure of model misfit and model norm as a function of iterations. (a) The first algorithm searches for an optimal trade-off parameter $\beta^{(k)}$ between each IRLS step, requiring a solution to multiple sub-inversions. (b) The second algorithm only adjusts $\beta^{(k)}$ once after each IRLS step. A new scaling parameter $\gamma^{(k)}$ is added to smooth the transition between the IRLS updates. The algorithm recovers a similar blocky model but is computationally cheaper, as indicated by the total number of beta iterations and CG solves.	67
Figure 4.7	(Top) Recovered models from two different algorithms used to implement the IRLS method for $q = 0$ with cooling of the threshold parameter ε . (Bottom) Measure of model misfit and model norm as a function of iterations. Both algorithms adjust β and scaling parameter $\gamma^{(k)}$ between each IRLS iteration. (a) In the first case, the threshold parameter ε is monotonically reduced until reaching the convergence criteria. The solution is sparser than the previous algorithm, even though ε is much larger than machine error. (b) In the second case, ε is decreased until reaching the target threshold ε^* . The solution is blocky, penalizing small model gradients.	71
Figure 4.8	(a) Distribution of model parameters and (b) model gradients recovered from the smooth l_2 -norm regularization. Both curves show a sharp corner around which the model functions vary rapidly. Similarly, a measure of (c) model norm s_{MS} and (d) model gradient norm s_{MGS} can be computed over a range of ε values, yielding a similar $L - curve$. The point of maximum curvature can be used to determine the optimal <i>effective zero</i> parameter ε_p and ε_q	73

Figure 4.9	(a) (Top) Recovered model using a mixed-norm regularization for $p = 0, q = 2, \varepsilon = 1e - 3$. (Bottom) Measure of model misfit and model norm as a function of iterations. The inversion converges to a sparse solution without apparent penalty on the gradient, indicative of an imbalance in the regularization. (b) Recovered model and convergence curves after rescaling of the regularization, yielding a model that is both sparse and smooth as expected from the applied mixed-norm	75
Figure 4.10	Contour maps for (a) the misfit function ϕ_d , (b) the model norm ϕ_s and (c) the norm of model gradients ϕ_x . (d) The total objective function $\phi(m)$ has a global minimum located at $\mathbf{m} = (0.5, 1.0)$ for a given small trade-off parameter ($\beta = 1e - 3$). The direction of update is shown for two starting models $\mathbf{m}^{(0)}$ (black and white dot). (e) Section through the objective function along the minimum of ϕ_d . The global minimum occurs where the partial gradients of $\frac{\partial \phi_s}{\partial m_1}$ and $\frac{\partial \phi_x}{\partial m_1}$ have equal and opposite signs.	77
Figure 4.11	(a) Partial gradients of approximated l_p -norm penalties for a fix stabilizing parameter $\varepsilon = 1e - 2$. Gradients for $p < 1$ are consistently larger on the interval $[0 < x_i < \sqrt{1 - \varepsilon^2}]$, making it hard to combine multiple norm penalties within the same objective function. (b) Function gradients after applying a scale of $\varepsilon^{(1-p/2)}$, forcing each l_p -norm to intersect at $m = \sqrt{\varepsilon}$	78
Figure 4.12	Recovered models for two different depth weighting formulations: (red) weighted sensitivity $\phi(\hat{m})$, (black) weighted regularization $\phi(m)$. (a) True and recovered models using the $\phi(\hat{m})$ and $\phi(m)$ formulations for penalty applied on the model gradients for $q = 0$ and (b) for $p = 0, q = 2$. The weighted sensitivity formulation $\phi(\hat{m})$ increases the influence the regularization function with distance along the x -axis, skewing the model towards the right.	81

Figure 4.13	(a) Model error $\ m - m^*\ _1$ and (b) misfit function for the 441 inverted models using a range of regularization with mixed-norm penalty on the model for $0 \leq p \leq 2$ and on model gradients for $0 \leq q \leq 2$. (c) The largest model error ($\ \delta\mathbf{m}\ _1$) was obtained with the mixed-norm for $p = 0, q = 2$, compared to (d) the optimal solution found with $p = 1.5$ and $q = 0.4$.	83
Figure 4.14	(a) Nine of the 441 inverted models for a range of mixed-norm penalties on the model and its gradient for $0 \leq p \leq 2$ and $0 \leq q \leq 2$.	84
Figure 4.15	(Left) Improved solution for the 1-D problem after applying a localized mixed-norm penalty, where the regularization is divided into two regions with independent l_p -norm regularization: (left) $p = q = 0$, (right) $p = 1, q = 2$. (Right) Convergence curves for the mixed-norm S-IRLS inversion.	86
Figure 4.16	(a) Synthetic 2-D model made up of a square block and a smooth Gaussian function. (b) Example of a kernel function for $e^{-\omega r} \cdot \cos(2\pi\omega r)$ and (c) data generated from $\mathbf{d} = \mathbf{F} \mathbf{m}$. Five percent random Gaussian noise is added.	88
Figure 4.17	(a) Recovered model for $p = q_x = q_z = 1$ penalizing finite difference gradients in orthogonal directions, yielding right-angled anomalies. (b) Recovered model for the same norms but penalizing the absolute gradient of the model ($ \nabla\mathbf{m} $) recovering round edges.	90
Figure 4.18	Distribution of l_p -norm on the model and model gradients over the 2-D model domain. The original boundary of each region was smoothed in order to get a slow transition and reduce visible artifacts. Regions were chosen to cover a larger area than the anomalies to simulate a blind-inversion.	90

Figure 4.19	(a) Smooth l_2 -norm solution used to initiate the IRLS iterations. (b) Recovered model using the mixed-norm regularization after seven S-IRLS iterations. The contour line (red) marks the value of ε_p , judged to be the <i>effective zero</i> value of the model ($m_i \leq 5e-2$). Both models (a) and (b) fit the data within 2% of the target misfit ϕ_d^* . (c) Dual plots showing the distribution of model parameters and the gradient of the l_0 -norm penalty function $\hat{g}_p(m)$ as a function of S-IRLS iterations. High penalties are applied to progressively smaller model values. The final model nicely recovers both the blocky and smooth Gaussian anomaly.	91
Figure 4.20	(a) Iso-surface (0.002 SI) and (b) sections through the recovered susceptibility model after five IRLS iterations for ($p = 0, q = 2$). The final model is substantially closer to the true solution.	93
Figure 4.21	Comparison between (a) observed and (b) predicted data from the recovered susceptibility model using compact norms for ($p = 0, q = 2$). (c) Normalized data residuals are within two standard deviations.	93
Figure 4.22	(a) Iso-surface (0.002 SI) and (b) sections through the recovered susceptibility model after nine IRLS iterations for ($p = 0, q = 1$). Sparsity constraints on the model and model gradients yield a simple and blocky model.	94
Figure 4.23	Comparison between (a) observed and (b) predicted data from the recovered susceptibility model using compact norms for ($p = 0, q = 1$). (c) Normalized data residuals are within two standard deviations.	94

Figure 4.24	(a) (Left) Horizontal section through the recovered susceptibility model at 25 m depth below topography from the smooth l_2 -norm regularization. (Right) Iso-surface of susceptibility values around 0.002 SI. (b) Recovered model using the mixed-norm S-IRLS algorithm. Magnetic dykes are better recovered, imaged as continuous plates and extending vertically at depth. Susceptibility values for DO-27 and DO-18 have increased, showing as compact vertical pipes.	97
Figure 4.25	Horizontal section through the mixed-norm models applied to four sub-regions with smooth transition across zones.	99
Figure 4.26	(a) Observed and predicted data over the TKC kimberlite complex. (b) Residuals between observed and predicted data normalized by the estimated uncertainties (10 nT). Both the smooth and mixed-norm inversions reproduce the data within four standard deviations.	100
Figure 5.1	Schematic representation of the Cooperative Magnetic Inversion (CMI) algorithm. Input and output parameters are indicated by a dash arrow.	103
Figure 5.2	(a) Inverted equivalent-source layer and (b-f) predicted \hat{x} , \hat{y} , \hat{z} -component, TMI and magnetic amplitude data for the synthetic model.	104
Figure 5.3	(a) Iso-surface (0.002 SI) and (b) sections through the recovered effective susceptibility model. This effective susceptibility model is used to construct a weighting matrix to constrain the MVI.	105
Figure 5.4	Comparison between (a) observed and (b) predicted data from the recovered effective susceptibility model. The inversion can predict most of the data within one standard deviation.	105
Figure 5.5	(a) Iso-surface (0.005 SI) and (b) sections through the recovered magnetization model from the CMI algorithm ($p = 2$, $q = 2$). The inversion recovers both the arc and the block anomaly as distinct objects.	106

Figure 5.6	Comparison between (a) observed and (b) predicted data from the recovered susceptibility model. (c) Normalized data residuals are within two standard deviations.	106
Figure 5.7	(a) Iso-surface (0.01 SI) and (b) sections through the recovered effective susceptibility model from the amplitude inversion with sparsity constraint applied ($p = 0, q = 1$). The l_p -norm constraint considerably reduces the complexity of the model, although the model is still stretched vertically.	108
Figure 5.8	Comparison between (a) observed and (b) predicted amplitude data from the recovered compact effective susceptibility model ($p = 0, q = 1$). (c) Normalized data residuals are within two standard deviations.	108
Figure 5.9	(a) Iso-surface (0.01 SI) and (b) sections through the recovered magnetization model from the CMI algorithm. Compact norms ($p = 0, q = 2$) were applied during the amplitude inversion. The l_p -norm constraint considerably reduces the complexity of the model.	109
Figure 5.10	Comparison between (a) observed and (b) predicted data from the recovered magnetization CMI model. (c) Normalized data residuals are within two standard deviations. Correlated residuals are no longer seen.	109
Figure 6.1	Topography and known kimberlites over the Lac de Gras region. The location of the main operations for the Ekati and Diavik mines are shown (red dot).	112
Figure 6.2	Topography, aeromagnetic survey and known kimberlites over the Lac de Gras region. The location of the main operations for the Ekati and Diavik mines is marked with a red dot.	114
Figure 6.3	Regional data (a) before and (c) after removal of a regional trend. The median value within selected regions (boxes) were used to compute (b) a first-order polynomial trend.	115
Figure 6.4	TMA data after removal of the regional signal and tiling configuration used for the regional inversion.	117

Figure 6.5	Interpreted dykes (line) from the property-scale magnetization inversion and known kimberlite pipes location (circle). The 16 pipes chosen for a deposit-scale inversion are labeled.	118
Figure 6.6	Comparison between (a) observed and (b) merged predicted data from the recovered magnetization model over the Ekati Property. (c) Normalized residual data show horizontal striations due to variations in magnetic data between adjacent lines, which could not be accounted for by the inversion. Most of the predicted data from the individual tiles can fit the observed data within one standard deviation. (d) Forward modeled data from the merged magnetization model and (c) normalized data residual. It appears that a large portion of the low frequency content has been lost during the merging step. More research is required in order to preserve the long wavelength information.	119
Figure 6.7	(a) Local data collected over the Misery pipe showing a local western trend. (b) Regional field data are computed from a local inversion. Most of the signal comes from a dyke running north-south along the western edge of the local tile. (c) Residual data after regional field removal, showing a clear reversely magnetized anomalous corresponding with the location of the Misery pipe.	122
Figure 6.8	Horizontal sections through the recovered (a) induced and (b) remanent magnetization model. Several pipes with strong remanence can easily be identified as discrete circular anomalies.	124
Figure 6.9	Age of 11 pipes from the Ekati region with respect to Earth's polarity reversal.	126

Acknowledgments

First of all I need to thank my supervisor Dr. Doug Oldenburg for his mentorship and his contagious passion for research. It has been an honor and privilege to learn from a pillar in Earth sciences. Special acknowledgment to Dr. Kristofer Davis for his support, beginning four years ago with my undergraduate honors thesis.

Thanks to all the UBC-GIF members for the interesting conversations and friendship. Sincere gratitude for the collaboration and mentorship that I have received over the years from several senior geoscientists: Peter Kowalczyk, Nigel Phillips and all the staff at Mira Geoscience, Dr. Peter Fullagar at Fullagar Geophysics, Dr. Randy Enkin at the GSC and Stéphane Lorrain at Environnement Illimité. Special thanks to Jon Carlson at Dominion Diamond Ekati Corporation, for granting me access to the rock physical property database. Also to Jennifer Pell and Brooke Clements at Peregrine Diamonds Ltd for providing the TKC dataset.

Last but not least, extra special thanks goes to my editor-in-chief, part-time lover and full-time best friend Chantelle Bellrichard.

Chapter 1

Introduction

Magnetic methods have a long and rich history in Earth sciences (Nabighian et al., 2005). As early as the 19th century it was understood that particular rocks were the source of local magnetic anomalies measurable at the surface. Since the development of the fluxgate magnetometer during World War II, magnetic surveys have contributed greatly to what we know about the earth. At a global scale, marine surveys over magnetically polarized oceanic plates have shaped our understanding of continental drift and plate tectonics (Vine and Matthews, 1963). By the 1990s, Global Positioning System (GPS) and increasingly sensitive instruments led to a proliferation of airborne surveys that gave us the ability to map geological features remotely. Later advancements in numerical computing contributed to the adoption of 3-D magnetic inversion as a tool to image magnetic bodies at depth. This master's thesis focuses on improving current inversion methods using a particular strategy. I explore how various inversion codes can be used cooperatively, yielding a more robust and flexible imaging tool in exploration geophysics.

But first, a review of the magnetic properties of rocks. Assuming no free currents and steady state, Maxwell's equations can be written as:

$$\nabla \cdot \vec{B} = 0 \quad (1.1)$$

$$\nabla \times \vec{H} = 0 \quad (1.2)$$

$$\vec{B} = \mu \vec{H}, \quad (1.3)$$

where \vec{B} is the magnetic flux density in Tesla (T) and \vec{H} is the magnetic field in units of amperes per meter (A/m). In matter, the magnetic permeability μ relates the magnetic flux density \vec{B} to the field \vec{H} such that

$$\mu = \mu_0(1 + \kappa), \quad (1.4)$$

where κ is the magnetic susceptibility, a dimensionless positive number describing the ability of certain material to become magnetized under an applied field. In free space, the term magnetic field is often used interchangeably to describe \vec{B} and \vec{H} since linearly related by the magnetic permeability of free space μ_0 ($4\pi \times 10^{-7}$ T·m/A).

Factors responsible for rock magnetism can be divided into an *induced* and a *remanent* component such that:

$$\vec{M} = \kappa \vec{H} + \vec{M}_{NRM} \quad (1.5)$$

$$\vec{H} = \frac{1}{\mu_0}(\vec{B}_0 + \vec{B}_A), \quad (1.6)$$

where \vec{M} (A/m) is the magnetization per unit volume of a rock, the quantity of interest in mineral exploration. The inducing field can be further decomposed in the primary geomagnetic flux \vec{B}_0 and anomalous *local* flux \vec{B}_A . The Natural Remanent Magnetization (\vec{M}_{NRM}) describes the ability of matter to retain a net magnetization component in the absence of an inducing field.

Generated within the Earth's core (Campbell, 1997), the inducing field \vec{B}_0 varies between 30,000 nT at the equator, to over 50,000 nT near the poles as illustrated in Figure 1.1. From the geochronological record, we also know that the Earth's field went through at least 92 magnetic field reversals over the last 100 million years. The current orientation of the field, pointing down towards the North pole, is defined as the *normal* polarity. In most cases, the induced component from the primary field \vec{B}_0 is the main driver for the magnetic response of rocks. Secondary fields \vec{B}_A are generally much weaker than the primary field and arise from the interaction of neighbouring magnetic objects. Secondary fields may be responsible for *self-demagnetization* effects as studied by Lelièvre (2003). In the majority of cases, the strength of \vec{B}_0 largely dominates any other secondary

Table 1.1: Magnetic susceptibility for various rock types.

Rock Type	Magnetic Susceptibility $\kappa \times 10^{-6} SI$
Granite	20 - 40,000
Slates	0 - 1,200
Basalt	500 - 80,000
Oceanic basalts	300 - 36,000
Limestone (with magnetite)	10 - 25,000
Gneiss	0 - 3,000
Sandstone	35 - 950
Hematite (ore)	420 - 10,000
Magnetite (ore)	$7 \times 10^4 - 1.4 \times 10^7$
Magnetite (crystal)	1.5×10^8

fields, forcing the magnetization to be aligned parallel to the geomagnetic field. As a rule of thumb, secondary fields only become important for rocks with magnetic susceptibility $\kappa > 1$. Table 1.1 summarizes the magnetic susceptibility of major rock types, varying over several orders of magnitude. Magnetite is by far the most abundant magnetic mineral, followed by pyrrhotite and ilmenite (Clark, 1991). All those minerals are commonly associated with iron oxides, and potentially good indicators for mineral deposits.

Natural Remanent Magnetization \vec{M}_{NRM} is a permanent magnetization moment preserved in the absence of an inducing field. The reader is encouraged to refer to Blakely (1996) and Clark (1991) for a more in-depth explanation of chemical, thermal and biological processes responsible for the remanent component of various rock types. While the induced component is linearly related to the inducing field, nothing can be assumed about the NRM orientation and it remains challenging to estimate. If aligned with the inducing field direction, the NRM is indistinguishable from a purely induced response, resulting in an over-estimation of the magnetic susceptibility of rocks. For large NRM components, perpendicular or anti-parallel to the inducing field, the magnetic response of compact objects may get distorted, potentially resulting in false interpretation about the distribution and geometry of rock units. Direct geological interpretation of magnetic data that ignores the effect

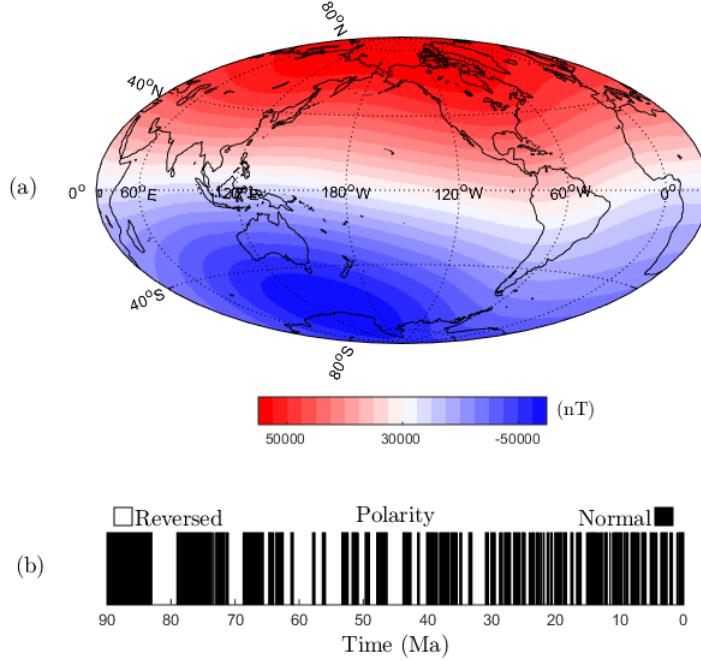


Figure 1.1: (a) Dipolar magnetic field of the Earth in its normal orientation. (b) Polarity of the field inferred from the geological record over the past 90 Ma (Cande and Kent, 1995).

of remanence has long been recognized as problem in mineral exploration. This brings up an important question: How can we better recover the location and geometry of magnetized objects without knowledge of the total magnetization direction \vec{M} ?

From Gauss's law, the relation between the observed magnetic field and rock magnetization is expressed as:

$$\vec{b}(r) = \frac{\mu_0}{4\pi} \int_V \nabla \nabla \frac{1}{r} \cdot \vec{M} dV , \quad (1.7)$$

where \vec{b} is the magnetic field (T) as measured at some distance r from a magnetic anomaly with magnetization per unit volume \vec{M} (A/m). Since \vec{b} is usually small, it is commonly measured in units of nano-Tesla (nT). The majority of magnetic data

consist of Total Magnetic Intensity (TMI) measurements which can be written as:

$$b^{TMI} = |\vec{B}_0 + \vec{B}_A| , \quad (1.8)$$

where we measure the magnitude of the field rather than the individual components. In most cases we are only interested in the anomalous local fields. Under the assumption that $|\vec{B}_A| \ll |\vec{B}_0|$, the anomalous field is approximated to be parallel with the direction of the inducing field $\hat{\vec{B}}_0$. The Total Magnetic field Anomaly (TMA) is given by:

$$\begin{aligned} b^{TMA} &= |\vec{B}_0 + \vec{B}_A| - |\vec{B}_0| \\ &\simeq \vec{B}_A \cdot \hat{\vec{B}}_0 . \end{aligned} \quad (1.9)$$

Figure 1.2 gives an example of TMA data measured on a plane above a magnetized sphere. The main challenge in interpreting magnetic data is to characterize the magnetic sources when nothing is known about the location and magnetic properties (κ, \vec{M}) of buried objects. Accurately imaging the location and geometry of magnetic objects is a core problem in exploration geophysics.

Early geophysical studies relied primarily on filtering techniques to infer geological structures and identify potential targets (Cowan et al., 2000). In order to reduce the complexity of the problem, the magnetization direction is in most cases assumed to be purely induced along \vec{B}_0 , neglecting self-demagnetization and remanence effects. Inversion codes, such as the UBC-MAG3D from Li and Oldenburg (1996), also rely on this assumption. While the induced component may dominate in most cases, recent petrophysical studies seem to indicate that the effect of remanent magnetization may be more important than previously thought (Enkin, 2014). This is especially true for specific types of mineral deposits such as Banded Iron-Formations (BIF) and diamondiferous kimberlite (Dransfield et al., 2003; Li et al., 2010). Strong remanent magnetic components can hinder the interpretation of magnetic data thus leading to false drilling targets.

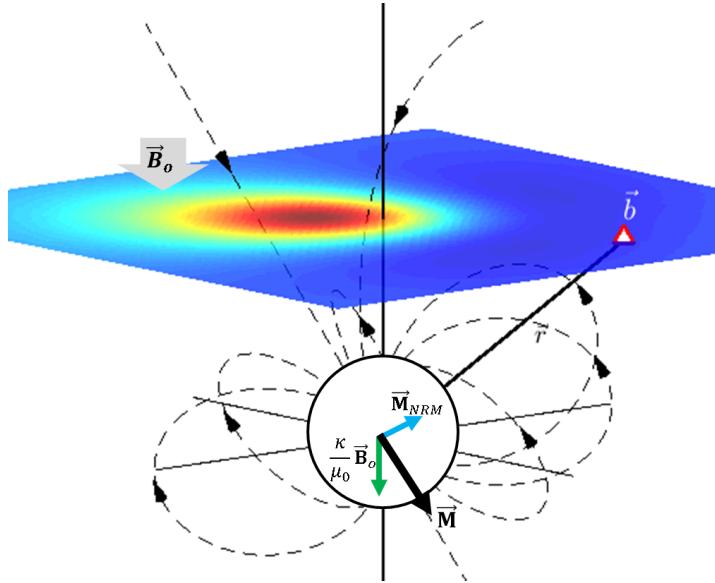


Figure 1.2: Magnetic field data \vec{b} observed on a plane over a magnetized sphere located at the origin. The orientation and magnitude of magnetization of the sphere is function of the magnetic susceptibility κ , the inducing field \vec{B}_0 and remanent components \vec{M}_{NRM} , neglecting self-demagnetization effects.

1.1 Magnetic methods

Several studies have been dedicated to the inversion of magnetic data in the presence of remanence. As summarized by Li (2012), the proposed inverse methods can be divided into three categories. The first category estimates the magnetization direction from the data in a pre-processing step. Strategies such as the Helbig's method (Phillips, 2003) and the cross-correlation (Dannemiller and Li, 2006) are used upstream of standard inversion codes. Adapted from the 2-D analytical signal of Nabighian (1972), Roest et al. (1992) estimate the orientation of remanence in 3-D. In most cases, data are transformed to the wavenumber domain in order to extract magnetic field components. These methods are simple and robust for simple and isolated anomalies, but become impractical when the data are acquired over rough terrain or complicated geology.

The second category deals with magnetic data that are weakly dependent on the orientation of magnetization. It has been shown that magnetic amplitude data

are weakly dependent on the magnetization direction in 3-D (Nabighian, 1972). Magnetic amplitude data are simply calculated as:

$$|\vec{b}| = (b_x^2 + b_y^2 + b_z^2)^{1/2}, \quad (1.10)$$

where b_x , b_y and b_z are the three components of the anomalous magnetic field. Inverting magnetic amplitude data can provide a robust estimate for the location of magnetized bodies as shown by Shearer (2005). Magnetic amplitude data are successfully inverted over a kimberlite deposit, recovering the true location of kimberlite pipes and associated magnetic dykes. Because most magnetic surveys only record TMI data, the three components of the field must be derived in post-processing. Two strategies have been proposed in the literature to extract magnetic field components. The first relies on the fact that magnetic data are potential field data, hence knowledge of any component of the field on an infinite plane above the source is sufficient to calculate the remaining components. Data are interpolated onto a uniform grid, which is then used to calculate the components of the field using Fourier transform techniques. This may be a problem however for air-borne surveys over steep topography or acquired at low latitude. Alternatively, the equivalent-source method has been proposed by Li and Oldenburg (2010) to alleviate those constraints. Taking advantage of the non-uniqueness of potential field data, TMA data are inverted for a layer of magnetic sources then used to forward model three-component magnetic data.

The third method directly solves for the orientation of magnetization without making any assumptions about the location or geometry of causative bodies. In the Magnetic Vector Inversion (MVI) proposed by Lelièvre and Oldenburg (2009), the magnetization vector is decomposed in its induced and orthogonal components. The MVI is closely related to the method proposed by Kubota and Uchiyama (2005), and later borrowed by Ellis et al. (2012). This method results in a large under-determined inverse problem, with roughly three times the number of free-variables over conventional susceptibility inversion codes. Recovered magnetization models are generally smooth and become overly complicated for direct geological interpretation. As pointed out by Lelièvre (2009), any *a priori* information from surface or borehole measurements may greatly reduce the non-uniqueness of

the problem. Unfortunately this kind of information is rarely available in greenfield settings or is only available in sparse samples.

Building upon the work of Shearer (2005), Liu et al. (2015) propose a hybrid three-step process to invert for the location, orientation and magnetic susceptibility distribution in 2-D. In the first step, the location of magnetic material is inverted from amplitude data. Next, the orientation of magnetization is found via a correlation method. The orientation of magnetization is then used in a standard inversion code to recover a susceptibility model. The combined method greatly reduces ambiguity about the location and orientation of isolated targets, but becomes challenging over multiple anomalies with variable magnetization directions. The same synthetic model was applied to the MVI method, resulting in a poor recovery for the location and intensity of magnetization. This result is surprising considering the satisfactory solution obtained in previous studies (Ellis et al., 2012; Lelièvre and Oldenburg, 2009). The combined approach of Liu et al. (2015) is interesting however, as it links together complementary algorithms.

1.2 Thesis outline

Following the same line of thought as Liu et al. (2015), I propose a Cooperative Magnetic Inversion (CMI) algorithm that directly incorporates the amplitude inversion of Shearer (2005) into the MVI algorithm of Lelièvre and Oldenburg (2009). Magnetic field amplitude data are computed from the Equivalent Source technique proposed by Li and Oldenburg (2010), removing the need to grid the data. Moreover, I introduce a Scaled Iterative Re-weighted Least Squares (S-IRLS) regularization method to recover blocky and sparse solutions. My thesis is divided into the following chapters:

In Chapter 2 and 3, I provide further details about rock magnetism and review the theory related to inverse problems in the context of exploration geophysics. I revisit three inversion codes from the literature: the magnetic susceptibility inversion from (Li and Oldenburg, 1996), the magnetic amplitude inversion from Shearer (2005), and the Magnetic Vector Inversion (MVI) from (Lelièvre, 2009). Each of these codes are tested on a synthetic example with complicated magnetization distribution. I also review the equivalent-source method from Li and Oldenburg

(2010).

In Chapter 4, I introduce a mixed l_p -norm regularization function for the recovery of compact and sparse solutions, increasing the flexibility of current inversion codes. The method uses a Scaled Iterative Re-weighted Least Squares (S-IRLS) formulation to approximate any l_p -norm penalties on the interval $0 \leq p \leq 2$. Sparse constraints are applied on both the model and model gradients independently. The mixed-norm regularization is implemented on the magnetic problem, and tested on an airborne magnetic dataset from the Tli Kwi Cho kimberlite complex, Northwest Territories.

In Chapter 5, I formulate a Cooperative Magnetic Inversion (CMI) method, combining the inversion codes presented in Chapter 3 and 4. The algorithm is tested on the same synthetic example for comparison.

Finally in Chapter 6, I apply the CMI algorithm to a large aeromagnetic data set over the Ekati Property, Northwest Territories. I design a tiled inversion scheme to automate the inversion process and to reduce the computational cost. Individual models are then merged back onto a global mesh for analysis. A bulk estimate of magnetization direction is compared to values published in the literature. The polarities of magnetization over 11 pipes are compared to the estimated age of emplacement. The analysis is extended to various dyke swarms, providing regional geologic information.

The work presented in this thesis is significant for two reasons. First, the mixed-norm regularization function introduced in Chapter 4 can considerably improve the flexibility of inversion algorithms, not limited to potential field problems. The S-IRLS method allows for a combination of sparse norms on the model and model gradients independently, granting access to a wider range of solutions than previously offered by globally convex functions.

Secondly, it is the first time that both the amplitude inversion and MVI algorithms are combined in a cooperative inversion method in 3-D. Structural information gained by the amplitude inversion is used directly to constrain the MVI algorithm, reducing the non-uniqueness of the solution. From a practical standpoint, it is an automated process that streamlines the inversion workflow. This, in turn, can reduce the overall time required to process magnetic data and promote the expansion of inversion methods in mineral exploration.

Chapter 2

Magnetostatic Problem

The general theory describing the magnetic static problem can be derived from Maxwell's equations. Assuming no free currents and no time varying electric field, the equations describing the electromagnetic field simplify to:

$$\nabla \cdot \vec{B} = 0 \quad (2.1)$$

$$\nabla \times \vec{H} = 0 \quad (2.2)$$

$$\vec{B} = \mu \vec{H}, \quad (2.3)$$

where \vec{B} is the magnetic flux density in Tesla (T) and \vec{H} is the magnetic field in units of amperes per meter (A/m). In matter, the magnetic permeability μ relates the magnetic flux density \vec{B} to the field \vec{H} such that:

$$\mu = \mu_0(1 + \kappa), \quad (2.4)$$

where μ_0 is the magnetic permeability of free space. The magnetic susceptibility κ is a dimensionless positive number describing the ability of certain material to become magnetized under an applied field. From 2.2, the magnetic field can be formulated via a potential field formulation such that:

$$\vec{H} = \nabla \phi, \quad (2.5)$$

Because Maxwell's equations forbid magnetic monopoles from 2.1, we approximate the potential field ϕ by a magnetic dipole. The potential from a dipole moment \vec{m} located at some location r_Q as observed at location r_P is given by:

$$\phi(r) = \frac{1}{4\pi} \vec{m} \cdot \nabla \left(\frac{1}{r} \right) \quad (2.6)$$

$$\begin{aligned} r &= |r_Q - r_P| \\ &= \sqrt{(x_P - x_Q)^2 + (y_P - y_Q)^2 + (z_P - z_Q)^2} \\ \vec{m} &= [m_x, m_y, m_z] . \end{aligned}$$

Going from a discrete dipole moment to continuous magnetization allows us to write:

$$\phi(r) = \frac{1}{4\pi} \int_V \vec{M} \cdot \nabla \left(\frac{1}{r} \right) dV \quad (2.7)$$

$$\vec{M} = \begin{bmatrix} M_x \\ M_y \\ M_z \end{bmatrix} ,$$

where \vec{M} is the magnetization vector per unit volume in A/m. Taking the gradient of 2.7, the magnetic flux density can be expressed as:

$$\vec{b}(P) = \frac{\mu_0}{4\pi} \int_V \vec{M} \cdot \nabla \nabla \left(\frac{1}{r} \right) dV \quad (2.8)$$

$$\vec{b}(P) = \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix} ,$$

where $\vec{b}(P)$ is the magnetic vector measurement at location P .

In a geophysical context, we are interested in identifying discrete volumes of magnetic material. As presented in Sharma (1966), the integral equation can be evaluated analytically for the magnetic field of a rectangular prism. The integral 2.8

then becomes:

$$\vec{b}(P) = \mathbf{T} \cdot \vec{M}, \quad (2.9)$$

where the tensor matrix \mathbf{T} takes the form:

$$\mathbf{T} = \begin{pmatrix} T_{xx} & T_{xy} & T_{xz} \\ T_{yx} & T_{yy} & T_{yz} \\ T_{zx} & T_{zy} & T_{zz} \end{pmatrix}. \quad (2.10)$$

Note that only five of the nine elements forming \mathbf{T} are independent and need to be computed. Expanding 2.9 in terms of components of the field yields:

$$\begin{aligned} b_x &= T_{xx} M_x + T_{xy} M_y + T_{xz} M_z \\ b_y &= T_{yx} M_x + T_{yy} M_y + T_{yz} M_z \\ b_z &= T_{zx} M_x + T_{zy} M_y + T_{zz} M_z. \end{aligned} \quad (2.11)$$

From 2.8, I divide the earth into nc discrete prisms, each of which has a constant magnetization such that:

$$\vec{b}(P) = \sum_{j=1}^{nc} \mathbf{T}_j \cdot \vec{M}_j. \quad (2.12)$$

From the superposition principle, the \hat{x} -component of magnetic flux density $\vec{b}(P)$ corresponds to the cumulative contribution of all nc cells such that:

$$b_x(P) = \begin{bmatrix} T_{xx}^1 & \dots & T_{xx}^{nc} & T_{xy}^1 & \dots & T_{xy}^{nc} & T_{xz}^1 & \dots & T_{xz}^{nc} \end{bmatrix} \begin{bmatrix} \mathbf{M}_x \\ \mathbf{M}_y \\ \mathbf{M}_z \end{bmatrix} \quad (2.13)$$

where, $\mathbf{M}_x = \begin{bmatrix} M_x^1 \\ \vdots \\ M_x^{nc} \end{bmatrix}$ $\mathbf{M}_y = \begin{bmatrix} M_y^1 \\ \vdots \\ M_y^{nc} \end{bmatrix}$ $\mathbf{M}_z = \begin{bmatrix} M_z^1 \\ \vdots \\ M_z^{nc} \end{bmatrix}$,

and likewise for the $b_y(P)$ and $b_z(P)$ components. In compact matrix vector notation, 2.13 can be written as:

$$\vec{b}(P) = \mathbf{T} \vec{\mathbf{M}}, \quad (2.14)$$

where the augmented tensor matrix $\mathbf{T} \in \mathbb{R}^{3 \times (3nc)}$ multiplies a vector of magnetization direction $\vec{\mathbf{M}} \in \mathbb{R}^{3nc}$.

Because magnetic flux measurements can be recorded at multiple locations, 2.14 is further augmented by a factor N such that:

$$\vec{\mathbf{b}} = \begin{bmatrix} \mathbf{b}_x \\ \mathbf{b}_y \\ \mathbf{b}_z \end{bmatrix} \quad (2.15)$$

where, $\mathbf{b}_x = \begin{bmatrix} b_x(P_1) \\ \vdots \\ b_x(P_N) \end{bmatrix}$ $\mathbf{b}_y = \begin{bmatrix} b_y(P_1) \\ \vdots \\ b_y(P_N) \end{bmatrix}$ $\mathbf{b}_z = \begin{bmatrix} b_z(P_1) \\ \vdots \\ b_z(P_N) \end{bmatrix}$.

This ordering defines the rows of the final matrix used for the forward calculations of magnetic data:

$$\vec{\mathbf{b}} = \mathbf{T} \vec{\mathbf{M}}$$

where, $\vec{\mathbf{M}} = \begin{bmatrix} \mathbf{M}_x \\ \mathbf{M}_y \\ \mathbf{M}_z \end{bmatrix}$, $\mathbf{T} = \begin{bmatrix} \mathbf{T}_{xx} & \mathbf{T}_{xy} & \mathbf{T}_{xz} \\ \mathbf{T}_{yx} & \mathbf{T}_{yy} & \mathbf{T}_{yz} \\ \mathbf{T}_{zx} & \mathbf{T}_{zy} & \mathbf{T}_{zz} \end{bmatrix}$ (2.16)

$$\vec{\mathbf{b}} \in \mathbb{R}^{(3N)}, \mathbf{T} \in \mathbb{R}^{(3N) \times (3nc)}, \vec{\mathbf{M}} \in \mathbb{R}^{(3nc)}.$$

Most geophysical surveys do not collect the components of the magnetic field \vec{b} , but rather its amplitude, or Total Magnetic Intensity (TMI) such that:

$$b^{TMI} = |\vec{B}_0 + \vec{B}_A|, \quad (2.17)$$

where \vec{B}_0 is the primary geomagnetic field and \vec{B}_A are anomalous *local* fields. In mineral exploration, we are only interested in the anomalous field arising from magnetized rocks. The quantity of interest is referred to as Total Magnetic Anomaly, or b^{TMA} :

$$b^{TMA} = |\vec{B}_0 - \vec{B}_A|. \quad (2.18)$$

Assuming that $\frac{|\vec{B}_A|}{|\vec{B}_0|} \ll 1$, the anomalous field is approximated as:

$$\begin{aligned} b^{TMA} &\simeq \vec{B}_A \cdot \hat{\vec{B}}_0 \\ &= |\vec{B}_0 + \vec{B}_A| - |\vec{B}_0|. \end{aligned} \quad (2.19)$$

The anomalous data \mathbf{b}^{TMA} can be obtained from 2.16 by a projection matrix acting on \mathbf{T} such that:

$$\begin{aligned} \mathbf{b}^{TMA} &= \mathbf{P} \mathbf{T} \vec{\mathbf{M}} \\ \mathbf{P} &= \frac{1}{|\vec{\mathbf{B}}_0|} \begin{bmatrix} B_{0x} \mathbf{I} & B_{0y} \mathbf{I} & B_{0z} \mathbf{I} \end{bmatrix} \end{aligned} \quad (2.20)$$

$$\mathbf{P} \in \mathbb{R}^{N \times (3N)},$$

where the projection matrix \mathbf{P} computes the inner-product between each magnetic field measurements and the inducing field, and \mathbf{I} is an $N \times N$ identity matrix.

2.1 Magnetization

We have so far remained general and have omitted any details regarding the magnetization vectors \vec{M} . In matter, the total magnetization per unit volume can be expressed as:

$$\vec{M} = \kappa \vec{H} + \vec{M}_{NRM}, \quad (2.22)$$

where \vec{M}_{NRM} is known as the Natural Remanent Magnetization and the magnetic susceptibility κ is the intrinsic physical property of rocks describing their ability to become magnetized under an applied field \vec{H} . I will here assume that the induced magnetization is isotropic. The inducing field \vec{H} can be further divided in two parts:

$$\vec{H} = \vec{H}_0 + \vec{H}_A, \quad (2.23)$$

where \vec{H}_0 is the Earth's geomagnetic field and \vec{H}_A are the anomalous local fields. The geomagnetic field \vec{H}_0 is generally dominant and believed to be generated within the core Campbell (1997). Smaller diurnal variations can be observed due to the movement of charged particles within the upper atmosphere.

Under the assumption of no free currents, secondary fields \vec{H}_A arise from the in-

teraction of magnetic objects, or *self-demagnetization* effects. For equidimensional objects, secondary induced fields oppose the geomagnetic field and reduce the overall induced magnetization. For an elongated magnetic body, the total magnetization direction may be deflected towards the principal axis. Self-demagnetization effects usually become important as κ gets large ($\kappa \geq 0.1$). For the remainder of this research project I will assume that self-demagnetization effects are negligible.

Lastly, the Natural Remanent Magnetization, or \vec{M}_{NRM} component, is a permanent dipole moment preserved in the absence of an inducing field. Certain rocks, known as ferromagnetic material, can retain a net magnetization vector, which in some cases will reflect the orientation of the Earth's field during formation. The orientation and magnitude of the NRM component is generally unknown and difficult to distinguish from the induced component. From laboratory measurements, the strength of remanence is often expressed as a ratio with respect to the induced component:

$$Q = \frac{|\vec{M}_{NRM}|}{\kappa |\vec{H}_0|}, \quad (2.24)$$

where Q stands for the Koenigsberger ratio. The NRM component has been found to play an important role in many geological settings. Due to frequent changes in the Earth's polarity and various geological processes, rock magnetization directions can vary greatly within a given region. Accurately determining the orientation of magnetization is the principal focus of this research project.

While most magnetic methods proposed in the literature assume a purely induced response, a more general approach must account for variability in magnetization direction. From a geophysical standpoint, we would like to infer the distribution of magnetic material from the observed field data. Modeling magnetic objects though the inverse problem without knowledge about the magnetization direction has proven to be difficult and remains a field of active research in geophysics.

Chapter 3

Inverse Problem

3.1 Regularized inversion

In chapter 2, I introduced the linear forward calculation for the anomalous magnetic data \mathbf{b}^{TMA} generated by volumes of magnetization. In a more general case, the discretized system of equations relating the model to the geophysical data can be written as:

$$\mathbb{F}[\mathbf{m}] = \mathbf{d}, \quad (3.1)$$

where \mathbb{F} is a generic operator relating the geophysical data $\mathbf{d} \in \mathbb{R}^N$ to discrete model parameters $\mathbf{m} \in \mathbb{R}^{nc}$. In some cases, \mathbb{F} is linear, but can be non-linear as is the case with amplitude data. We are usually interested in finding a solution to the inverse problem:

$$\mathbf{m} = \mathbb{F}^{-1}\mathbf{d}, \quad (3.2)$$

namely to recover some model parameters \mathbf{m} responsible for a set of observations \mathbf{d} . At this point, the magnetization model \mathbf{m} is general. In this thesis, \mathbf{m} will take different forms depending on assumptions made, coordinate system chosen or data transformation.

It is difficult to solve the inverse problem for two reasons. First, the inverse problem is often *ill – posed* as the number of unknown parameters largely exceeds the number of observations. The solution is highly non-unique. Secondly, field data are generally corrupted by random noise \mathbf{e} such that the true linear system

should be written as:

$$\begin{aligned}\mathbb{F}[\mathbf{m}] &= \mathbf{d}^{\text{obs}} \\ \mathbf{d}^{\text{obs}} &= \mathbf{d} + \mathbf{e}.\end{aligned}\tag{3.3}$$

Even if the problem is linear and the matrix \mathbb{F} is full ranked and invertible, the problem is said to be *ill-conditioned*. Any solution that satisfies 3.3 is unstable, as small changes in the noise can induce large changes in model values. Moreover, most geophysical methods have a strong spatial dependency related to the distance between the source and observation location. Consequently, few components describing the system \mathbb{F} can be significantly larger, which can adversely impact numerical solvers.

The inverse problem described by 3.3 may admit an infinite number of solutions that are geologically unrealistic, and the solution may not be very stable with respect to the observed data. As first introduced by Tikhonov and Arsenin (1977), the inverse problem can be formulated as a *regularized* least-squares problem of the form:

$$\begin{aligned}\min_m \phi(m) \\ \phi(m) &= \phi_d + \beta \phi_m \\ \phi_d &= \|\mathbf{W}_d (\mathbb{F}[\mathbf{m}] - \mathbf{d}^{\text{obs}})\|_2^2 \\ \phi_m &= \mathbf{R},\end{aligned}\tag{3.4}$$

where the optimal solution is found at the minimum of the objective function $\phi(m)$. The misfit function ϕ_d measures the residual between predicted and observed data \mathbf{d}^{obs} normalized by the estimated uncertainties \mathbf{W}_d :

$$\mathbf{W}_d = \begin{bmatrix} 1/\sigma_1 & 0 & \dots & 0 \\ 0 & 1/\sigma_2 & 0 & \vdots \\ \vdots & \ddots & & 0 \\ 0 & \dots & 0 & 1/\sigma_N \end{bmatrix},\tag{3.5}$$

where σ_i are assigned standard deviations. Assuming the noise to be Gaussian

and uncorrelated, the misfit function follows a chi-squared distribution with an expected value of N . Under this assumption, the expected data misfit is also equal to the number of data N .

The model objective function ϕ_m is added to stabilize and constrain the solution. The trade-off parameter β balances the relative influence between the misfit function and any *a priori* information prescribed by a chosen regularization \mathbf{R} . There has been much research done on designing robust and effective regularization functions. In the original magnetic inversion work of Li and Oldenburg (1996), the regularization function involves a measure of model *smallness* and *smoothness*. The general objective function takes the form:

$$\phi(m) = \phi_d + \beta \left[\alpha_s \int_V w_s(r) |m(r) - m^{ref}|^2 dV + \sum_{i=x,y,z} \alpha_i \int_V w_i(r) \left| \frac{\partial m(r)}{\partial x_i} \right|^2 dV \right], \quad (3.6)$$

where α_s , α_x , α_y and α_z are adjustable constants balancing the relative contribution between the various components of the model objective function. The first integral measures the deviation from a reference model m^{ref} . The three following integrals measure the spatial gradients of the model $m(r)$ in Cartesian coordinates. The weighting functions $w_s(r)$, $w_x(r)$, $w_y(r)$ and $w_z(r)$ are cell-based penalties added to the system to reflect specific characteristics expended from the solution. For potential field problems, it is common to resort to a sensitivity-based weighting in order to compensate for the natural decay of the kernel functions, such that:

$$\begin{aligned} w_s(r) &= w_r(r) \tilde{w}_s(r) \\ w_i(r) &= w_r(r) \tilde{w}_i(r), \end{aligned} \quad (3.7)$$

where $w_r(r)$ is a general distance weighting, and \tilde{w}_s , \tilde{w}_i are customizable weights that reflect any available *a priori* information. Note that the distance weighting $w_r(r)$ is applied through the regularization function rather than directly to the sensitivity matrix as prescribed by Li and Oldenburg (1996). Reasons for this change will be discussed in Section 4.4.1 while experimenting with different l_p -norm regularization functions.

I discretize 3.6 on a tensor mesh made of rectangular prims such that:

$$\phi(m) = \phi_d + \beta \left[\|\mathbf{W}_s (\mathbf{m} - \mathbf{m}^{ref})\|_2^2 + \sum_{i=x,y,z} \|\mathbf{W}_i \mathbf{G}_i \mathbf{m}\|_2^2 \right], \quad (3.8)$$

where the diagonal matrices \mathbf{W}_s and $\mathbf{W}_i \in \mathbb{R}^{nc \times nc}$ contain dimensional scales that arise from the discretization and cell-based weights ($w_s(r)$, $w_i(r)$). The global scaling constants α_s and α_x are also absorbed as cell weights. The measure of spatial gradients are calculated by a generic forward difference scheme such that:

$$\mathbf{G}_i = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & -1 & 1 \\ 0 & \dots & 0 & 1 & -1 \end{bmatrix}, \quad (3.9)$$

where the gradient operator $\mathbf{G}_i \in \mathbb{R}^{nc \times nc}$ calculates the horizontal changes in model parameters \mathbf{m} . I voluntarily removed all spatial dimensions out of the gradient operators for reasons that will be discussed in Section 4.4.1. The last row is using a backward difference in order to get a square matrix. The banding structure of the gradient operators is different for \mathbf{G}_x , \mathbf{G}_y and \mathbf{G}_z to account for the ordering of the cells in the tensor mesh.

For ease of notation, I write the model objective function in compact form as:

$$\phi_m = \|\mathbf{W}_m(\Delta\mathbf{m})\|_2^2, \quad (3.10)$$

such that:

$$\mathbf{W}_m^T \mathbf{W}_m = \mathbf{W}_s^T \mathbf{W}_s + \sum_{i=x,y,z} \mathbf{G}_i^T \mathbf{W}_i^T \mathbf{W}_i \mathbf{G}_i,$$

as well as:

$$\Delta\mathbf{m} = \mathbf{m} - \mathbf{m}^{ref},$$

to define the deviation between the model parameters \mathbf{m} and the reference model \mathbf{m}^{ref} .

3.1.1 Iterative solver

As previously stated, my goal is to minimize the objective function described by 3.8. The minimum solution is found where the partial gradients of the function are vanishing such that:

$$\mathbf{g}(\mathbf{m}) = \frac{\partial \phi(m)}{\partial m} = 0. \quad (3.11)$$

Taking the partial derivatives of 3.8 with respect to the model parameters yield:

$$\mathbf{J}^T \mathbf{W}_d^T \mathbf{W}_d [\mathbb{F}[\mathbf{m}] - \mathbf{d}^{obs}] + \beta \mathbf{W}_m^T \mathbf{W}_m (\Delta \mathbf{m}) = 0, \quad (3.12)$$

where \mathbf{J} is the Jacobian of the forward operator:

$$\mathbf{J} = \frac{\partial \mathbb{F}[\mathbf{m}]}{\partial \mathbf{m}}. \quad (3.13)$$

More generally, we are interested in solving the nonlinear system described by 3.12, which can be done with second-order methods. Newton's method computes a series of model updates $\delta \mathbf{m}$ by solving the system:

$$\mathbf{H} \delta \mathbf{m} = -\mathbf{g}(\mathbf{m}^{(j)}), \quad (3.14)$$

where \mathbf{H} is the Hessian, or second-order derivatives of the objective function. The optimization problem is solved iteratively such that:

$$\mathbf{m}^{(j+1)} = \mathbf{m}^{(j)} + \alpha \delta \mathbf{m}, \quad (3.15)$$

where the superscript (j) denotes the solver iterations for a model perturbation $\delta \mathbf{m}$ scaled by a step length α .

Computing the true Hessian can be computationally expensive. For least-squares problems, the Gauss-Newton method can be used to approximate the Hessian such that:

$$\frac{\partial^2 \phi}{\partial m^2} \approx \tilde{\mathbf{H}} = \mathbf{J}^T \mathbf{W}_d^T \mathbf{W}_d \mathbf{J} + \beta \mathbf{W}_m^T \mathbf{W}_m, \quad (3.16)$$

where the assumption is made that the change in Jacobian is small between each iteration:

$$\mathbf{J}(\mathbf{m}^{(j)} + \delta \mathbf{m}) \approx \mathbf{J}(\mathbf{m})^{(j)}, \quad (3.17)$$

and that the model update is approximated to be along the gradient direction such that:

$$\mathbb{F}[\mathbf{m} + \delta\mathbf{m}] \approx \mathbb{F}[\mathbf{m}] + \mathbf{J}\delta\mathbf{m}. \quad (3.18)$$

Combining 3.12 and 3.16, and ignoring the superscript (j) yields:

$$\begin{aligned} \tilde{\mathbf{H}} \delta\mathbf{m} &= -\mathbf{g}(\mathbf{m}) \\ (\mathbf{J}^T \mathbf{W}_d^T \mathbf{W}_d \mathbf{J} + \beta \mathbf{W}_m^T \mathbf{W}_m) \delta\mathbf{m} &= \\ &- \mathbf{J}^T \mathbf{W}_d^T \mathbf{W}_d [\mathbb{F}[\mathbf{m}] - \mathbf{d}^{obs}] - \beta \mathbf{W}_m^T \mathbf{W}_m (\Delta\mathbf{m}). \end{aligned} \quad (3.19)$$

Solving 3.19 gives a step direction. Each Gauss-Newton step requires a solution to linear system of the form $\mathbf{A} \mathbf{x} = \mathbf{b}$. In our case, the left-hand-side of 3.19 can be assumed to be a symmetric positive definite matrix. Hence it is possible find a unique solution by directly computing the inverse of the pseudo-Hessian. From a practical standpoint however, the computational cost of this operation on a dense matrix is $O(nc^3)$, which can rapidly become prohibitive for a large system of equations.

Krylov Space methods have been proposed as an alternative to direct solvers. Iterative solvers, such as the Conjugate Gradient (CG) method, apply a series of orthogonal steps towards the minimum of a function. Each step is said to be A -orthogonal to the previous ones, contributing to the fast convergence of CG over simpler gradient descent methods. The reader is encouraged to read Shewchuk (1994) for a comprehensive review of the method. The algorithm is presented in Table 3.1. It can be shown that for a symmetric positive definite matrix \mathbf{A} , the CG method converges to a unique solution in at most $O(nc)$ operations. It is rarely required to solve the system exactly however. My algorithm uses a stopping criteria for the minimum CG update ($\delta > 1e-4$), after which a Gauss-Newton step direction $\delta\mathbf{m}$ is returned. The step length α is calculated by a line-search method as presented in Table 3.2. The Gauss-Newton steps are repeated until the model updates falls below some threshold value, $|\alpha\delta\mathbf{m}| < \gamma$, where γ is some small value. In this thesis, I fix $\gamma = 1e-4$, which experimentally has proven to be a good compromise between computation cost and accuracy.

In order to reduce the amount of memory required to store the dense matrix $\tilde{\mathbf{H}}$

Table 3.1: Conjugate Gradient algorithm

Initialize: $\mathbf{d}_{(0)} = \mathbf{r}_{(0)} = \mathbf{b} - \mathbf{A} \mathbf{x}_{(0)}$
<i>while</i> : $\ \mathbf{r}\ > \delta$
$\alpha_{(i)} = \frac{\mathbf{r}_{(i)}^T \mathbf{r}_{(i)}}{\mathbf{d}_{(i)}^T \mathbf{A} \mathbf{d}_{(i)}}$
$\mathbf{x}_{(i+1)} = \mathbf{x}_{(i)} + \alpha_{(i)} \mathbf{d}_{(i)}$
$\mathbf{r}_{(i+1)} = \mathbf{r}_{(i)} - \alpha_{(i)} \mathbf{A} \mathbf{d}_{(i)}$
$\beta_{(i+1)} = \frac{\mathbf{r}_{(i+1)}^T \mathbf{r}_{(i+1)}}{\mathbf{r}_{(i)}^T \mathbf{r}_{(i)}}$
$\mathbf{d}_{(i+1)} = \mathbf{r}_{(i+1)} + \beta_{(i+1)} \mathbf{d}_{(i)}$

Table 3.2: Line-search

Initialize: $\alpha = 1$, $\hat{\mathbf{m}} = \mathbf{m}^{(j)}$
<i>while</i> : $\phi(\hat{\mathbf{m}}) \geq \phi(\mathbf{m}^{(j)})$
$\alpha = \alpha/2$
$\hat{\mathbf{m}} = \mathbf{m}^{(j)} + \alpha \delta \mathbf{m}$
$\mathbf{m}^{(j+1)} = \hat{\mathbf{m}}$

in 3.19, the Gauss-Newton steps can be formulated as an overdetermined problem of the form:

$$\begin{bmatrix} \mathbf{W}_d \mathbf{J} \\ \sqrt{\beta} \mathbf{W}_s \\ \sqrt{\beta} \mathbf{W}_x \mathbf{G}_x \\ \sqrt{\beta} \mathbf{W}_y \mathbf{G}_y \\ \sqrt{\beta} \mathbf{W}_z \mathbf{G}_z \end{bmatrix} \delta \mathbf{m} = - \begin{bmatrix} \mathbf{W}_d (\mathbb{F}[\mathbf{m}] - \mathbf{d}^{obs}) \\ \mathbf{W}_s (\Delta \mathbf{m}) \\ \mathbf{W}_x \mathbf{G}_x (\Delta \mathbf{m}) \\ \mathbf{W}_y \mathbf{G}_y (\Delta \mathbf{m}) \\ \mathbf{W}_z \mathbf{G}_z (\Delta \mathbf{m}) \end{bmatrix}. \quad (3.20)$$

The CG steps presented in Table 3.1 are altered slightly in order to calculate the residual $\mathbf{r}_{(i)}$ and descent direction $\mathbf{d}_{(i)}$. Memory savings are important as I avoid forming a dense system $\tilde{\mathbf{H}} \in \mathbb{R}^{nc \times nc}$, but only require sparse matrix-vector products.

3.1.2 Preconditioner

As explained in Section 3.1.1, I want to solve iteratively a normal equation of the form $\mathbf{Ax} = \mathbf{b}$. The performance of iterative solvers depends strongly on the

condition number κ of the left-hand side matrix \mathbf{A} such that:

$$\kappa = \frac{\lambda_{\max}}{\lambda_{\min}}, \quad (3.21)$$

where λ_{\max} and λ_{\min} are the largest and smallest eigenvalues of \mathbf{A} . It can be shown that an upper bound on the convergence rate of CG is function of this condition number such that:

$$\omega \leq \frac{\kappa - 1}{\kappa + 1}, \quad (3.22)$$

where the convergence rate ω gets worse as the condition number increases. In order to accelerate the convergence, it is common to resort to a *preconditioner* to improve the spectral properties of \mathbf{A} (van der Vorst, 2003). The basic idea is to pre-multiply the linear system such that:

$$\mathbf{P} \mathbf{A} \mathbf{x} = \mathbf{P} \mathbf{b}, \quad (3.23)$$

where the resulting condition number of $\mathbf{P} \mathbf{A}$ is smaller than the original matrix \mathbf{A} . Assuming that \mathbf{A} is invertible, the perfect preconditioner is its inverse \mathbf{A}^{-1} , in which case $\mathbf{A}^{-1}\mathbf{A}$ is the identity matrix with a condition number of 1. Computing this preconditioner would obviously be redundant, since if I know \mathbf{A}^{-1} , then I would have already solved the problem.

I can decompose the matrix \mathbf{A} in its lower and diagonal blocks:

$$\mathbf{A} = \mathbf{L} + \mathbf{D} + \mathbf{L}^T, \quad (3.24)$$

where in our case the matrix \mathbf{D} is the diagonal of the pseudo Hessian described by Equation 3.19. The Jacobi preconditioner is the simplest choice where I set $\mathbf{P} = \mathbf{D}^{-1}$. In this research project, the matrix \mathbf{A} is a positive definite matrix, hence \mathbf{D} is assumed to have diagonal elements that are strictly non-zero positive numbers. The preconditioning matrix \mathbf{P} is updated before each Gauss-Newton step and pre-multiplied ahead of the Conjugate Gradient solver presented in Table 3.1.

3.1.3 Bound constraints

In many cases, the solution \mathbf{m} is expected to be bounded on a specific interval. In particular for the magnetic problem, the magnetic susceptibility κ can only be a positive real number on the interval $\kappa_i \in [0, \infty)$. Three strategies have been proposed in the literature. In Li and Oldenburg (2003), positive constraints are applied via a log-barrier penalty term such that:

$$\phi(\lambda) = \phi_d + \beta\phi_m - 2\lambda \sum_{i=1}^{nc} \ln(\kappa_i), \quad (3.25)$$

where the log barrier parameter λ controls the influence of the logarithmic penalty function. The method solves a series of non-linear problems by monotonically reducing λ until convergence to a stable solution has been reached. As $\kappa_i \rightarrow 0$, the $\log(\kappa_i)$ tends to be large and negative, hence strongly penalized during the minimization process. The same approach can be used to impose an upper bound on the model parameters.

In a second approach, the model values are parameterized into a quantity that can only be positive. Lelièvre (2003) proposes transforming the model parameters such that:

$$m_i = \kappa_i^{(1/2)}, \quad (3.26)$$

in which case if $m_i \in (-\infty, \infty)$, then $\kappa_i \in [0, \infty)$, hence a strictly positive value. This method has the advantage of reducing the size of the system to be solved, but can only be used as a lower bound.

The third strategy, which is used in this work, is the Projected Gradient method as presented in Vogel (2002). Within the line-search algorithm presented in Table 3.2, model parameters outside the bounds are replaced and set to be inactive for the following Gauss-Newton iteration.

All three strategies make the inverse problem non-linear with respect to \mathbf{m} . The Projected Gauss-Newton method is the simplest to implement however within an iterative solver framework. It does not require additional parameters to be adjusted and can be used to impose both an upper and a lower bound. All the algorithms presented in this thesis were implemented with the Projected Gauss-Newton method.

3.2 Magnetic susceptibility inversion

As formulated in 2.20, the observed Total Magnetic Anomaly data \mathbf{b}^{TMA} are related to discrete magnetization parameters $\vec{\mathbf{M}}$ by a large under-determined system of equations $\mathbf{P} \mathbf{T} \in \mathbb{R}^{N \times (3*nc)}$. A long standing strategy, as adopted by Li and Oldenburg (1996), Pilkington (1997) and others, is to neglect the effect of self-demagnetization and remanence and invert directly for the magnetic susceptibility of rocks. Assuming a uniform magnetization direction parallel to the Earth's field \vec{H}_0 , the linear system 2.11 can be written as:

$$\begin{aligned} b_x &= [T_{xx} H_{0x} + T_{xy} H_{0y} + T_{xz} H_{0z}] \kappa \\ b_y &= [T_{xx} H_{0x} + T_{xy} H_{0y} + T_{xz} H_{0z}] \kappa \\ b_z &= [T_{xx} H_{0x} + T_{xy} H_{0y} + T_{xz} H_{0z}] \kappa, \end{aligned} \quad (3.27)$$

where the magnetic field response is linearly related to the effective susceptibility parameter κ . For a larger system involving N observations over a model space discretized into nc cells, 3.27 is written as:

$$\begin{aligned} \mathbf{b}^{TMA} &= \mathbf{P} \mathbf{T} \mathbf{H} \kappa \\ \mathbf{H} &= \begin{bmatrix} H_{0x} \mathbf{I} \\ H_{0y} \mathbf{I} \\ H_{0z} \mathbf{I} \end{bmatrix}, \quad \kappa = \begin{bmatrix} \kappa_1 \\ \kappa_2 \\ \vdots \\ \kappa_M \end{bmatrix} \end{aligned} \quad (3.28)$$

$$\mathbf{H} \in \mathbb{R}^{(3nc) \times nc}, \quad \kappa \in \mathbb{R}^{nc},$$

where \mathbf{I} is the identity matrix and \mathbf{H} is a block diagonal matrix containing the orientation of the inducing field \vec{H}_0 in Cartesian coordinates. Since the inducing field \vec{H}_0 is assumed to be constant, the matrix product $\mathbf{P} \mathbf{T} \mathbf{H}$ can be directly incorporated into a linear system of equations such that:

$$\mathbf{b}^{TMA} = \mathbf{F} \kappa, \quad (3.29)$$

where $\mathbf{F} \in \mathbb{R}^{(N) \times nc}$ is the linear forward operator relating the discrete magnetic susceptibility values κ to the magnetic field observations \mathbf{b}^{TMA} .

3.2.1 Synthetic example

Having defined the linear system relating the anomalous field data \mathbf{b}^{TMA} to a discrete magnetic susceptibility model κ , I illustrate the forward calculation with a synthetic example shown in Figure 3.1. The model consists of a folded anomaly with induced magnetization of 3 A/m, arching around a discrete block with magnetization of 2 A/m. The arc-shaped anomaly dips 20° towards the south. From the relation between induced magnetization and ambient field:

$$\vec{M}_{ind} = \kappa \vec{B}_0 / \mu_0 , \quad (3.30)$$

the susceptibility of both anomalies are 0.075 and 0.05 SI respectively, subject to a 50,000 nT inducing field \vec{H}_0 . I create this model in order to simulate complex geology with varying shapes and depth extent. The model is discretized on a uniform 20 meters cell size mesh. A grid of 342 observation stations is placed on a plane 20 meters above the top of the mesh. I generate \mathbf{b}^{TMA} data using 3.29 as shown in Figure 3.2. Data are corrupted with random Gaussian noise, 1 nT standard deviation, to simulate field data conditions.

I then follow with the inverse problem, namely to recover a distribution of susceptibility values from the \mathbf{b}^{TMA} data as presented in Li and Oldenburg (1996). Similar to the general system 3.4, the objective function to be minimized is written as:

$$\phi(\kappa) = \|\mathbf{W}_d(\mathbf{F}\kappa - \mathbf{d}^{obs})\|_2^2 + \beta \|\mathbf{W}_m\kappa\|_2^2 . \quad (3.31)$$

In this case, \mathbf{W}_d is simply the identity matrix since the noise is known to have a standard deviation of exactly one. I also set the reference mode \mathbf{m}^{ref} to zero, hence forcing the solution to be *small* and *smooth*. Table 3.3 summarizes the inversion parameters used.

I minimize 3.31 iteratively following the procedure presented in Section 3.1.1. Figure 3.3 shows the convergence curve for the data misfit $\phi_d^{(k)}$ and model norm $\phi_m^{(k)}$ as a function of β iterations. The inversion achieved the target misfit after seven

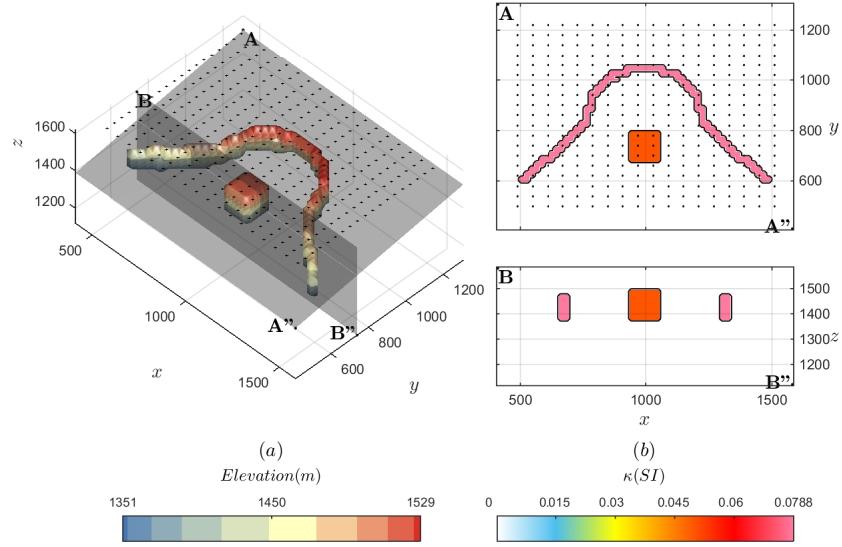


Figure 3.1: (a) Synthetic susceptibility model consisting of a folded anomaly ($\kappa = 0.075$ SI) arching around a discrete block ($\kappa = 0.05$ SI) .

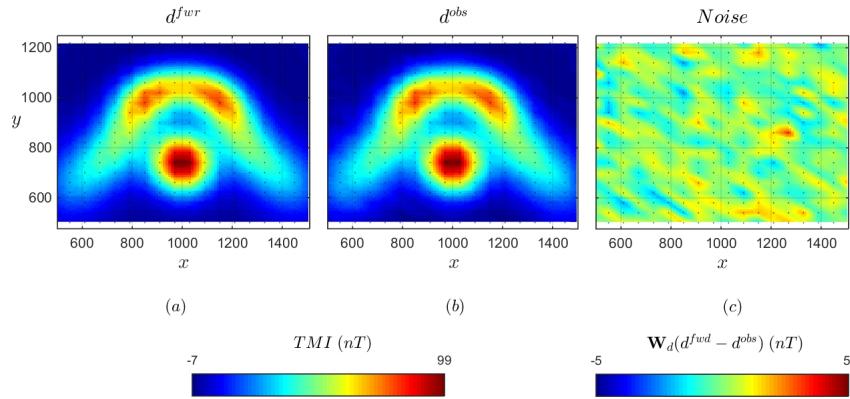


Figure 3.2: (a) Data generated from the synthetic susceptibility model subject to a vertical 50,000 nT inducing field. (b) Data are then corrupted with (c) random Gaussian noise, 1 nT standard deviation.

Table 3.3: Inversion parameters.

Core cell size	$20 \times 20 \times 20$ m
nc cells	82,000
N data	342
$\alpha_s, \alpha_x, \alpha_y, \alpha_z$	2.5e-3, 1, 1, 1
w_s, w_x, w_y, w_z	1, 1, 1, 1
Uncertainties	1 nT

iterations, after which the misfit function levels off. In cases where the true noise level is unknown, it is common practice to chose the point of maximum curvature on the misfit curve as the optimal model. Letting the inversion progress further down the misfit curve comes at the risk of fitting some of the noise, potentially introducing artifacts in the model.

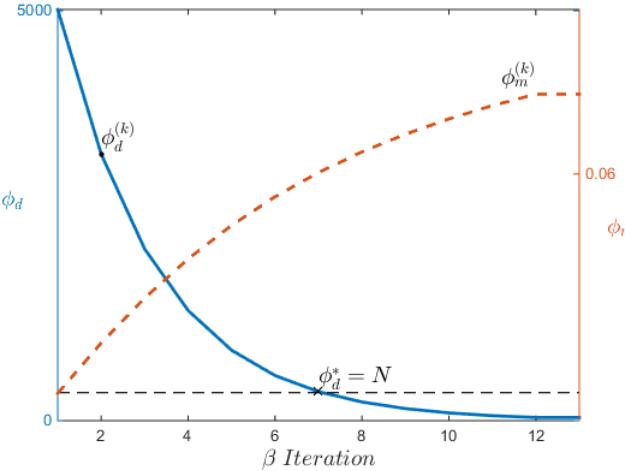


Figure 3.3: Convergence curve showing the data misfit $\phi_d^{(k)}$ and model norm $\phi_m^{(k)}$ as a function of β iterations. The inversion achieves target misfit after the 7th iteration, which in this case also corresponds to the point of maximum curvature on the misfit curve. Attempting to further lower the data residual comes at the risk of fitting some of the Gaussian noise.

Sections of the recovered susceptibility model from the 7th iteration are shown

in Figure 3.4. Both anomalies are recovered at roughly the right location and at depth. As expected from the l_2 -norm regularization, the model is smoothly varying spatially. Peak susceptibility values are underestimated by roughly 20% due to the smooth constraint. The model can predict the data well within one standard deviation on average (Fig. 3.5). It is worth noting that there are correlated residuals directly above the magnetic anomaly. Even though the global misfit criterion is satisfied, there is clearly information in the data that is not being captured. One option would be to lower the target misfit, but at the risk of fitting some of the random noise. I tested that strategy on this example, which returned speckly models with isolated susceptibility anomalies near the surface. A second option would be to manually reduce the uncertainties only over the correlated residuals. This kind of processing is time consuming and somewhat arbitrary. I will argue that the smooth regularization prevents the inversion from fitting the high frequency content. In Chapter 4 I introduce a compact norm regularization in order to test this hypothesis.

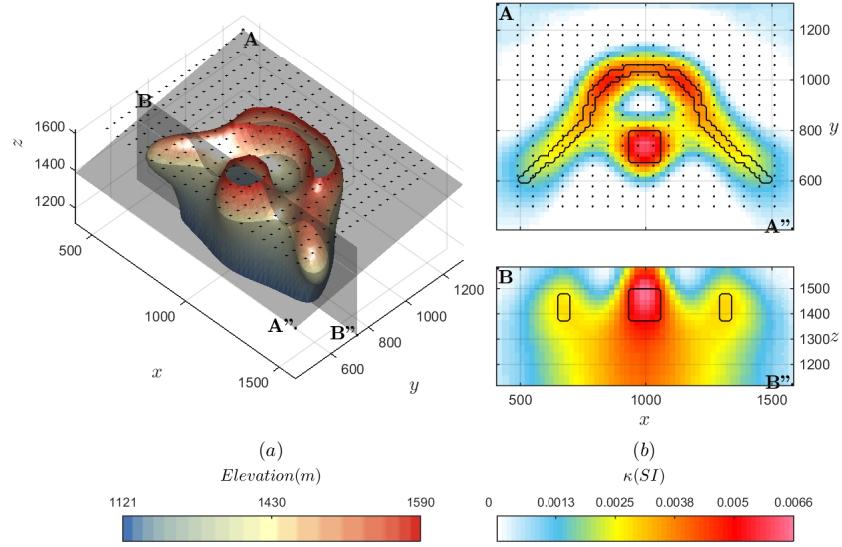


Figure 3.4: (a) Iso-surface (0.002 SI) and (b) sections through the recovered susceptibility model for a purely induced response. The model is smooth but recovers the arc and block anomaly at roughly the right depth.

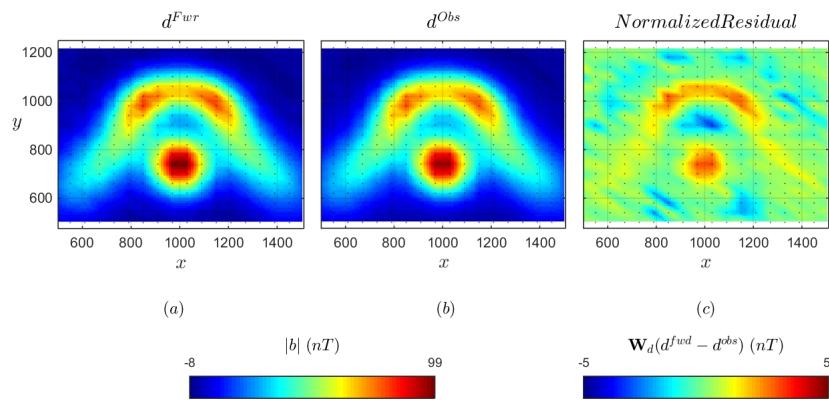


Figure 3.5: Comparison between (a) observed and (b) predicted data from the recovered susceptibility model. (c) The normalized data residuals appear to be correlated with the location of the magnetic body.

3.2.2 Effect of remanent magnetization

Recent studies have shown that the assumption of a purely induced response may not be adequate in many geological settings (Buchan et al., 2009; Enkin, 2014). As presented in Lelièvre (2009), incorrect assumptions regarding the orientation of magnetization can make the inversion process unreliable. To simulate this problem, I alter the synthetic model presented in Figure 3.1. I change the orientation of magnetization along the arc-shaped anomaly as shown in Figure 3.6 to simulate a geological folding of a ferromagnetic object. The magnetization inclination is orientated at 45° from the horizontal, with variable declinations between $[-45^\circ N ; 45^\circ N]$. Data are generated from the new magnetization model and corrupted with random Gaussian noise, 1 nT standard deviation (Fig 3.7). Compared to the purely induced response, the positive anomaly from the arc shifts towards the center while creating a large negative anomaly on the outside boundary of the arc.

Following the same procedure, I invert this new data set while still assuming a purely vertical magnetization direction. As presented in Figure 3.8, the inversion poorly recovers the location of the arc-shaped body. A broad susceptibility anomaly is created on the periphery of the data set, pushing susceptibilities outward in order to account for the large negative data.. The signal from the arc also affects the center block anomaly, that is now recovered deeper than the true location. As seen in Figure 3.9, the inversion algorithm has difficulty fitting the strong negative data. In a mineral exploration context, the model presented in 3.8 could result in false drilling targets —costly both in time, resources and confidence in geophysical methods. This is a well known problem in mineral exploration, thus a need for inversion methods that can account for remanence.

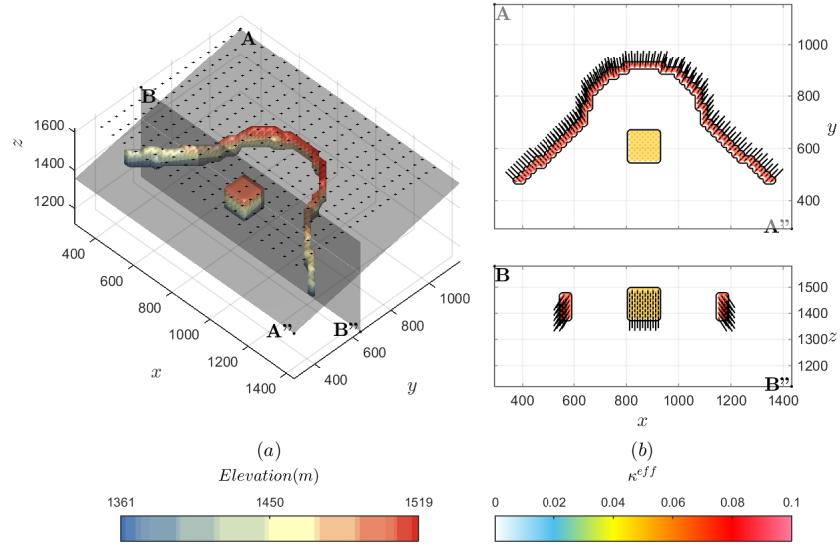


Figure 3.6: Perspective view and sections through the synthetic magnetization model. The arc-shaped anomaly is magnetized at 45° from horizontal and with variable declinations directions between $[-45^\circ N ; 45^\circ N]$.

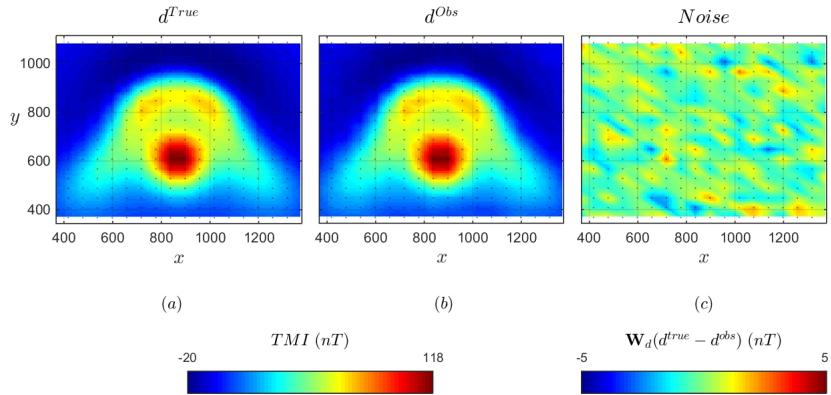


Figure 3.7: (a) Data generated from the synthetic magnetization model. (b) Observed data are corrupted with (c) random Gaussian noise, 1 nT standard deviation.

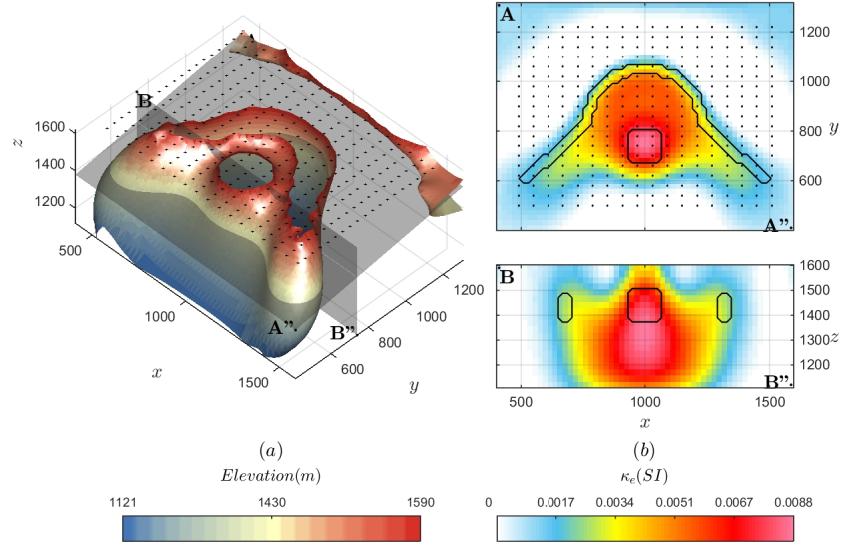


Figure 3.8: (a) Iso-surface (0.002 SI) and (b) sections through the recovered susceptibility model assuming no remanence. The arc-shaped anomaly is poorly recovered and magnetic susceptibilities are pushed at depth and outwards.

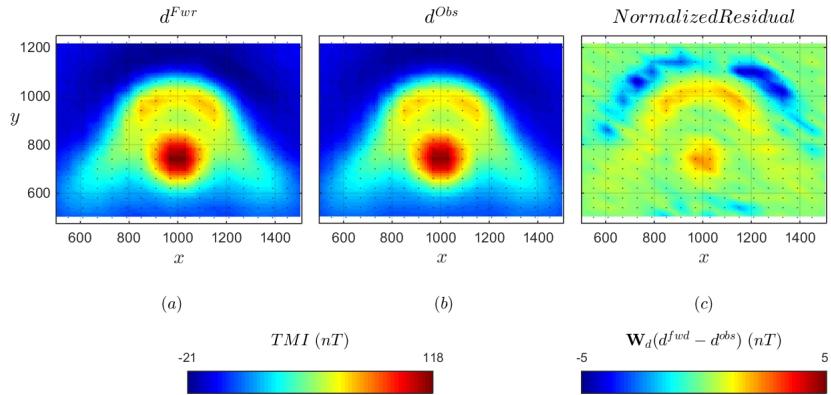


Figure 3.9: Comparison between (a) observed and (b) predicted data from the recovered susceptibility model assuming a purely induced response. (c) The inversion has a harder time fitting the large negative fields along the arc.

3.3 Magnetic vector inversion

As shown in the previous example, the assumption of a purely induced magnetization direction has its limitations. In his PhD thesis, Lelièvre (2009) introduces a Magnetization Vector Inversion (MVI) procedure in order to recover both the location and orientation of magnetization. From the general expression for the magnetic field of a prism in 2.11, the forward problem is formulated as:

$$\vec{b} = \mathbf{T} \cdot \vec{M}_p + \mathbf{T} \cdot \vec{M}_s + \mathbf{T} \cdot \vec{M}_t , \quad (3.32)$$

where I define three orthogonal components of magnetization. The primary \hat{p} component is set to be aligned with the inducing field, the second \hat{s} component has the same azimuth as the primary but is rotated 90° in inclination. The third component \hat{t} lies on the xy -plane and perpendicular to \hat{p} and \hat{s} . I define the transformation from Cartesian to {p,s,t} coordinate system by a double rotation such that:

$$\begin{bmatrix} \hat{p} & \hat{s} & \hat{t} \end{bmatrix} = \mathbf{R}_z(\theta) \mathbf{R}_x(\phi) \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} , \quad (3.33)$$

where θ and ϕ are the declination and inclination of the inducing field as defined in a Cartesian coordinate, positive counter-clockwise. For an inducing field \vec{H}_0 oriented 0°I, 0°D, the components of magnetization \hat{p} , \hat{s} and \hat{t} would be aligned with the Cartesian components \hat{y} , \hat{z} and \hat{x} respectively.

For the b_x component of the magnetic field due to a single prism magnetized along the {p,s,t} direction can be written as:

$$b_x = \begin{bmatrix} T_{xx} & T_{xy} & T_{xz} \end{bmatrix} \left(\begin{bmatrix} H_{\hat{p}_x} \\ H_{\hat{p}_y} \\ H_{\hat{p}_z} \end{bmatrix} \kappa_p + \begin{bmatrix} H_{\hat{s}_x} \\ H_{\hat{s}_y} \\ H_{\hat{s}_z} \end{bmatrix} \kappa_s + \begin{bmatrix} H_{\hat{t}_x} \\ H_{\hat{t}_y} \\ H_{\hat{t}_z} \end{bmatrix} \kappa_t \right) , \quad (3.34)$$

and likewise for the b_y and b_z components of the magnetic field. The magnetic data \vec{b} is now function on three orthogonal components of *effective susceptibility* κ_p , κ_s and κ_t and their respective direction of magnetization \vec{H}_p , \vec{H}_s and \vec{H}_t defined by the

$\{p, s, t\}$ coordinate system such that:

$$\begin{aligned}\vec{H}_p &= |\vec{H}_0| \hat{p} \\ \vec{H}_s &= |\vec{H}_0| \hat{s} \\ \vec{H}_t &= |\vec{H}_0| \hat{t}.\end{aligned}\tag{3.35}$$

Just as I did for the magnetic susceptibility problem, I augment the linear system for the magnetic response of nc magnetized prisms as observed at N data locations. Equation 3.32 becomes:

$$\begin{aligned}b^{TMA} &= \mathbf{F} \hat{\kappa} \\ &= \mathbf{P} \mathbf{T} \begin{bmatrix} \mathbf{H}_p & \mathbf{H}_s & \mathbf{H}_t \end{bmatrix} \begin{bmatrix} \kappa_p \\ \kappa_s \\ \kappa_t \end{bmatrix}.\end{aligned}\tag{3.36}$$

where $\hat{\mathbf{F}} \in \mathbb{R}^{N \times (3nc)}$ has now three times the number of variables as the susceptibility inversion described in 3.29. The block diagonal matrices $\mathbf{H}_p, \mathbf{H}_s$ and $\mathbf{H}_t \in \mathbb{R}^{3nc \times nc}$ relate the three effective susceptibility parameters κ_p, κ_s and $\kappa_t \in \mathbb{R}^{nc}$ to the system of equation \mathbf{T} in Cartesian coordinates.

The objective function to be minimized becomes:

$$\phi(\hat{\kappa}) = \|\mathbf{W}_d(\mathbf{F}\hat{\kappa} - \mathbf{d}^{obs})\|_2^2 + \beta \|\hat{\mathbf{W}}_m \hat{\kappa}\|_2^2.\tag{3.37}$$

where the weighting matrices and gradient operators comprised by $\hat{\mathbf{W}}_m$ are block diagonal matrices such that:

$$\begin{aligned}\hat{\mathbf{W}}_{\square} &= \begin{bmatrix} \mathbf{W}_{\square} & \dots & 0 \\ \vdots & \mathbf{W}_{\square} & \vdots \\ 0 & \dots & \mathbf{W}_{\square} \end{bmatrix} \\ \hat{\mathbf{G}}_{\square} &= \begin{bmatrix} \mathbf{G}_{\square} & \dots & 0 \\ \vdots & \mathbf{G}_{\square} & \vdots \\ 0 & \dots & \mathbf{G}_{\square} \end{bmatrix}.\end{aligned}$$

Each operator is three times the size as those used for the susceptibility inversion.

The inversion process follows the same method as the susceptibility inversion and uses the data presented in Figure 3.7. The only difference is that effective susceptibilities are allowed to go negative. Figure 3.10 presents sections of the recovered magnetization model, with color scale representing the total effective susceptibility κ_e where:

$$\kappa_e = (\kappa_p^2 + \kappa_s^2 + \kappa_t^2)^{1/2}. \quad (3.38)$$

The inverted model recovers the approximate orientation of magnetization in the block anomaly and part of the arc. The result is smooth however, making it hard to distinguish the extremities of the arc. Effective susceptibilities associated with the center block extend deeper than the true model. Large effective susceptibility values are found near the top of the mesh, leaving holes in the iso-surface.

The current inversion strategy may work well in simple cases, or when substantial *a priori* information is provided. For more general cases, the smooth model presented in Figure 3.10 could lead to false interpretation due to the lack of boundaries. There is still need to develop a robust algorithm that could reduce the non-uniqueness of the MVI formulation with minimal input required from the user.

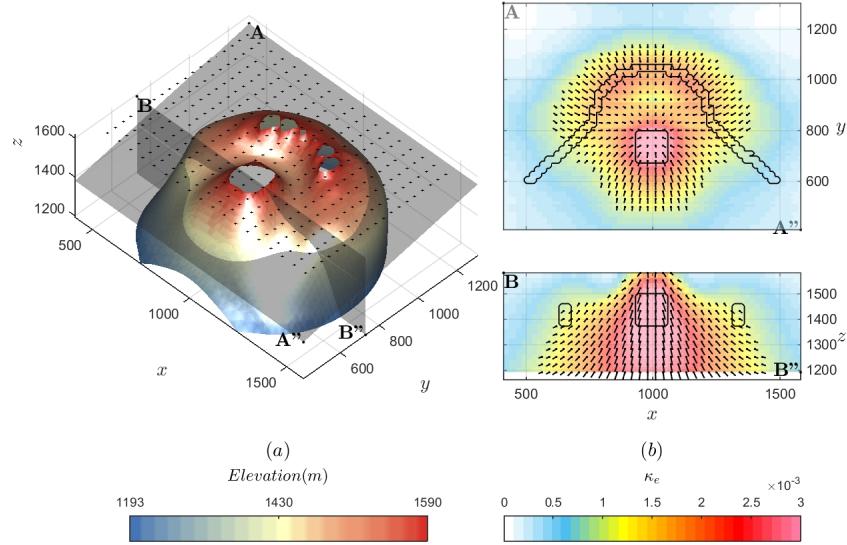


Figure 3.10: (a) Iso-surface ($\kappa_e = 0.001$) and (b) sections through the recovered magnetization model from the MVI method. The inversion recovers the true orientation of magnetization inside the block, but the thin arc is poorly resolved.

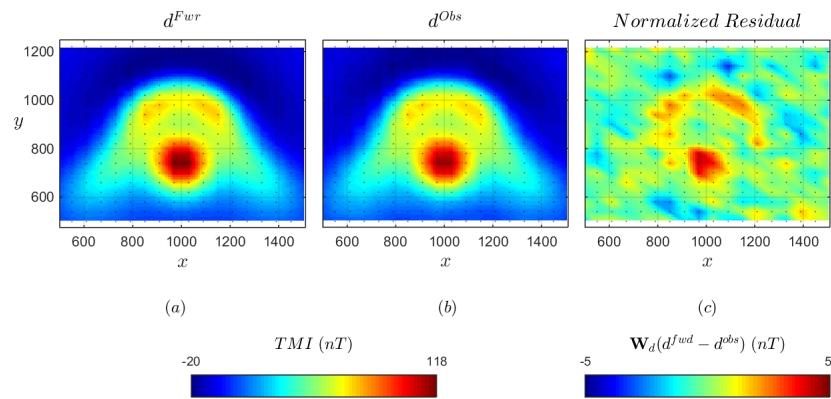


Figure 3.11: Comparison between (a) observed and (b) predicted data from the recovered magnetization model from the MVI method. The model can replicate the data at the same level achieved by the purely induced problem.

3.4 Magnetic amplitude inversion

The inversion of magnetic quantities, weakly dependent on the magnetization direction, have been proposed as an alternative to the full magnetic vector inversion approach (Li et al., 2010). In this section I review the work done by Shearer (2005) in inverting magnetic amplitude data. Amplitude data have the interesting property of being weakly dependent on the orientation of magnetization (Nabighian, 1972), which can be used to get an estimate for the location of magnetized material. I illustrate the idea with a single vertical dipole located at the origin. The magnetic field \vec{b} measured at some distance \vec{r} can be written as:

$$\vec{b} = \frac{3\mu_0 m}{4\pi|r|^3} \left(\sin \phi \cos \phi [\cos \theta \hat{x} + \sin \theta \hat{y}] + (\cos^2 \phi - \frac{1}{3}) \hat{z} \right), \quad (3.39)$$

where θ is the angle on the xy -plane, ϕ is the angle with respect to the z -axis and m is the dipole moment. Computing the magnitude of the field yields:

$$\begin{aligned} |\vec{b}| &= \sqrt{b_x^2 + b_y^2 + b_z^2} \\ &= \frac{\mu_0 m}{4\pi|r|^3} \sqrt{3 \cos^2(\phi) + 1}. \end{aligned} \quad (3.40)$$

The amplitude of the field $|\vec{b}|$ is strongly dependent on the radial distance $|r|$, but weakly dependent on ϕ , the inclination of the dipole. If we were to collect magnetic field data at a constant height above the dipole, the location of maximum amplitude would roughly occur above the location of the dipole, and the amplitude of the field would vary by at most a factor two. Hence amplitude data may reduce some of the ambiguity related to the horizontal distribution of magnetized material.

Recall from Chapter 2, I have defined the component of the magnetic field at some location P due to a prism with uniform magnetization \vec{M} as:

$$\begin{aligned} b_x &= [T_{xx} T_{xy} T_{xz}] \vec{M} \\ b_y &= [T_{yx} T_{yy} T_{yz}] \vec{M} \\ b_z &= [T_{zx} T_{zy} T_{zz}] \vec{M}, \end{aligned}$$

where the forward operator \mathbf{T} maps the components of the magnetic field due to a

distribution of magnetized prisms \vec{M} . Just as in the MVI method, the magnetization direction is unknown in most cases. The advantage of working with amplitude data is that the magnetization direction does not have to be known exactly in order to get a good estimate of $|\vec{b}|$. Under the assumption that the magnetization vector is in most part parallel to the geomagnetic field direction, equation 2.22 can be simplified to:

$$\begin{aligned}\vec{M} &= \kappa(\vec{H}_0 + \vec{H}_s) + \vec{M}_{rem} \\ &\approx \vec{H} \kappa_e ,\end{aligned}$$

where the effective susceptibility κ_e is a unitless quantity representing the total magnetization of each cell. Assuming a uniform inducing field \vec{H} , the anomalous magnetic field measurement due to a magnetized cell is approximated by:

$$\begin{aligned}\vec{b}(P) &= \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix} = \begin{pmatrix} T_{xx} & T_{xy} & T_{xz} \\ T_{yx} & T_{yy} & T_{yz} \\ T_{zx} & T_{zy} & T_{zz} \end{pmatrix} \begin{bmatrix} \kappa_e H_x \\ \kappa_e H_y \\ \kappa_e H_z \end{bmatrix} \\ &= \begin{bmatrix} F_x \\ F_y \\ F_z \end{bmatrix} \kappa_e .\end{aligned}\tag{3.41}$$

Note the parallel with the κ_e used in the MVI method to describe the magnetization components \hat{p}, \hat{s} and \hat{t} . Because the geomagnetic field is generally much larger than any secondary fields, I approximate the magnetization direction to be parallel to the inducing field ($\vec{H} \simeq \vec{H}_0$). From the linear relation described above, I write the forward calculation of magnetic amplitude data as:

$$\begin{aligned}|\vec{b}| &= \left[b_x^2 + b_y^2 + b_z^2 \right]^{1/2} \\ &= \left[(F_x \kappa_e)^2 + (F_y \kappa_e)^2 + (F_z \kappa_e)^2 \right]^{1/2} .\end{aligned}\tag{3.42}$$

Taking the partial derivative in terms of model parameter κ_e yields:

$$\begin{aligned}\frac{\partial |\vec{b}|}{\partial \kappa_e} &= \frac{\partial}{\partial \kappa_e} [b_x^2 + b_y^2 + b_z^2]^{1/2} \\ &= \frac{1}{|\vec{b}|} \left[b_x \frac{\partial b_x}{\partial \kappa_e} + b_y \frac{\partial b_y}{\partial \kappa_e} + b_z \frac{\partial b_z}{\partial \kappa_e} \right] \\ &= \frac{1}{|\vec{b}|} [b_x F_x + b_y F_y + b_z F_z].\end{aligned}\quad (3.43)$$

The inverse problem is clearly non-linear with respect to the model parameter κ_e . A solution is found iteratively as outlined in section 3.1.1. At the k^{th} iteration, the sensitivity relating the i^{th} amplitude data due to the j^{th} prism is given by:

$$J_{ij} = \frac{\vec{b}^{(k)}}{|\vec{b}^{(k)}|} \cdot \begin{bmatrix} F_x \\ F_y \\ F_z \end{bmatrix}, \quad (3.44)$$

where the magnetic data $\vec{b}^{(k)}$ is computed from an effective susceptibility found at a previous iteration. In order to have a Jacobian defined at the first iteration, I choose a small number larger than zero for the starting model ($\kappa^{(0)} = 1e-4$).

In mineral exploration, amplitude data must be derived directly from TMA data, which will be covered in Section 3.5. For this synthetic example, amplitude data are generated directly from 3.42 as I already know the true magnetization model. Data are corrupted with the same Gaussian noise, 1 nT standard deviation. Once again, positivity constraints are enforced on values of effective susceptibility.

Figure 3.12 presents the recovered effective susceptibility model after reaching the target misfit. Compared to the result presented in Figure 3.8, high κ_e values are recovered along the arc, and at the right depth inside the center block. The amplitude model closely resembles the model obtained from the purely induced response obtained in Section 3.2.1. I note however that the amplitude inversion has the tendency to stretch the model vertically. This remains an open question.

Comparing the observed and predicted data, the magnetic amplitude inversion fits the data generally well, within one standard deviation as shown in Figure 5.4. Once again, the highest residuals are correlated with the location of the anomaly,

which will be tackled in Chapter 4.

While neither the MVI or amplitude inversion managed to recover the location and orientation of magnetization exactly, both methods brought complementary information. From the MVI method, I am recovering a better estimate of the magnetization direction. With the amplitude inversion, I get a closer estimate of the true location of magnetic anomalies. The next chapter explores different regularization functions in order to further reduce the non-uniqueness of the MVI method. My goal is to combine those methods into a cooperative inversion work-flow in order to a recover a simpler and more accurate solution.

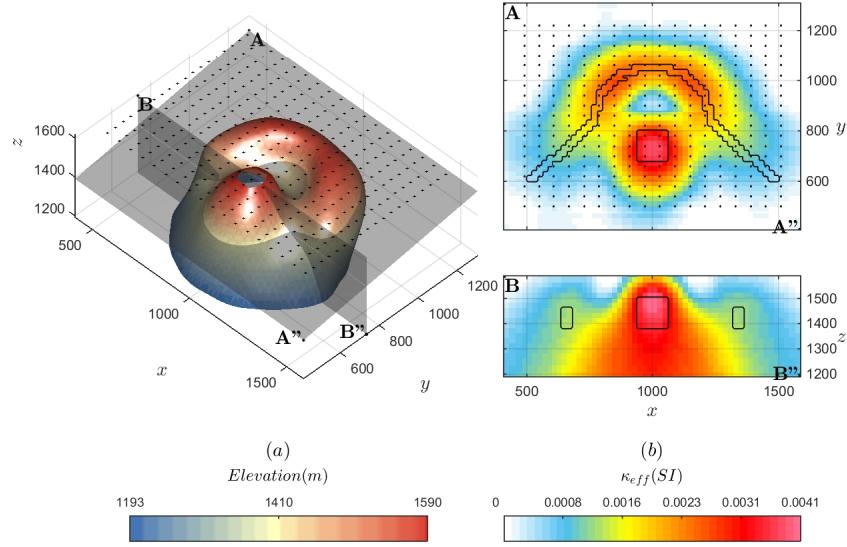


Figure 3.12: (a) Iso-surface (0.002 SI) and (b) sections through the recovered effective susceptibility model. The arc-shaped and block anomalies are recovered at the right location, but smoothly stretched vertically.

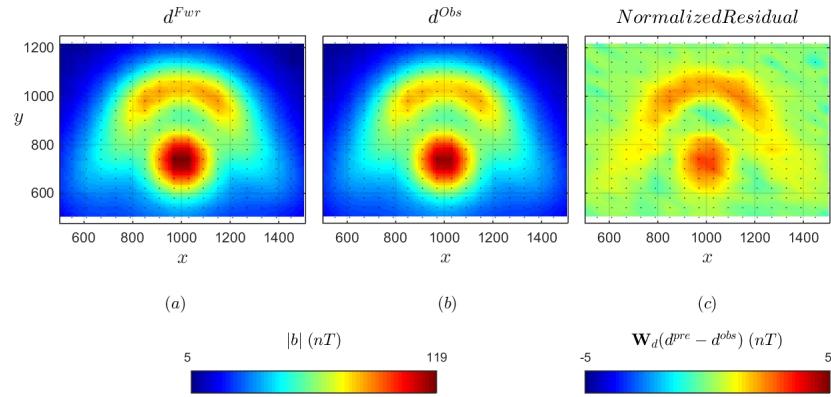


Figure 3.13: Comparison between (a) observed and (b) predicted data from the recovered effective susceptibility model. The inversion can predict most of the data within one standard deviation.

3.5 Equivalent source method

The magnetic amplitude inversion method introduced in 3.4 requires magnetic field components $[b_x, b_y, b_z]$ in order to calculate amplitude data such that:

$$|\mathbf{b}| = \left[\mathbf{b}_x^2 + \mathbf{b}_y^2 + \mathbf{b}_z^2 \right]^{1/2}. \quad (3.45)$$

While technically possible to measure three-component magnetic data, the vast majority of current and past magnetic surveys consist of TMI measurements. This is largely due to the difficulty in determining the location and orientation of three-component receivers. Two methods have dominated the literature in order to extract vector components directly from TMI data, either in the frequency domain or by the equivalent source method.

As demonstrated by Bhattacharyya (1964), TMI data can be expressed as a 2-D Fourier series expansion of the form:

$$\begin{aligned} \tilde{\mathbf{b}}(x, y, z) = & \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} e^{-2\pi z \left(\frac{m^2}{L_x^2} + \frac{n^2}{L_y^2} \right)^{1/2}} \\ & \left(A_m \cos 2\pi m \frac{x}{L_x} + B_m \sin 2\pi m \frac{x}{L_x} \right) \\ & \left(C_n \cos 2\pi n \frac{y}{L_y} + D_n \sin 2\pi n \frac{y}{L_y} \right), \end{aligned} \quad (3.46)$$

where L_x and L_y are the fundamental wavelengths in the x and y directions. The coefficients A_m, B_m, C_n and D_n are computed from the data for each wavenumber m and n . The harmonic representation of the data in (3.46) can be seen as a product of two functions: an exponential function controlling the amplitude of the signal and an harmonic function for the spatial distribution.

Converting data from the spatial to the wavenumber domain requires two important assumptions: that the data are located on a plane and distributed over a uniform grid. In most cases however, magnetic surveys are carried along unevenly spaced grids and over rugged topography. Even if acquired on a plane above a flat topography, the data need to be interpolated and smoothed. The transformation becomes even more complicated when data are collected at different elevations, as is often the case for airborne surveys with overlapping flight lines.

The Equivalent Source method has been suggested as an alternative to horizontal gridding methods (Dampney, 1969). The technique makes use of the inherent ambiguity of potential fields in determining the source location. It can be shown that any magnetic response can be explained by an arbitrary distribution of sources. In discrete form, an equivalent source model \mathbf{m}_{es} can be calculated from the least-squares problem:

$$\min_m \|\mathbf{F} \mathbf{m}_{\text{es}} - \mathbf{d}\|_2 . \quad (3.47)$$

As demonstrated by Dampney (1969) and later revised by Li and Oldenburg (2010), the distance between the data and the equivalent source is limited by the frequency content of the signal. Most recent work by Li et al. (2014) addresses striation artifacts when applying a reduction to the pole at low-latitude. They advocate for a positivity constraint formulation and prove the existence of an all-positive equivalent source layer in 2-D.

I replicate the synthetic example presented in Li et al. (2014) as shown in Figure 3.14. The model consists of 200 unit cubes with susceptibility of 0.01 SI placed in a non-susceptible background. Data are generated on a plane exactly 1 unit above the anomaly, assuming a purely vertical inducing field of 35,000 nT. Random Gaussian noise with a standard deviation of 1 nT is added to the field components, from which TMI and amplitude data are calculated. Using the formulation in Li et al. (2014), I invert for an equivalent source with a positivity constraint. The equivalent source layer is placed at a depth that is half the data spacing, in this case at half a unit below the data plane. Figure 3.15 presents the equivalent source layer, as well as the residual between the observed and predicted data. All three components of the field, and consequently $|\mathbf{b}|$ data, are well recovered within the noise level. Note however that some of the high frequency content related to the edges of the anomaly is lost in the process.

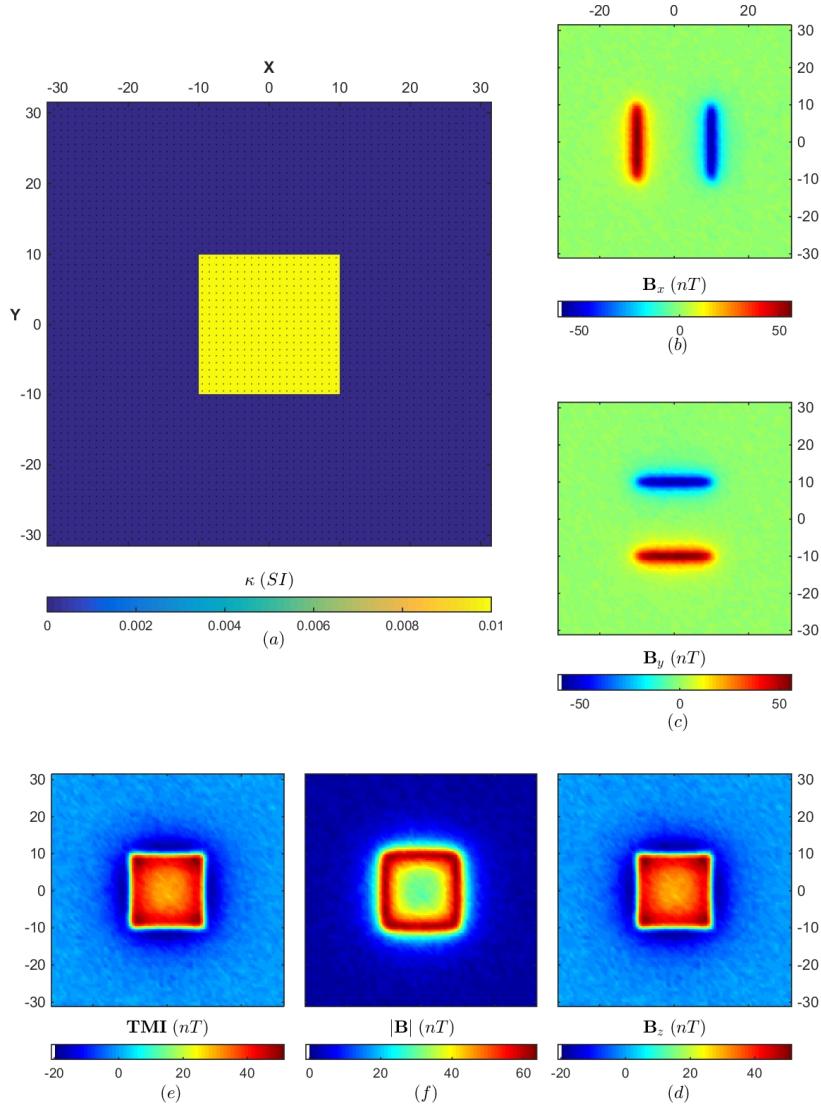


Figure 3.14: (a) Synthetic model consisting of 200 unit cubes of susceptible material in a non-susceptible background. Data are generated on a plane one unit above the source location, assuming a purely vertical inducing field. Various components of the fields are shown in figure (b) to (f).

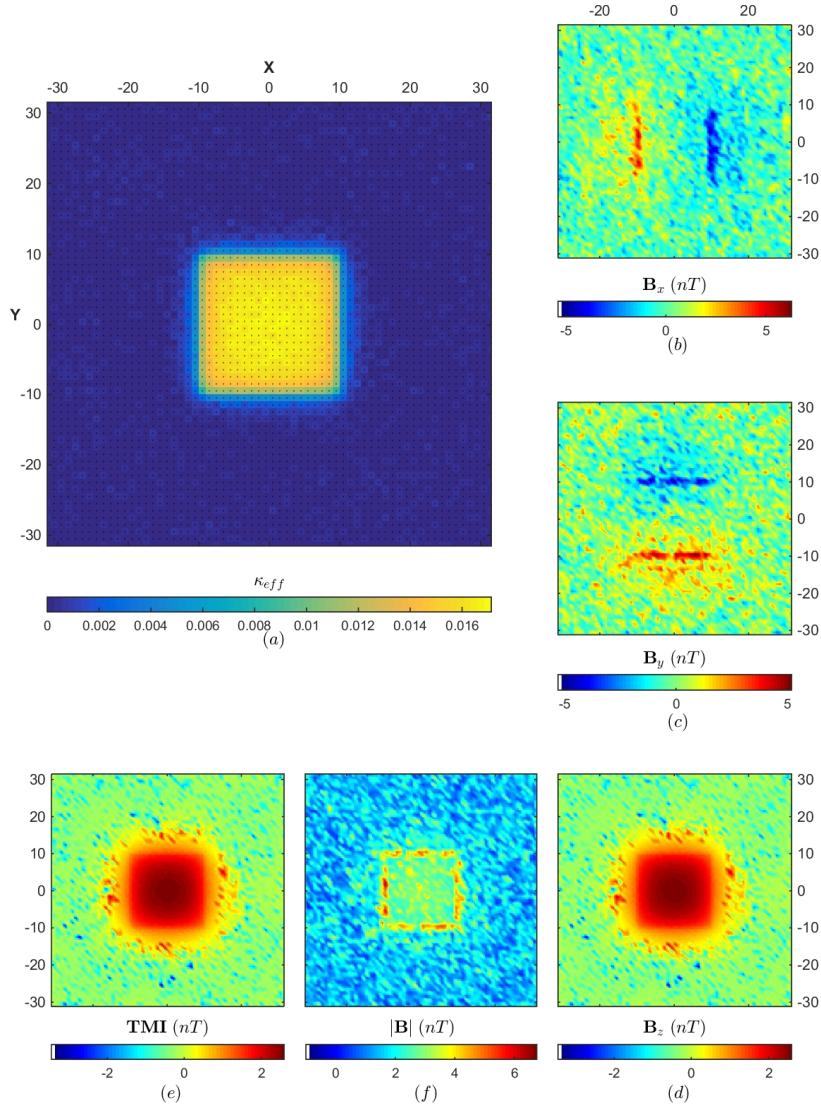


Figure 3.15: (a) Recovered equivalent source layer from TMI data using a positivity constraint. The residuals between observed and predicted data are shown in figure (b) to (f) for various components of the field. Each component is well recovered within the noise level.

3.5.1 Comment for future research

While testing the equivalent source method, I experimented on different data distributions to test the stability of the algorithm. Issues arose in cases where data coverage would partially cover the magnetic anomaly as illustrated in Figure 3.16. In this example, a portion of data are removed over one of the corners of the magnetic source. An equivalent-source layer is then calculated with the same parameters that were previously used for Figure 3.15. Large correlated artifacts are created on the \mathbf{b}_x and \mathbf{b}_y component of the field along the missing portion of the data. Consequently, a strong narrow anomaly is recovered on $|\mathbf{b}|$ with amplitude in the 40 nT range, well above the noise level. The same experiment was repeated for various inducing field orientations, transferring the correlated artifacts to other components of the field accordingly. This suggests some level of ambiguity in the components of the field if the anomalous response is not fully captured by the data. Such artifacts, if ignored during the inversion process, can have notable consequences on the effective susceptibility model. It is to my knowledge the first time that this issue is raised, and it will require further research. For the remainder of this research project, I continue using the equivalent source method of Li et al. (2014), but I take special care in removing edge data with large magnetic anomalies.

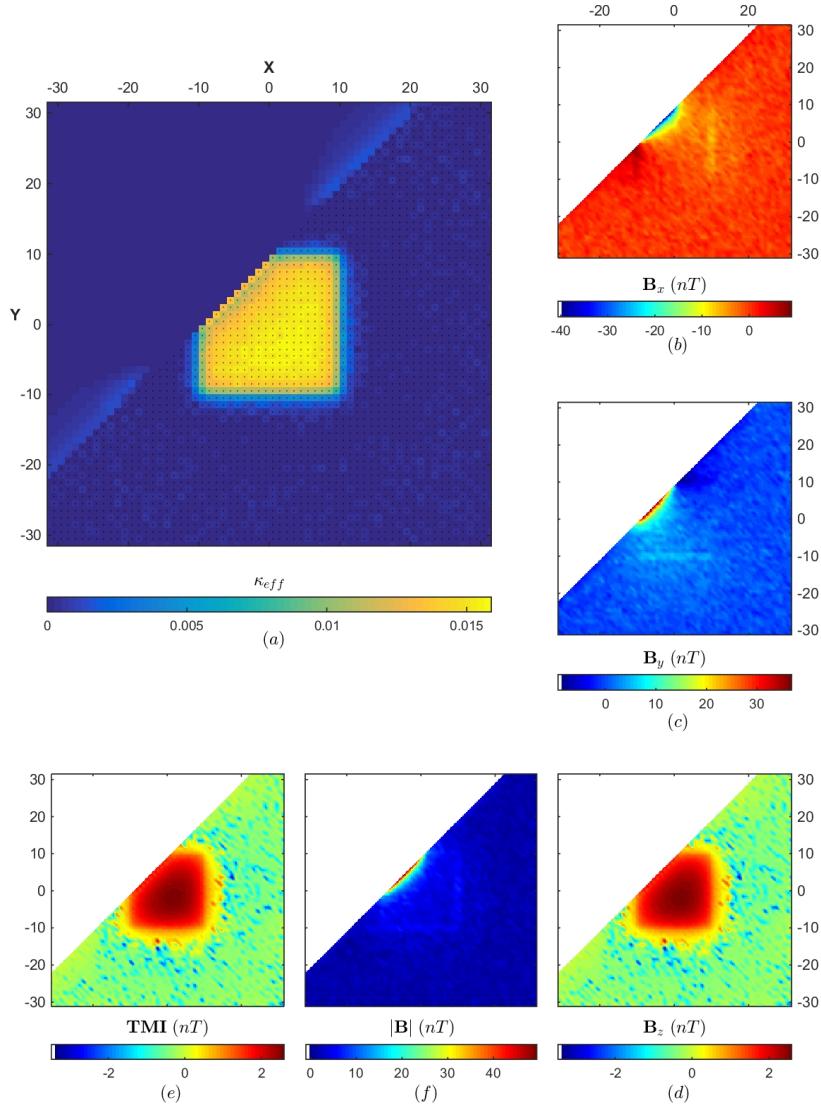


Figure 3.16: (a) Recovered equivalent source layer from TMI data after removing a portion of data over the corner of the magnetic anomaly. The residuals between observed and predicted data are shown in figure (b) to (f) for various components of the field. Note the large correlated artifacts recovered on the \mathbf{b}_x , \mathbf{b}_y and $|\mathbf{b}|$ components.

Chapter 4

Mixed L_p -norm Regularization

As presented in Chapter 3, magnetic inverse problems are generally formulated as a regularized least-squares problem. The regularization has the dual purpose of improving the conditioning of the linear systems as well as to impose constraints on the solution. In this chapter, I explore different norm measures in order to impose sparse and blocky constraints on the solution. I introduce a new regularization algorithm based on the Iterative Re-weighted Least-Squares (IRLS) method to improve the flexibility and robustness of current inversion methods. The algorithm is tested in 1-D, 2-D and 3-D on various synthetic examples. Finally, I implement the algorithm on an airborne magnetic survey over the Tli Kwi Cho kimberlite deposit.

4.1 Least-norm and regularized least-squares

As a general introduction to inverse problems, I consider the linear system of equation of the form:

$$\mathbf{F} \mathbf{m} = \mathbf{d}, \quad (4.1)$$

where the discretized forward operator $\mathbf{F} \in \mathbb{R}^{N \times nc}$ is a linear system of equations relating a set of model parameters $\mathbf{m} \in \mathbb{R}^{nc}$ to the observed data $\mathbf{d} \in \mathbb{R}^N$. From a geophysical standpoint, we are interested in solving the inverse problem:

$$\mathbf{m} = \mathbf{F}^{-1} \mathbf{d}, \quad (4.2)$$

where a set of unknown model parameters can be recovered from the observed data. In most cases, the inverse of \mathbf{F} cannot be computed directly as $N \ll nc$, giving rise to an undetermined system of equations. There are an infinite number of possible solutions satisfying 4.1, which corresponding to the null-space of \mathbf{F} , with $(nc - N)$ degrees of freedom. The choice of a specific answer is subjective and depends on characteristics expected from the true model.

One simple option would be to find the smallest possible model that also minimizes the data residual. The *least-norm* solution \mathbf{m}_{ln} can be found from:

$$\mathbf{m}_{\text{ln}} = \mathbf{F}^T (\mathbf{F} \mathbf{F}^T)^{-1} \mathbf{d}, \quad (4.3)$$

where \mathbf{m}_{ln} marks the point of shortest distance between the null-space of \mathbf{F} and the origin such that:

$$\mathbf{m}_{\text{ln}} \perp \mathcal{N}(\mathbf{F}). \quad (4.4)$$

An equivalent answer can be found via an optimization problem of the form:

$$\min_m \phi(m) \quad (4.5)$$

$$\begin{aligned} \phi(m) &= \phi_d + \beta \phi_m \\ \phi_d &= \|\mathbf{F}\mathbf{m} - \mathbf{d}\|_2^2 \\ \phi_m &= \|\mathbf{m}\|_2^2, \end{aligned}$$

where ϕ is our objective function. As first introduced by Tikhonov and Arsenin (1977), a trade-off parameter β balances the relative importance between the data residual ϕ_d and the Euclidean norm of the model \mathbf{m} . The minimum of the objective function is found at the location of vanishing gradients such that:

$$\begin{aligned} \frac{\partial \phi(m)}{\partial m} &= 0 \\ (\mathbf{F}^T \mathbf{F} + \beta \mathbf{I}) \mathbf{m} &= \mathbf{F}^T \mathbf{d}, \end{aligned} \quad (4.6)$$

where \mathbf{I} is the identity matrix.

I illustrate those concepts with a simple 2-variable example where I attempt to

solve:

$$m_1 + 2 m_2 = 1 . \quad (4.7)$$

Equation 4.7 can be formulated as an under-determined linear system of the form:

$$\mathbf{F} = \begin{bmatrix} 1 & 2 \end{bmatrix}, \mathbf{d} = 1. \quad (4.8)$$

Figure 4.1(a) displays contours along the surface formed by the l_2 -norm measure of data residual $\phi_d = \|\mathbf{F}\mathbf{m} - \mathbf{d}\|_2^2$ over a range of model values \mathbf{m} . The misfit function can be seen as a trough with the minimum laying along the null space of \mathbf{F} . Any solution along the minimum of ϕ_d can satisfy 4.1. The least-norm solution is marked with a solid dot, which is the closest distance between the null space of \mathbf{F} and the origin. Similarly, Figure 4.1(b) depicts the l_2 -norm regularization function ϕ_m as a function of model values \mathbf{m} , forming a symmetric paraboloid centered at the origin.

The full objective function $\phi(m)$ shown in Figure 4.1(c) can be interpreted conceptually as the sum of two competing objectives. On one hand we want to find the best model reproducing the data, which in this case is any solution along the null-space of \mathbf{F} . On the other hand, we impose some constraints on the magnitude of \mathbf{m} in a l_2 -norm sense, pulling the solution towards zero. The shift between the least-norm solution and the regularized least-squares depends on the regularization parameter β . It can be shown that the solution to 4.6 converges to the least-norm solutions exactly as $\beta \rightarrow 0$ as shown in Figure 4.2(a). For a small enough β , the solution converges to the global minimum at $\mathbf{m}_{\text{ln}} = [0.2, 0.4]$, regardless of the starting model.

It is also important to note that the model norm regularization can be shifted away from the origin if we were to minimize:

$$\phi_m = \|\mathbf{m} - \mathbf{m}^{\text{ref}}\|_2^2. \quad (4.9)$$

In this case, the minimum gradient of the objective function would occur at \mathbf{m}^{ref} . We would therefore look for a solution that both minimizes the data residual, while also approximating the expected model value. For simplicity, I consider the case where the reference model is at origin.

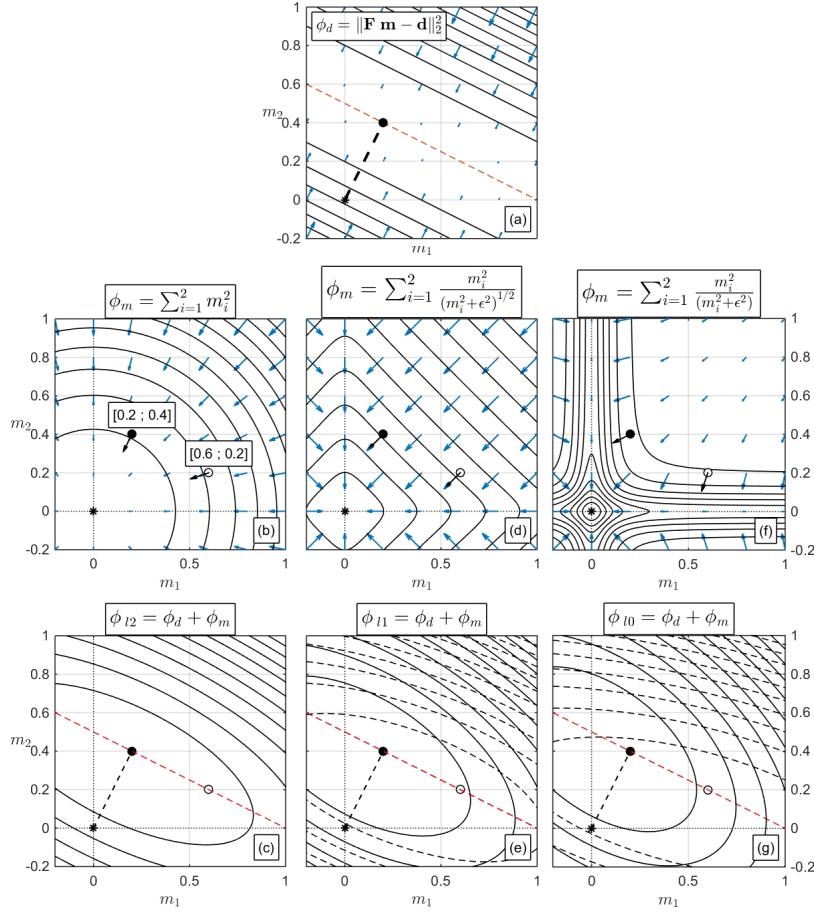


Figure 4.1: Comparative contour maps for various objective functions over a range of model values $[m_1; m_2]$. (a) The minimum of the misfit function ϕ_d forms a line spanned by $\mathcal{N}(\mathbf{F})$ (red dash). The *least – norm* solution is marked as a solid dot. (Middle) Regularization functions and gradient directions (blue arrows) for approximated l_p -norm measures of the model for (b) $p = 2$, (d) $p = 1$ and (f) $p = 0$. The gradient directions are shown for two different starting models (black arrows). (bottom) Contour maps of the initial objective functions $\phi(m) = \phi_d + \phi_m$ for the same two starting models: $\mathbf{m}_1^{(0)}$ (solid) and $\mathbf{m}_2^{(0)}$ (dash). (c) In the l_2 -norm case , the function has a global minimum regardless of the starting model, while for non-linear functions for (e) $p = 1$ and (g) $p = 0$, the objective function changes with respect to the starting model.

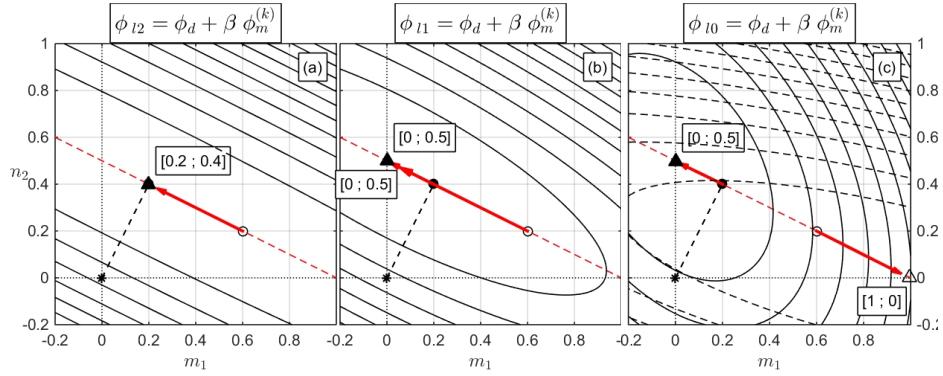


Figure 4.2: Contour maps for various objective functions after convergence of the IRLS algorithm. (a) Final model obtained with the l_2 -norm penalty on the model for two starting models at $\mathbf{m}_1^{(0)} = [0.2; 0.4]$ and $\mathbf{m}_2^{(0)} = [0.6; 0.2]$ for a fixed trade-off parameter ($\beta = 1e - 4$). In both cases, the solution converges to the global minimum, which is also the *least – norm* solution at $\mathbf{m}_{\text{ln}} = [0.2; 0.4]$. (b) Solution with the l_1 -norm penalty for the same starting models and trade-off parameter, converging to a global minimum at $\mathbf{m}^* = [0; 0.5]$. This solution is sparse and can reproduce the data. (c) The same experiment is repeated for the l_0 -norm penalty, converging to two different solutions depending on the relative magnitude of the starting model parameters. Both solutions are sparse and honor the data.

4.2 L_p -norm and iterative re-weighted least squares

I have so far only considered the Euclidean norm of the model parameters, penalizing the square of model values. Likewise, Li and Oldenburg (1996) include an l_2 -norm penalty on the model gradients, yielding smooth models. In some cases, the solution may be expected to be sparse and blocky, either in terms of model parameters or spatial gradients. There have been several studies dedicated to the use of non- l_2 measures to recover models with sharp edges. Most methods proposed in the literature made use of the globally convex l_1 -norm, with proven convergence to a global minimizer (Daubechies et al., 2010; Farquharson and Oldenburg, 1998; Sun and Li, 2014). Others have proposed approximations to the non-convex l_0 -

norm, methods such as the *compact* regularization of Last and Kubik (1983) and the *minimum support* functional of Portniaguine (1999), to name a few. My goal is to further generalize the proposed methods in order to explore a wide range of solutions for any combinations of l_p -norm penalty applied on the model and model gradients independently.

I begin with a generalized expression for the model objective function, such that (4.5) becomes:

$$\phi_m = \sum_{i=1}^{nc} \rho(x_i), \quad (4.10)$$

where ρ is some norm measure of a model function $\mathbf{x}(m)$ commonly chosen to be the model itself or some measure of spatial gradients. The general l_p -norm measure can be written as

$$\phi_m = \sum_{i=1}^{nc} |x_i|^p, \quad (4.11)$$

where for $p = 2$, I recover the standard regularized inversion presented in (4.5). Several approximations to the l_p -norm have been proposed, such as the Ekblom norm (Ekblom, 1973):

$$\phi_m = \sum_{i=1}^{nc} (x_i^2 + \varepsilon^2)^{p/2}, \quad (4.12)$$

where a small number ε is added to guarantee that the function is continuous and differentiable as $\mathbf{x} \rightarrow 0$. Figure 4.3(a) presents various norms for a fix threshold values ($\varepsilon = 1e-2$). The derivative of (4.12) is given by:

$$\begin{aligned} \frac{\partial \phi_m}{\partial m} &= \sum_{i=1}^{nc} \rho'(x_i) \frac{\partial x_i}{\partial m} \\ &= p \frac{x_i}{(x_i^2 + \varepsilon^2)^{1-p/2}} \frac{\partial x_i}{\partial m}. \end{aligned} \quad (4.13)$$

Expression (4.13) is clearly non-linear with respect to the model function x_i . As first introduced by Lawson (1961), the norm can be linearized by the *Iteratively Re-weighted Least-Squares* (IRLS) method such that:

$$\phi_m^{(k)} = \frac{1}{2} \sum_{i=1}^{nc} r_i x_i^2 \quad (4.14)$$

$$\frac{\partial \phi_m^{(k)}}{\partial m} = r_i x_i \frac{\partial x_i}{\partial m}, \quad (4.15)$$

where we added the superscript $\square^{(k)}$ to denote the IRLS iterations. The weights $r(x)$ are computed from model values obtained at a previous iteration such that:

$$r_i = \left((x_i^{(k-1)})^2 + \varepsilon^2 \right)^{p/2-1}, \quad (4.16)$$

where $r(x) \in \mathbb{R}^{nc}$.

The goal of the IRLS method is to approximate the l_p -norm by solving a series of locally convex least-squares problems. The general objective function to be minimized takes the form:

$$\begin{aligned} \phi(m) &= \phi_d + \beta \phi_m^{(k)} \\ &= \|\mathbf{F}\mathbf{m} - \mathbf{d}\|_2^2 + \beta \|\mathbf{R} \mathbf{x}(m)\|_2^2, \end{aligned} \quad (4.17)$$

where the diagonal matrix $\mathbf{R} \in \mathbb{R}^{nc \times nc}$ holds the IRLS weights such that:

$$R_{ii} = r_i^{1/2}. \quad (4.18)$$

At each k^{th} iteration, we seek a minimum along the gradient of the the objective function. Replacing the regularization function from equation 4.6 we get:

$$\begin{aligned} \mathbf{F}^T \mathbf{F} \mathbf{m} + \beta \mathbf{g}(x) &= \mathbf{F}^T \mathbf{d} \\ \mathbf{g}(x) &= \mathbf{R}^T \mathbf{R} \mathbf{x}(m) \frac{\partial \mathbf{x}(m)}{\partial m}, \end{aligned} \quad (4.19)$$

where I explicitly define the gradient of the approximated l_p -norm regularization function $\mathbf{g}(x)$. I voluntarily neglect the constant of differentiation p from equation (4.15) for two reasons. First, note that in the special case where $p = 0$, the regularization function would vanish and reduce 4.19 to a simple least-squares problem. Secondly, for any $p \neq 0$, the constant would simply get absorbed by the trade-off parameter β . Other parameters will be introduced in the following section to handle scaling issues arising from mixed-norm regularization functions.

Figure 4.3(b) and (c) presents the IRLS weights $\mathbf{r}(x)$ and gradient functions $\mathbf{g}(x)$ for a range of p -value. For $p = 2$ and $\mathbf{x}(m) = \mathbf{m}$, we obtain the smooth regu-

larization function presented in Figure 4.1(a). It is important to note that for a small l_p -norm (i.e $p < 1$), the IRLS weights and gradients rapidly increase as $x_i \rightarrow \varepsilon$. The behavior of the regularization function around the threshold parameter ε is important for reasons that will be addressed in Section 4.3.4.

I illustrate the various l_p -norms on the same 2-variable inverse problem presented in Section 4.1. Figure 4.1(d) and (e) show the regularization and the objective function for $p = 1$, over a range of model parameters. I point out that, compared to the globally convex l_2 -norm, the shape of the objective function now depends on the starting model $\mathbf{m}_1^{(0)}$ and $\mathbf{m}_2^{(0)}$.

For $p = 0$, I recover the *compact* regularization function put forward by Last and Kubik (1983), which has been borrowed by many researchers in geophysics Ajo-Franklin et al. (2007); Barbosa and Silva (1994); Stocco et al. (2009). Figure 4.1(f) and (g) presents the approximated l_0 -norm and corresponding objective function over the same range of model parameters. Similarly, the minimum of the objective function depends on the initial model $\mathbf{m}^{(0)}$ used to construct the regularization.

Although difficult to prove analytically, experiments have shown that the l_0 -norm can yield a sparser solution than the convex l_1 -norm (Chartrand, 2007). I use the same 2-variable example to illustrate this idea. For simplicity, I impose a fix threshold parameter ($\varepsilon = 1e-8$) and a fix trade-off parameter ($\beta = 1e-4$). Here, I am only interested in the final solution depending on the choice of l_p -norm, and as a function of starting model $\mathbf{m}^{(0)}$. In the next section, I will provide strategies to efficiently determine those parameters.

In the first experiment, I choose the starting model $\mathbf{m}_1^{(0)}$ to be the least-norm solution at $\mathbf{m}_{1n} = [0.2; 0.4]$, marked as a solid dot (Fig. 4.2). In both cases the l_1 -norm and l_0 -norm converge to the same optimal solution at $\mathbf{m}^* = [0.0; 0.5]$. The model \mathbf{m}^* is interesting as it is sparse with a single non-zero parameter, while also having the smallest norm possible.

In the second experiment, the initial model is also chosen to satisfy the target data misfit, but this time with a relatively smaller value on the second variable such that $\mathbf{m}_2^{(0)} = [0.6; 0.2]$, marked as a white dot. Because both the l_1 -norm and l_0 -norm penalize small model parameters, the regularization forces the solution to be sparse along $\mathcal{N}(\mathbf{F})$. Note the clear difference between the globally convex l_1 -

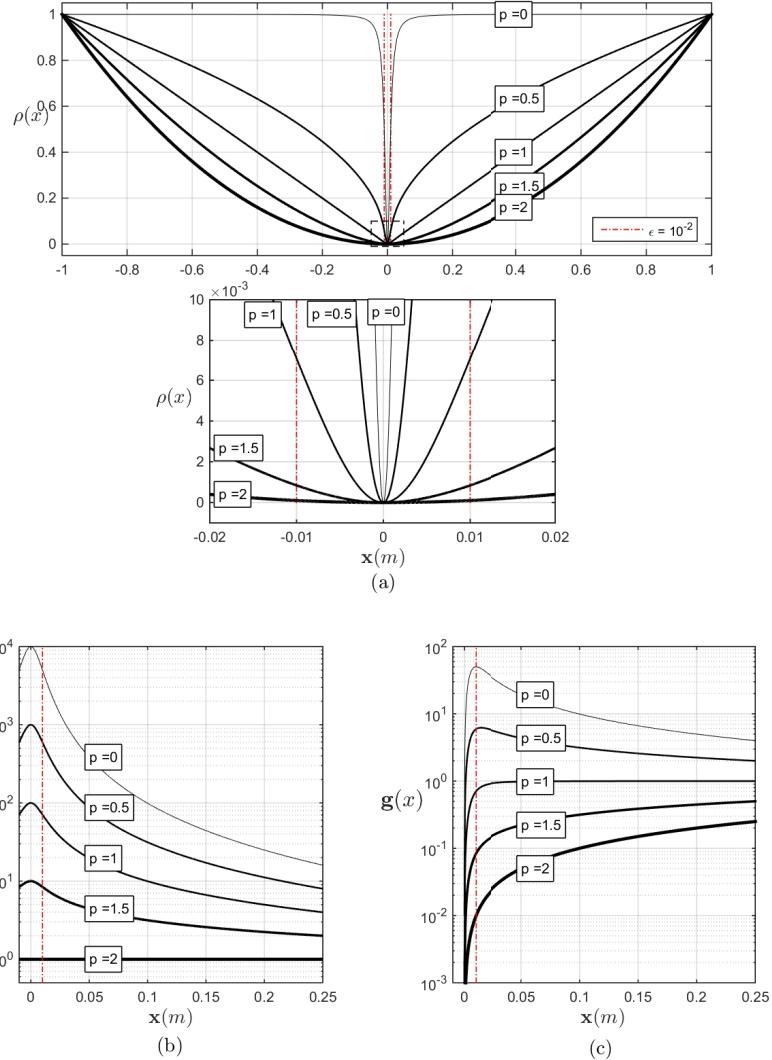


Figure 4.3: (a) Penalty function $\rho(x)$ for different approximate l_p -norm measures, and enlarged view near the region of influence of ϵ , making the l_p -norm continuous at the origin. (b) IRLS weights $\mathbf{r}(x)$ as a function of model function $\mathbf{x}(m)$ and p -values and for a fix threshold parameter ($\epsilon = 1e-2$). (c) Gradients $\mathbf{g}(x)$ of the model penalty function for various p -values. Note that the gradients are on a logarithmic scale due to the rapid increase as $x_i \rightarrow \epsilon$ for $p < 1$.

norm, converging back to \mathbf{m}^* , compared to the non-convex l_0 -norm, reaching the m_1 -axis at $\mathbf{m}^{(k)} = [1; 0]$. The l_0 -norm has therefore two possible solutions that do not depend on the overall magnitude of $\mathbf{m}^{(0)}$, but rather on the *relative* magnitude of its components $[m_1 ; m_2]$. The l_0 -norm acts as a binary selector, influenced by the largest elements of a starting model.

The choice of specific norm should therefore reflect the expected character of the solution, and the chosen algorithm should allow access to the full range of norms for $0 \leq p \leq 2$. While it is simple to solve (4.17) for a 2-variable problem, finding a solution to large systems for non-convex norms ($p < 1$) has proven to be difficult and remains a field of active research. The following sections review potential complications for the non-convex cases, and review strategies to efficiently solve the IRLS method.

4.3 IRLS solver

The IRLS method presented in 4.17 is an iterative process, which depends on two key parameters: the trade-off parameter β and the threshold parameter ε . In this section, I explore four strategies for the implementation of the IRLS.

4.3.1 Basic IRLS algorithm

As illustrated with the previous 2-variable problem, the choice of a starting model $\mathbf{m}^{(0)}$ is especially important for non-convex norms $p < 1$. According to Chartrand (2007), it may be possible to compute a global minimizer from non-convex functions, as long as the initial model is close enough to the global optimum. Most methods proposed in the literature seem to agree on the l_2 -norm solution as a valid candidate (Ajo-Franklin et al., 2007; Portniaguine and Zhdanov, 2002; Sun and Li, 2014). Current inversion strategies already rely on the assumption that a smooth model is a good approximation of the true solution. Applying a sparse l_p -norm can then segregate the most important model parameters, which in turn can reduce the complexity of the solution. This is mostly interesting if the true solution is known to behave like a delta function.

The basic IRLS algorithm becomes a two step process, as summarized in Table 4.1. During Phase-I, the algorithm finds a smooth solution with the l_2 -norm

regularization. The trade-off parameter β is monotonically reduced until the model can predict the data near the target misfit ϕ_d^* . The final l_2 -norm solution provides a starting model $\mathbf{m}^{(0)}$ for the calculation of a weighting matrix \mathbf{R} .

In Phase-II of the IRLS method, model updates are computed iteratively by minimizing 4.17. This process is repeated until the algorithm converges to a stable solution. I define a *convergence* criteria as:

$$\delta\phi_m^{(k)} = \frac{|\phi_m^{(k)} - \phi_m^{(k-1)}|}{\phi_m^{(k)}} \times 100\%, \quad (4.20)$$

where the change in model norm falls below some pre-defined threshold, chosen to be 2% in all my experiments. In the simplest form of the algorithm, the threshold ϵ and trade-off parameter β remain constant throughout Phase-II. For now, I will follow the general consensus that ϵ has to be small, or near machine error ($\epsilon_p = \epsilon_q = 1e-8$) (Ajo-Franklin et al., 2007; Last and Kubik, 1983; Stocco et al., 2009). I will revisit this number in Section 4.3.4. In the method proposed by (Ajo-Franklin et al., 2007), a trade-off parameter β^* is chosen so that the initial IRLS solution predicts the data near the target misfit ϕ_d^* . This β^* then remains constant throughout the iterative process.

Table 4.1: IRLS Solver 1: Fix parameters

Phase I - L_2-norm iterations: $\{\mathbf{m}^{(0)} \mid \phi_d \simeq \phi_d^*\}$
Adjust β
$\min \phi(\mathbf{m}) \rightarrow \mathbf{m}^{(0)}$
Search: $\{\beta^* \mid \phi_d^{(k)} \simeq \phi_d^*\}$
Phase II - L_p-norm iterations: $\{\mathbf{m}^{(k)} \mid \delta\phi_m^{(k)} < 1\%\}$
Fix $\{\beta^*, \epsilon_p, \epsilon_q\}$
Update $r_i^{(k)}$
$\min \phi^{(k)} \rightarrow \mathbf{m}^{(k)}$

4.3.2 1-D synthetic example

I demonstrate this simple IRLS implementation on a synthetic 1D problem similar to the problem used by Li and Oldenburg (1996). The model consists of a rectangular pulse and a Gaussian function on the interval [0 1] and discretized with 200 uniform intervals. Data are generated from the equation:

$$d_j = \int_0^1 f_j(x)m(x) dx, \quad j = 0, \dots, N, \quad (4.21)$$

where the kernel functions relating the model and the data are defined as:

$$f_j(z) = e^{-jx} \cdot \cos(2\pi jx), \quad (4.22)$$

where x defines the distance along the x -axis.

The model and kernel functions for $j \in [1, 30]$ are shown in Figure 4.4. Two percent random noise is added to the data in order to simulate a true geophysical experiment. The data are weighted accordingly such that:

$$\phi_d = \|\mathbf{W}_d (\mathbf{F} \mathbf{m} - \mathbf{d})\|_2^2, \quad (4.23)$$

where the diagonal matrix \mathbf{W}_d holds the estimated uncertainties associated with each datum. Since the noise is assumed to be Gaussian and uncorrelated, the misfit function follows a chi-squared distribution with expected value of N , hence the target misfit.

I here generalize the objective function presented in Chapter 3 to allow for multiple l_p -norm regularization functions such that:

$$\phi(m) = \phi_d + \beta \left[\int_0^1 w_s |m - m^{ref}|^p dl + \int_0^1 w_x \left| \frac{\partial m}{\partial x} \right|^q dl \right]. \quad (4.24)$$

The only difference with the regularization function used in 3.6 is in the norms applied on the model and model gradients $|\cdot|^p$ and $|\cdot|^q$, where p and q can take any value on the interval $[0, 2]$. Since we are dealing with a 1-D problem, model gradients are only computed along the x -axis. I also apply a depth weighting function

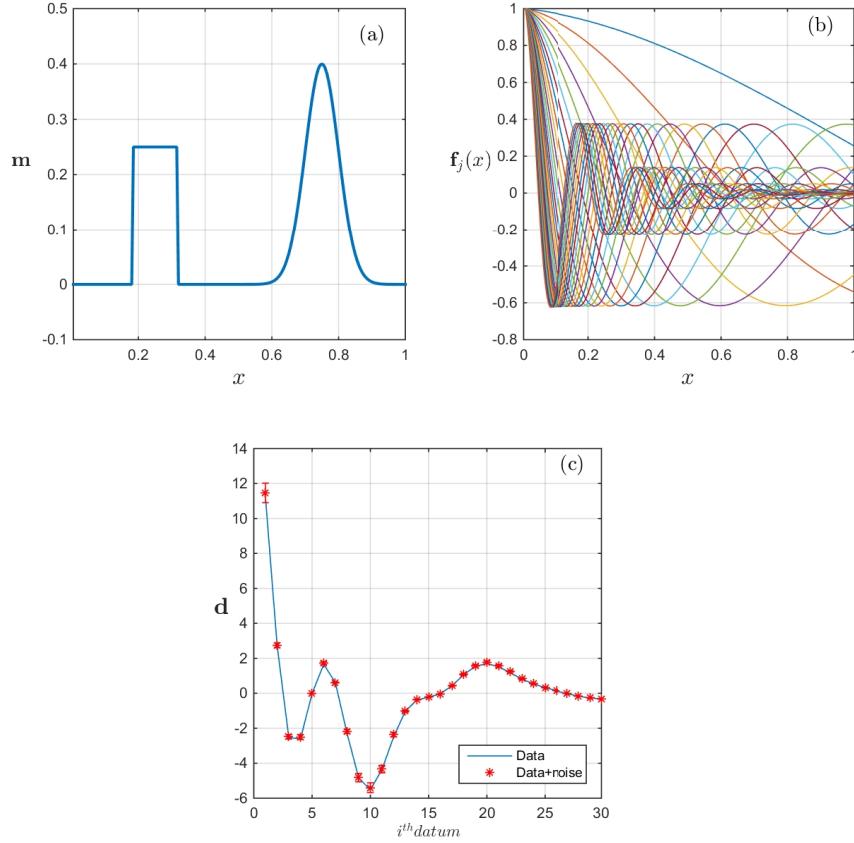


Figure 4.4: (a) Synthetic 1D model made up of a rectangular pulse and a Gaussian function. (b) Kernel functions consisting of exponentially decaying cosine functions of the form $f_j(x) = e^{-jx} \cdot \cos(2\pi jx)$. (c) Data generated from $\mathbf{d} = \mathbf{F} \mathbf{m}$, with 5% random Gaussian noise added.

to counter the decay of the kernel functions such that:

$$\begin{aligned} w_s &= \alpha_s e^{jx} \\ w_x &= \alpha_x e^{jx}. \end{aligned} \tag{4.25}$$

Combining the IRLS approximation in (4.17) and the discrete objective function

presented in 3.8, the linear 1-D equation 4.24 becomes:

$$\phi(m) = \|\mathbf{W}_d (\mathbf{F} \mathbf{m} - \mathbf{d})\|_2^2 + \beta \left[\|\mathbf{W}_s \mathbf{R}_s (\mathbf{m} - \mathbf{m}^{\text{ref}})\|_2^2 + \|\mathbf{W}_x \mathbf{R}_x \mathbf{G}_x \mathbf{m}\|_2^2 \right], \quad (4.26)$$

where the diagonal matrices $\mathbf{W}_s, \mathbf{W}_x$ are the cell weights, and \mathbf{G}_x is the spatial gradient operator presented in Chapter 3. The IRLS weights \mathbf{R}_s and \mathbf{R}_x are defined as:

$$\begin{aligned} R_{sii} &= \left[(m_i^{(k-1)})^2 + \varepsilon_p^2 \right]^{(p/2-1)/2} \\ R_{xii} &= \left[\left(\frac{\partial m_i^{(k-1)}}{\partial x} \right)^2 + \varepsilon_q^2 \right]^{(q/2-1)/2}, \end{aligned} \quad (4.27)$$

where $\varepsilon_p, \varepsilon_q$ are the stabilizing parameters for the sparsity constraint applied on the model and model gradient respectively.

As explained in Section 3.1.1, the minimum norm solution is found where $\frac{\partial \phi(m)}{\partial m} = 0$. Taking the partial derivatives of 4.26 with respect to the model parameters and setting the reference model to zero yields:

$$\left(\mathbf{F}^T \mathbf{W}_d^T \mathbf{W}_d \mathbf{F} + \beta [\mathbf{R}_s^T \mathbf{W}_s^T \mathbf{W}_s \mathbf{R}_s + \mathbf{G}_x^T \mathbf{R}_x^T \mathbf{W}_x^T \mathbf{W}_x \mathbf{R}_x \mathbf{G}_x] \right) \mathbf{m} = \mathbf{F}^T \mathbf{W}_d^T \mathbf{W}_d \mathbf{d}. \quad (4.28)$$

This linear system can be expressed as an overdetermined problem of the form:

$$\begin{bmatrix} \mathbf{W}_d \mathbf{F} \\ \sqrt{\beta} \mathbf{W}_s \mathbf{R}_s \\ \sqrt{\beta} \mathbf{W}_x \mathbf{R}_x \mathbf{G}_x \end{bmatrix} \mathbf{m} = \begin{bmatrix} \mathbf{W}_d \mathbf{d} \\ 0 \\ 0 \end{bmatrix}. \quad (4.29)$$

Solving the least-squares problem 4.29 yields a model update at the k^{th} iteration of the IRLS method. In this case, the left-hand side of 4.28 is linear with respect to $\mathbf{m}^{(k)}$ and it forms a symmetric positive definite matrix, which can be solved efficiently by the Conjugate Gradient descent method.

For the initial phase of the IRLS method, a solution is found with the globally convex l_2 -norm regularization, in which case \mathbf{R}_s and \mathbf{R}_x reduce to the identity matrix. Figure 4.5(a) presents the inverted result obtained with the l_2 -norm, as well

as the convergence curve after achieving the target data misfit ϕ_d^* . As expected from an l_2 -norm regularization, the solution is smooth and dispersed over the entire model domain.

From there, the algorithm proceeds with a sequence of IRLS iterations with l_0 -norm penalty on the model gradient ($\alpha_s = 0, q = 0$). The goal is to use the regularization to enforce a solution with sharp gradients and hopefully recover the rectangular pulse. The IRLS weights are initiated with the smooth l_2 -norm solution. The initial β^* is found by a search method where a solution to equation 4.29 is computed for multiple trials. The iterative process is repeated until the inversion reaches the convergence criteria specified by equation 4.20, while keeping the trade-off parameter β^* constant.

As shown in Figure 4.5(b), the updated solution with the l_0 -norm penalty recovers a blocky model with sharp edges. It is important to note that the final data residual $\phi_d^{(k)}$ is much larger than the target misfit ($\phi_d^* = 30$). Clearly the influence of the regularization has overtaken the inversion process. The algorithm likely converged to some local minimum, far from the global optimizer previously found with the smooth l_2 -norm. Since I know the true solution, I can measure the accuracy of the solution, or model error as:

$$\|\delta m\|_1 = \sum_{i=1}^{nc} |m_i^{(k)} - m_i^*|, \quad (4.30)$$

where $\mathbf{m}^{(k)}$ is the solution found at the k^{th} iteration. The final model error is larger than the one found with the smooth l_2 -norm regularization, hence it is a poor estimate of the true solution.

This example clearly illustrates some of the challenges related to non-convex objective functions. We have so far only formulated the basic algorithm behind the IRLS method, which has been used by many researchers in the past. Important details regarding the stability and predictability of the algorithm will now be addressed.

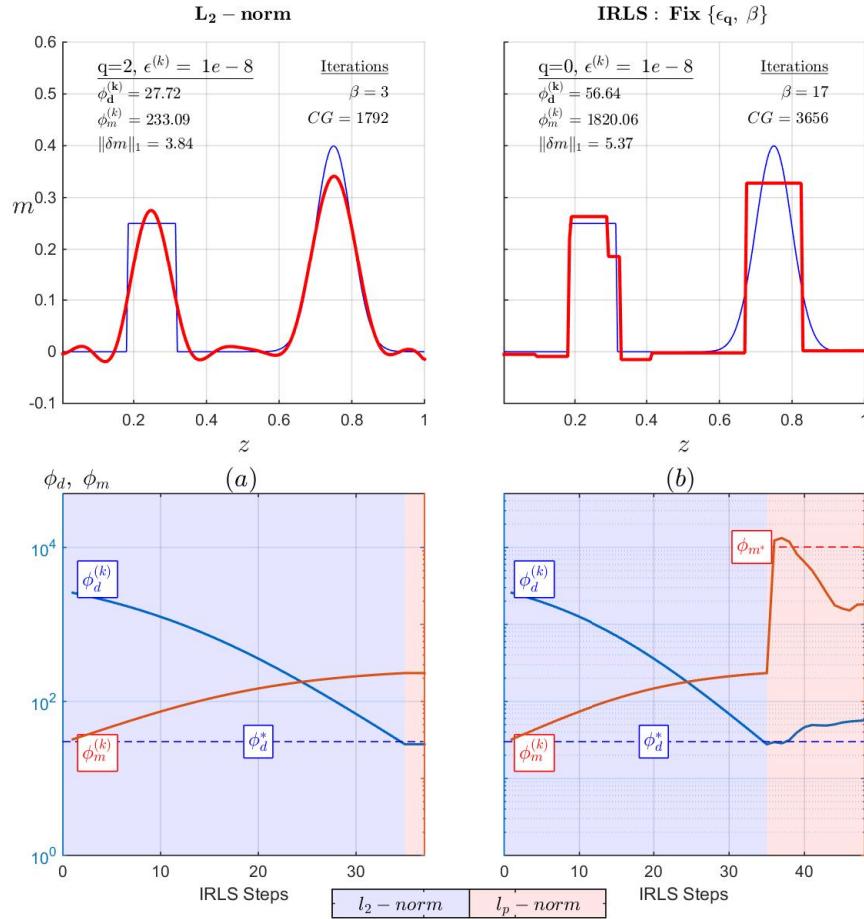


Figure 4.5: (a) Recovered models from the smooth l_2 -norm regularization, and (bottom) measure of model misfit and model norm as a function of iterations. Both the rectangular pulse and Gaussian function are recovered at the right location along the x -axis, but the solution is smooth and dispersed over the entire model domain. (b) Solution obtained from the IRLS with sparsity constraints on the model gradients ($q=0$), using the l_2 -norm model (a) to initiate the IRLS steps. The algorithm uses a fixed threshold parameter ($\epsilon = 1e - 8$) and fixed trade-off parameter β . The final solution is blocky, as expected from the norm chosen, but fails to achieve the target data misfit. Clearly the influence of the regularization function has overtaken the minimization process.

4.3.3 Regularization scaling

Following Tikhonov's approach, a solution to the inverse problem is found by progressively reducing the influence of the regularization until reaching the target data misfit ϕ_d^* . For strictly convex regularization functions, such as the l_2 -norm, the influence of the regularization functions is scaled linearly by the trade-off parameter β . For non-convex functions, in our case for $q = 0$, the iteration process involves non-linear transformations of the objective function, driven by the IRLS weights computed in (4.16). Those weights can change rapidly as $\varepsilon \rightarrow 0$, which directly impacts the influence of the model norm in a non-linear fashion. As demonstrated in Figure 4.5(b), special care must be taken in order to obtain a solution that satisfies the data while also being sparse.

As a second strategy, I experiment with a brute-force approach where the optimal trade-off parameter β is determined before each IRLS steps (Table 4.2). A solution to 4.19 is computed several times for a range of β -values until a suitable trade-off parameter is found. Figure 4.6(a) presents the model and convergence curve following this strategy. The solution is blocky, as expected from the l_0 -norm on model gradients, while also honoring the data within 1% of the target misfit. This iterative process is very expensive however, as it requires multiple β solves per IRLS iteration. For large non-linear problems, this type of approach would be computationally prohibitive.

Table 4.2: IRLS Solver 2: β -search

Phase I - L_2-norm iterations: $\{\mathbf{m}^{(0)}, \beta^{(0)} \mid \phi_d \simeq \phi_d^*\}$
Adjust β
$\min \phi(m) \rightarrow \mathbf{m}^{(0)}$
Phase II - L_p-norm iterations: $\{\mathbf{m}^{(k)} \mid \delta\phi_m^{(k)} < 1\%\}$
Fix $\{\varepsilon_p, \varepsilon_q\}$
Update $r_i^{(k)}$
Search: $\{\beta^{(k)} \mid \phi_d^{(k)} \simeq \phi_d^*\}$
$\min \phi^{(k)} \rightarrow \mathbf{m}^{(k)}$

As a third strategy, I relax the requirement for the model to predict the data *exactly* between each IRLS steps. The trade-off parameter is adjusted once after each IRLS step such that:

$$\beta^{(k+1)} = \beta^{(k)} * \frac{\phi_d^*}{\phi_d^{(k)}}. \quad (4.31)$$

I have found experimentally however that this type of posterior update is not sufficient to guarantee a smooth convergence. The initial IRLS update can change the model substantially, moving away from the global l_2 -norm solution. In order to preserve the relative importance between misfit and regularization functions while reducing the computational cost, I propose an iterative re-scaling of the model objective function such that:

$$\begin{aligned} \hat{\phi}_m^{(k)} &= \gamma^{(k)} \left(\|\mathbf{W}_s \mathbf{R}_s \mathbf{m}\|_2^2 + \|\mathbf{W}_x \mathbf{R}_x \mathbf{G}_x \mathbf{m}\|_2^2 \right) \\ \gamma^{(k)} &= \frac{\hat{\phi}_m^{(k-1)}}{\|\mathbf{W}_s \mathbf{R}_s \mathbf{m}^{(k-1)}\|_2^2 + \|\mathbf{W}_x \mathbf{R}_x \mathbf{G}_x \mathbf{m}^{(k-1)}\|_2^2}, \end{aligned} \quad (4.32)$$

where $\phi_m^{(k)}$ is the scaled model objective function at some k^{th} iteration, and $\gamma^{(k)}$ is a scalar computed at the beginning of each IRLS iterations. Effectively we are searching for model parameters that are close to the present solution. Since we are changing the *rule* by which the size of the model is measured, we attempt to account for it. I point out that for constant l_p -norm regularizations, the scaling parameter $\gamma^{(k)}$ is equal to one. Table 4.3 summarizes the proposed method.

As shown in Figure 4.6(b), the re-scaling procedure greatly reduces the total number of CG solves. Even though nothing guarantees that non-convex norms will converge to a *global* minimum, the re-scaling scheme proposed here forces local minima to be in the proximity of the l_2 -norm solution. In other words, the scaling parameter $\gamma^{(k)}$ helps preserve the character of all previous iterations, slowly converging to a new minimum, hopefully in the vicinity of the global l_2 -norm solution.

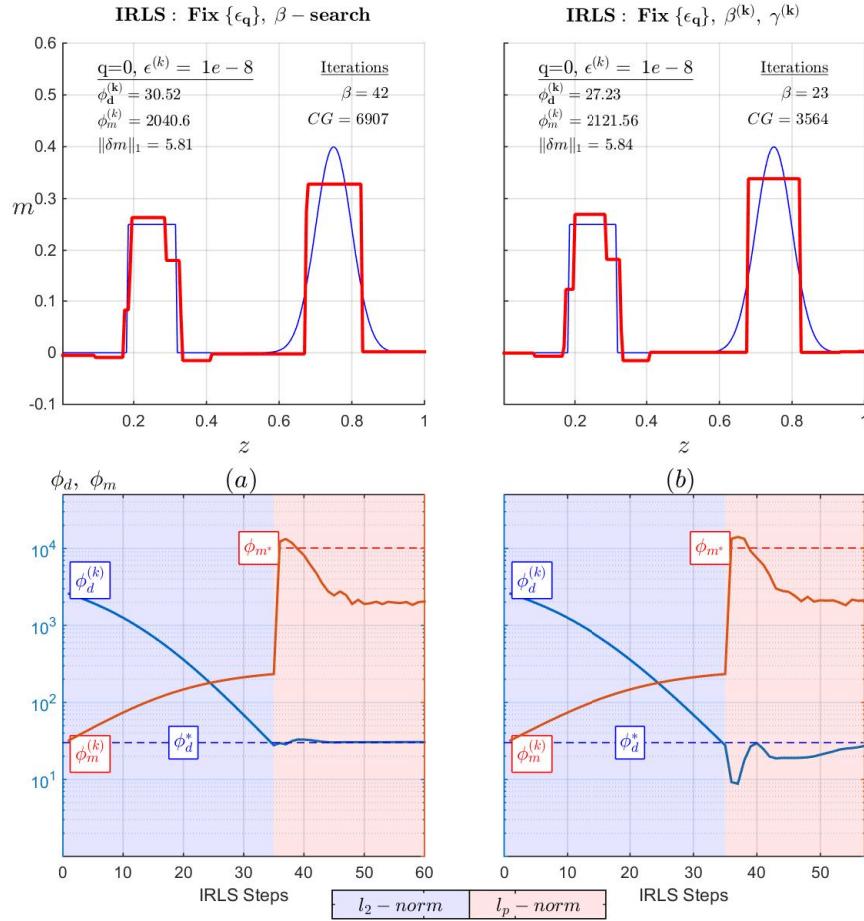


Figure 4.6: (Top) Recovered models from two different algorithms used to implement the IRLS method for $q = 0$ and a fix threshold parameter ($\epsilon = 1e-8$). (Bottom) Measure of model misfit and model norm as a function of iterations. (a) The first algorithm searches for an optimal trade-off parameter $\beta^{(k)}$ between each IRLS step, requiring a solution to multiple sub-inversions. (b) The second algorithm only adjusts $\beta^{(k)}$ once after each IRLS step. A new scaling parameter $\gamma^{(k)}$ is added to smooth the transition between the IRLS updates. The algorithm recovers a similar blocky model but is computationally cheaper, as indicated by the total number of beta iterations and CG solves.

Table 4.3: IRLS Solver 3: Scaled regularization

Phase I: L_2-norm iterations $\{ \mathbf{m}^{(0)}, \boldsymbol{\beta}^{(0)} \mid \phi_d \simeq \phi_d^* \}$ Adjust $\boldsymbol{\beta}$ $\min \phi(m) \rightarrow \mathbf{m}^{(0)}$
Phase II: L_p-norm iterations $\{ \mathbf{m}^{(k)} \mid \delta \phi_m^{(k)} < 1\%, \phi_d \simeq \phi_d^* \}$ Fix $\{\varepsilon_p, \varepsilon_q\}$ Update $r_i^{(k)}$ Scale $\hat{\phi}_m^{(k)}$ $\min \phi^{(k)} \rightarrow \mathbf{m}^{(k)}$ $\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} * \frac{\phi_d^*}{\phi_d^{(k)}}$

4.3.4 Threshold parameter ε

I have so far delayed providing details regarding the choice of threshold parameter ε , which has been the subject of disagreement among researchers (Ajo-Franklin et al., 2007; Barbosa and Silva, 1994; Last and Kubik, 1983; Stocco et al., 2009; Sun and Li, 2014). Little has been said in the literature on how to determine a specific value of ε .

In the classic work of Last and Kubik (1983), and many other research papers after, it has been suggested that the stabilizing parameter should be small ($\varepsilon < 10^{-8}$), or near machine error in order to approximate the l_p -norm well. Other researchers, such as in Ajo-Franklin et al. (2007) have observed severe instabilities with such a small value, and found experimentally that ε should be between 10^{-4} and 10^{-7} . This may be due to the large spread in weights along the diagonal of \mathbf{R} impacting the conditioning of the linear system described in equation (4.6). A large condition number can adversely affect the convergence rate of gradient descent solvers.

As an alternative to the regularization function used by Last and Kubik (1983), Gorodnitsky and Rao (1997) apply the IRLS weights directly to the sensitivity

matrix . The general objective function to be minimized takes the form:

$$\phi = \|\hat{\mathbf{F}}\hat{\mathbf{m}} - \mathbf{d}\|_2^2 + \beta \|\hat{\mathbf{x}}(m)\|_2^2, \quad (4.33)$$

where the weighted forward model operator is written as:

$$\hat{\mathbf{F}} = \mathbf{F} \operatorname{diag}[\mathbf{m}^{(k-1)}]. \quad (4.34)$$

The same technique was later revised by Portniaguine and Zhdanov (2002); Portniaguine (1999) and coined Minimum Support (MS) functional. The method was extended to penalties on the model gradients and named Mininum Gradient Support (MGS) functional. This formulation is interesting as it eliminates the stabilizing parameter ε . From a practical standpoint however, I have found issues when used in concert with other sensitivity based weighting, which will be addressed in Section 4.4.1. Moreover, this type of penalty is less flexible than the general IRLS formulation as I will demonstrate in Section 4.4.

It appears that the choice of ε is problem dependent and becomes a compromise between achieving the desired level of sparsity, while minimizing the numerical cost. Based on the method proposed by Chartrand (2007), I bring in a fourth algorithm using an ε -cooling strategy. The goal is to progressively change the penalty function, starting with a coarse approximation that resemble the l_2 -norm function. Following the smooth inversion, the threshold ε is initialized as a large value ($\varepsilon^{(0)} \approx \max(\mathbf{x}(m)) * 10$), then monotonically reduced between each IRLS step such that:

$$\varepsilon^{(k)} = \frac{\varepsilon^{(0)}}{2^k}. \quad (4.35)$$

The optimal threshold parameter ε is found after convergence of the algorithm as $\delta\phi_m^{(k)} \rightarrow 0$. In a third and final phase, the algorithm fixes $\varepsilon^{(k)}$ and continues re-adjusting the trade-off parameter $\beta^{(k)}$ until reaching the target misfit.

Figure 4.7 presents the recovered model after convergence of the algorithm. The solution is blocky, as expected from an l_0 -norm penalty on the model gradients. This example shows that ε can be much larger than machine error and still accomplish the same objective. By progressively reducing ε , changes in regularization between each IRLS step are reduced. The data residual remains close to

Table 4.4: IRLS Solver 4: ε -Cooling

Phase I: L_2-norm iterations $\{\mathbf{m}^{(0)}, \beta^{(0)} \mid \phi_d \simeq \phi_d^*\}$ Adjust β $\min \phi(m) \rightarrow \mathbf{m}^{(0)}$
Phase II: ϕ_m-iterations $\{\mathbf{m}^{(k)} \mid \delta\phi_m^{(k)} < 1\%\}$ $\varepsilon_p^{(k)} = \frac{\varepsilon_p^{(0)}}{2^k}, \varepsilon_q^{(k)} = \frac{\varepsilon_q^{(0)}}{2^k}$ Update $r_i^{(k)}$ Scale $\hat{\phi}_m^{(k)}$ $\min \phi^{(k)} \rightarrow \mathbf{m}^{(k)}$ $\beta^{(k+1)} = \beta^{(k)} * \frac{\phi_d^*}{\phi_d^{(k)}}$
Phase III: ϕ_d-iterations $\{\mathbf{m}^{(k)} \mid \phi_d \simeq \phi_d^*\}$ Fix $\{\varepsilon_p^{(k)}, \varepsilon_q^{(k)}\}$ Update $r_i^{(k)}$ Scale $\hat{\phi}_m^{(k)}$ $\min \phi^{(k)} \rightarrow \mathbf{m}^{(k)}$ $\beta^{(k+1)} = \beta^{(k)} * \frac{\phi_d^*}{\phi_d^{(k)}}$

the target misfit throughout the iteration process, indicative of a stable algorithm. Numerical experiments have shown that for large inverse problems, the cooling procedure can make the algorithm substantially cheaper and more stable than with a fix and small ε approach.

The method proposed above is advantageous as it does not require a specific choice of threshold parameters ε_p and ε_q . It does depend however on the specified convergence criteria $\delta\phi_m^{(k)}$. In some cases, the algorithm may converge too quickly before reaching the desired level of sparsity. Alternatively, the algorithm may become overly expensive if the convergence criteria is too restrictive or the solution oscillates around the minimum. Secondly, the stabilizing parameter ε can be interpreted as an *effective zero*, penalizing specific ranges of model parameters. It

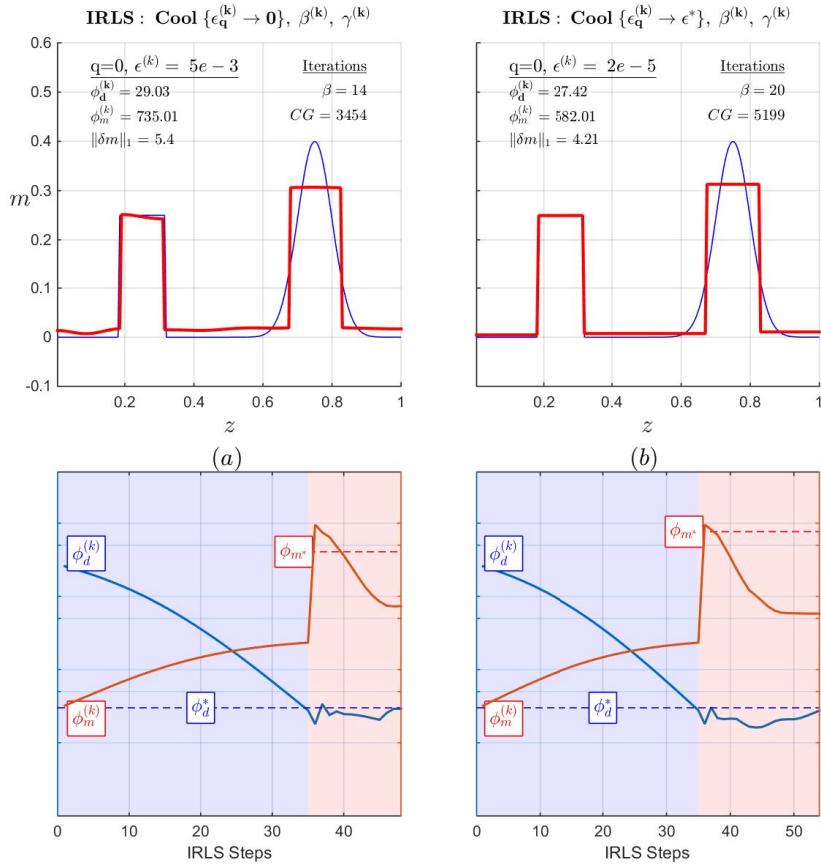


Figure 4.7: (Top) Recovered models from two different algorithms used to implement the IRLS method for $q = 0$ with cooling of the threshold parameter ϵ . (Bottom) Measure of model misfit and model norm as a function of iterations. Both algorithms adjust β and scaling parameter $\gamma^{(k)}$ between each IRLS iteration. (a) In the first case, the threshold parameter ϵ is monotonically reduced until reaching the convergence criteria. The solution is sparser than the previous algorithm, even though ϵ is much larger than machine error. (b) In the second case, ϵ is decreased until reaching the target threshold ϵ^* . The solution is blocky, penalizing small model gradients.

may therefore be necessary to determine a minimum threshold value in order to guarantee convergence and penalize the right model values.

Since the initialization of the IRLS requires a good approximation of the model via a least-squares solution, an estimate of the distribution of model parameters is available for analysis. The value of an optimal ε^* can be chosen directly by the user based on *a priori* information. Alternatively the choice can be based on the distribution of model values. Figure 4.8(a) presents the distribution of model and model gradients obtained from the smooth l_2 -norm inversion. Both curves display a sharp corner around which the model parameters rapidly change. Similarly, Zhdanov and Tolstaya (2004) suggest an L-curve based on the change in model norm such that:

$$\begin{aligned} s_{MS}(\varepsilon) &= \mathbf{m}^T \mathbf{R}_s^T \mathbf{R}_s \mathbf{m} \\ s_{MGS}(\varepsilon) &= \mathbf{m}^T \mathbf{G}_x^T \mathbf{R}_x^T \mathbf{R}_x \mathbf{G}_x \mathbf{m}, \end{aligned} \quad (4.36)$$

where the model norm s_{MS} and model gradient norm s_{MGS} are computed over a range of ε values as shown in 4.8(b). I found experimentally that both approaches were valid but did not always yield a well defined corner. More research may be needed to determine the most robust approach.

As a final experiment, I invert the 1-D example using the point of maximum curvature on the s_{MGS} curve as a minimum threshold for the model gradients ($\varepsilon_q^* = 2e - 5$). Figure 4.7(b) presents the model after convergence. I note that the result is more blocky than the one previously obtained with the convergence criteria. The inversion is, computationally, slightly more expensive due to the larger number of IRLS steps to reach the target threshold parameter ε_q^* .

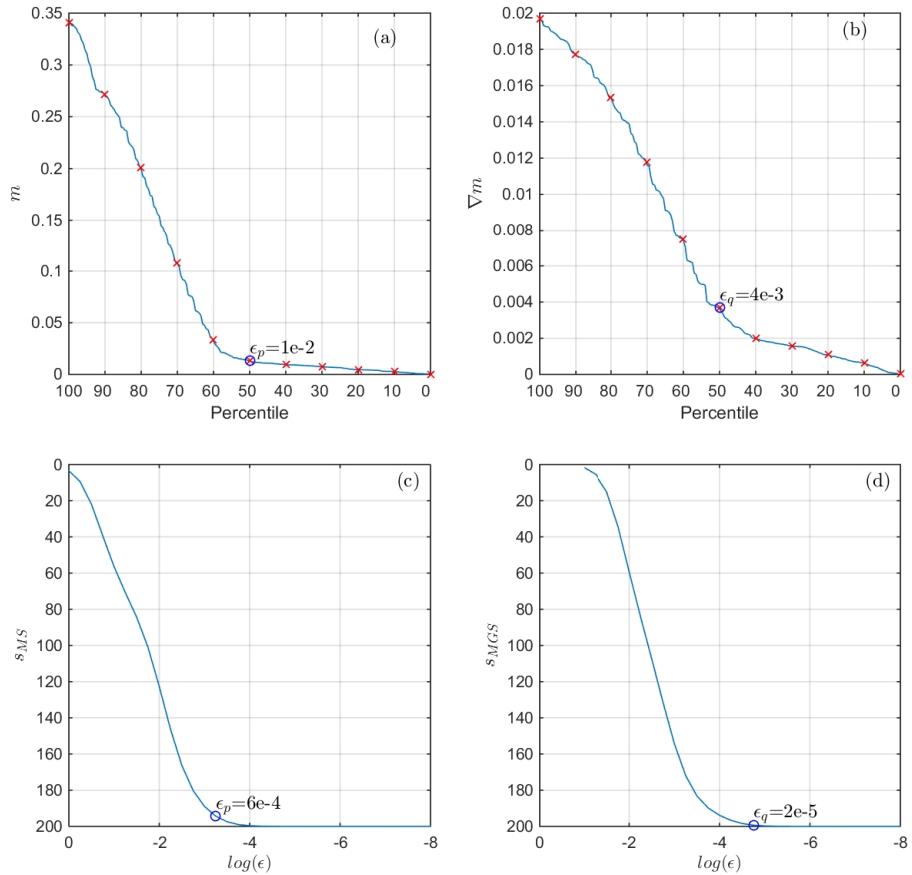


Figure 4.8: (a) Distribution of model parameters and (b) model gradients recovered from the smooth l_2 -norm regularization. Both curves show a sharp corner around which the model functions vary rapidly. Similarly, a measure of (c) model norm s_{MS} and (d) model gradient norm s_{MGS} can be computed over a range of ϵ values, yielding a similar L -curve. The point of maximum curvature can be used to determine the optimal *effective zero* parameter ϵ_p and ϵ_q .

4.4 Scaled-IRLS method (S-IRLS)

The blocky solution found previously is expected from an l_0 -norm penalty on the gradient. But it is clearly not appropriate to recover the true synthetic model, which is both smooth and sparse. Building upon the previous section, I explore different combinations of norms on the model and model gradients in order to *shape* the penalty function. This function should allow any available *a priori* information to be incorporated in the solution. My first attempt uses an l_0 -norm on the model combined with an l_2 -norm on the model gradients, or $\{p = 0, q = 2\}$. For this combination of norms, I would expect the solution to be both sparse and smooth, better approximating the width of the rectangular pulse and Gaussian anomaly. Figure 4.9(a) shows the solution after convergence of the IRLS algorithm. I here identify an important issue with the current IRLS method involving various norm measures within the same objective function. The solution is clearly dominated by the sparsity constraint. In this case, the l_0 -norm penalty on the model suppresses the l_2 -norm penalty on the model gradient, yielding a strictly sparse solution without smoothness constraint.

To understand the issue, it is useful to look at the linear system defined by 4.19. Recall that a solution is found along the gradient of the objective function, which from (4.6) can be written explicitly as:

$$\begin{aligned} \frac{\partial \phi(m)}{\partial m} &= (\mathbf{F}^T \mathbf{d} - \mathbf{F}^T \mathbf{F} \mathbf{m}) - \beta (\mathbf{g}_s(m) + \mathbf{g}_x(m)) \\ \mathbf{g}_s(m) &= \mathbf{R}_s^T \mathbf{W}_s^T \mathbf{W}_s \mathbf{R}_s \mathbf{m} \\ \mathbf{g}_x(m) &= \mathbf{G}_x^T \mathbf{R}_x^T \mathbf{W}_x^T \mathbf{W}_x \mathbf{R}_x \mathbf{G}_x \mathbf{m}. \end{aligned} \quad (4.37)$$

I divided (4.37) into three parts to highlight the different components of the total gradient direction. The first term $(\mathbf{F}^T \mathbf{d} - \mathbf{F}^T \mathbf{F} \mathbf{m})$ is related to the misfit function and depends solely on the ability of the model to reproduce the data \mathbf{d} . The second term $\beta (\mathbf{g}_s(m) + \mathbf{g}_x(m))$ is related to the regularization function, which depends on the IRLS weights \mathbf{R}_s and \mathbf{R}_x computed in 4.16.

To illustrate the importance of scaling between the function gradient terms I

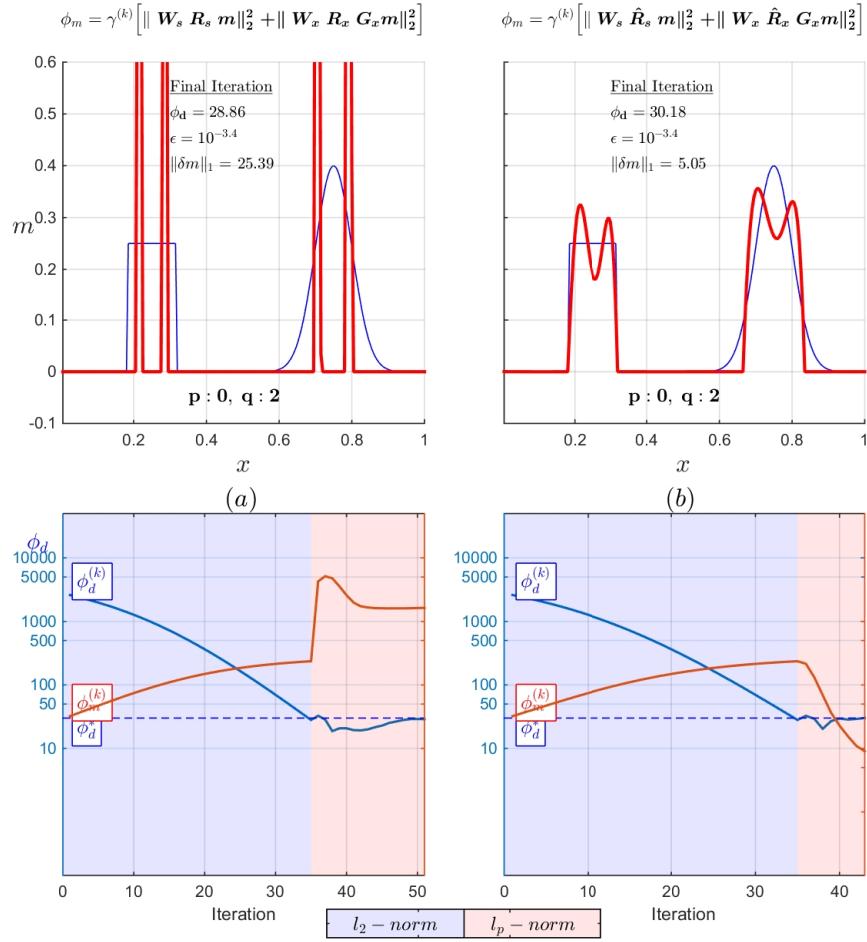


Figure 4.9: (a) (Top) Recovered model using a mixed-norm regularization for $p = 0, q = 2, \epsilon = 1e - 3$. (Bottom) Measure of model misfit and model norm as a function of iterations. The inversion converges to a sparse solution without apparent penalty on the gradient, indicative of an imbalance in the regularization. (b) Recovered model and convergence curves after rescaling of the regularization, yielding a model that is both sparse and smooth as expected from the applied mixed-norm .

form a slightly different 2-variable problem where:

$$\mathbf{F} = \begin{bmatrix} 0 & 1 \end{bmatrix}, \mathbf{d} = 1. \quad (4.38)$$

Figure 4.10(a) displays contours along the surface formed by the l_2 -norm measure of data residual $\|\mathbf{F}\mathbf{m} - \mathbf{d}\|_2^2$, where this time the null-space of \mathbf{F} lies parallel to the x -axis. I simplify the problem by having cell size and weights all equal to one so the objective function can be written as:

$$\begin{aligned} \phi(\mathbf{m}) &= \phi_d + \beta [\phi_s + \phi_x] \\ &= \|\mathbf{F}\mathbf{m} - \mathbf{d}\|_2^2 + \beta \left[\|\mathbf{m}\|_2^2 + \|\mathbf{G}_x \mathbf{m}\|_2^2 \right]. \end{aligned} \quad (4.39)$$

By setting a small trade-off parameter ($\beta = 1e-3$), I can focus on the solution space along the null-space of \mathbf{F} marked by the red dash line, as shown in Figure 4.10(e). The global minimum of the objective function lies at $\mathbf{m} = [0.5; 1]$, where the partial gradients of $\frac{\partial \phi_s}{\partial m_1}$ and $\frac{\partial \phi_x}{\partial m_1}$ have equal and opposite signs. Here the two regularization functions use an l_2 -norm measure, hence the optimal solution is at mid-distance from their respective minimums. This simple example shows that the actual value of the individual norms do not matter, but rather the relative magnitude of the gradients along the minimum of ϕ_d .

The behavior of the l_0 -norm for small p and ε values can easily explain the result obtain in Figure 4.9(a). Figure 4.11(a) compares the function gradients for different norm penalties as a function of p values. In this case, the function gradients \mathbf{g}_s measured with the l_0 -norm are always much larger than the function gradients \mathbf{g}_x measured with the l_2 -norm, except at $m_i = \{0, \sqrt{1 - \varepsilon^2}\}$. The solution can therefore only be sparse with few non-zero model values required to satisfy ϕ_d^* . The penalty on model gradients had no influence on the solution.

Just as I added the parameter $\gamma^{(k)}$ to balance the relative importance between the misfit function and the regularization, I also want to control the relative scaling between the various regularization functions in terms of their gradients. In other words, I want to find a scaling parameter that makes the gradient of any l_p -norm function to intersect, elsewhere than at $m_i = \{0, \sqrt{1 - \varepsilon^2}\}$.

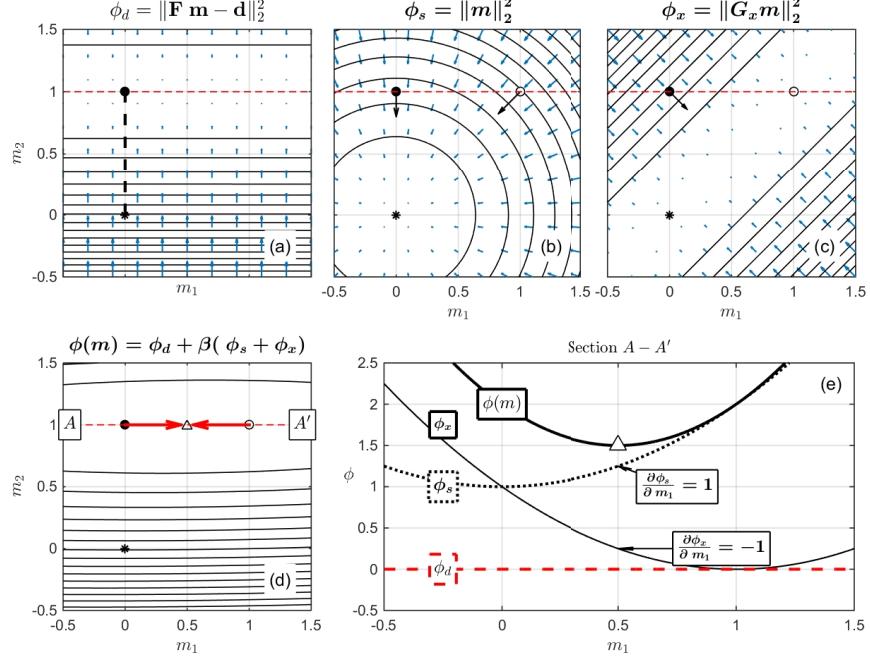


Figure 4.10: Contour maps for (a) the misfit function ϕ_d , (b) the model norm ϕ_s and (c) the norm of model gradients ϕ_x . (d) The total objective function $\phi(m)$ has a global minimum located at $\mathbf{m} = (0.5, 1.0)$ for a given small trade-off parameter ($\beta = 1e-3$). The direction of update is shown for two starting models $\mathbf{m}^{(0)}$ (black and white dot). (e) Section through the objective function along the minimum of ϕ_d . The global minimum occurs where the partial gradients of $\frac{\partial \phi_s}{\partial m_1}$ and $\frac{\partial \phi_x}{\partial m_1}$ have equal and opposite signs.

I here introduce a scaling parameter η :

$$\eta = \varepsilon^{(1-p/2)}, \quad (4.40)$$

where once again p denotes the l_p -norm penalty and ε is a small value used to approximate the norm. Adding the scaling parameter η to 4.13, the partial gradients

of the objective function become:

$$\frac{\partial \phi_m}{\partial m} = \frac{\epsilon^{(1-p/2)} x_i}{(x_i^2 + \epsilon^2)^{1-p/2}}. \quad (4.41)$$

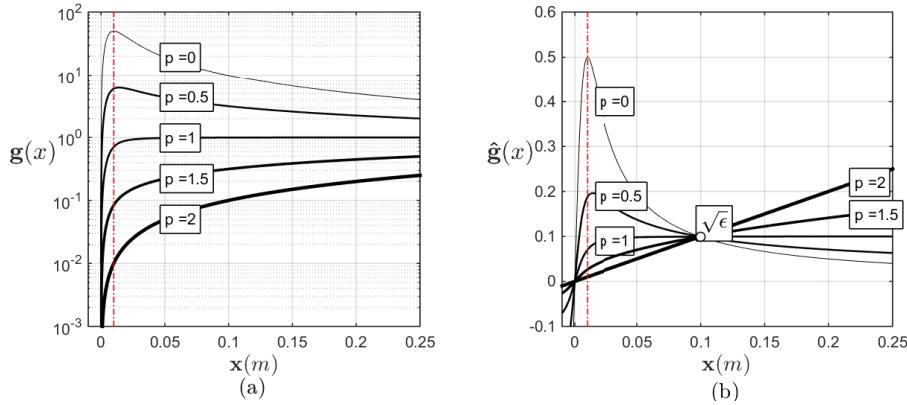


Figure 4.11: (a) Partial gradients of approximated l_p -norm penalties for a fix stabilizing parameter $\epsilon = 1e-2$. Gradients for $p < 1$ are consistently larger on the interval $[0 < x_i < \sqrt{1 - \epsilon^2}]$, making it hard to combine multiple norm penalties within the same objective function. (b) Function gradients after applying a scale of $\epsilon^{(1-p/2)}$, forcing each l_p -norm to intersect at $m = \sqrt{\epsilon}$.

Looking at the gradient for $x_i = \sqrt{\epsilon}$, 4.41 simplifies to:

$$\frac{\partial \phi_m}{\partial m} = \frac{\epsilon^{1/2}}{(1 + \epsilon)^{1-p/2}}.$$

Then assuming that $\epsilon \ll 1$:

$$\frac{\partial \phi_m}{\partial m} \approx \epsilon^{1/2}. \quad (4.42)$$

Hence after applying the scaling η , the gradient of the l_p -norm is independent of the p -value exactly at $m = \sqrt{\epsilon}$, and has also a gradient $\frac{\partial \phi}{\partial m} = \sqrt{\epsilon}$, as shown in Figure 4.3(c). Therefore, $\sqrt{\epsilon}$ is a critical point around which all norm penalties can influence the solution. In this manner, two penalty functions with different p -values can coexist within a regularization function and achieve different objectives.

Adding this final scaling to 4.32 yields a *Scaled Iterative Re-weighted Least Squares* (S-IRLS) regularization function such that:

$$\hat{\phi}_m^{(k)} = \gamma^{(k)} \left(\|\mathbf{W}_s \hat{\mathbf{R}}_s \mathbf{m}\|_2^2 + \|\mathbf{W}_x \hat{\mathbf{R}}_x \mathbf{G}_x \mathbf{m}\|_2^2 \right). \quad (4.43)$$

I define a Scaled-IRLS weighting matrix $\hat{\mathbf{R}}$ such that:

$$\begin{aligned} \hat{R}_{sii} &= \sqrt{\eta_p} \left[(m_i^{(k-1)})^2 + \varepsilon_p^2 \right]^{(p/2-1)/2} \\ \hat{R}_{xii} &= \sqrt{\eta_q} \left[\left(\frac{\partial m_i^{(k-1)}}{\partial x} \right)^2 + \varepsilon_q^2 \right]^{(q/2-1)/2} \\ \eta_p &= \varepsilon_p^{(1-p/2)} \\ \eta_q &= \varepsilon_q^{(1-q/2)}, \end{aligned} \quad (4.44)$$

where ε_p and ε_q are the stabilizing parameters for the l_p -norm of the model and model gradients respectively. Note that for $p = 2$, the scaling parameters η is equal to one, hence we recover the the classic l_2 -norm regularization.

Figure 4.9(b) presents the result after a re-scaling of the gradient descent. Both the l_0 -norm penalty on the model and the l_2 -norm penalty on the model gradients are represented, yielding a solution that is both sparse and smooth. I now have a flexible and robust regularization function. This allows us to explore a wide range of solutions, or l_p -space of solutions, combining various norms on the model and model gradients.

4.4.1 Cell-based weights (w_r, w_m)

In Li and Oldenburg (1996), a depth weighting function is added to counter the natural decay of potential fields. The rapid decay in sensitivity is an important problem in geophysics as most data sets are acquired from the surface. The same idea can be used to incorporate any *a priori* information regarding the spatial distribution of model parameters. In this section, I investigate the effect of having such cell-based weighting applied directly to the sensitivity matrix, compared to having it applied to the regularization function as previously formulated in 3.8 .

Following the weighted sensitivity formulation of Li and Oldenburg (1996),

the objective function with S-IRLS regularization takes the form:

$$\phi(\hat{\mathbf{m}}) = \|\mathbf{W}_d (\hat{\mathbf{F}} \hat{\mathbf{m}} - \mathbf{d})\|_2^2 + \gamma^{(k)} \left[\|\mathbf{W}_s \hat{\mathbf{R}}_s \hat{\mathbf{m}}\|_2^2 + \|\mathbf{W}_x \hat{\mathbf{R}}_x \mathbf{G}_x \hat{\mathbf{m}}\|_2^2 \right], \quad (4.45)$$

where:

$$\begin{aligned} \hat{\mathbf{F}} &= \mathbf{F} \mathbf{W}_r^{-1} \\ \hat{\mathbf{m}} &= \mathbf{W}_r \mathbf{m} \\ \mathbf{W}_r &= \text{diag} \left(e^{jz/(2\pi)} \right), \end{aligned}$$

where the matrix \mathbf{W}_r hold the inverse exponential describing the decay of the kernel functions presented in 4.22. Similar sensitivity based weighting is used in the *Minimum Support* functional of Portniaguine and Zhdanov (2002).

Using both formulations, I invert the same 1-D example presented in Figure 4.7 and 4.9. In the first experiment, I apply a sparsity constraint on the model gradients for $q = 0, \alpha_s = 0$. Figure 4.12(a) compares the true and recovered models after convergence. I observe the solution obtained with the $\phi(\hat{\mathbf{m}})$ formulation is skewed in the direction of the sensitivity weighting. Similarly, Figure 4.12(b) shows the recovered model for the combination of $p = 0, q = 2$. While the recovered anomaly over the rectangular pulse is both smooth and sparse, I note that the solution tends to favor a strictly sparse model further at depth over the Gaussian function. On the other end, the solution obtained with the $\phi(\mathbf{m})$ formulation remains smooth and sparse over both anomalies equally.

Issues encountered with the $\phi(\hat{\mathbf{m}})$ formulation are due to the IRLS weights computed in 4.27. Written explicitly in terms of weighted model parameters, the linearized IRLS norm from 4.14 can be written as:

$$\phi_{\hat{\mathbf{m}}}^{(k)} = \sum_{i=1}^{nc} \frac{(w_i x_i)^2}{\left[(w_i x_i^{(k-1)})^2 + \varepsilon^2 \right]^{1-p/2}}. \quad (4.46)$$

Factoring out the cell-based weight w_i , equation 4.46 becomes:

$$\phi_{\hat{m}}^{(k)} = \sum_{i=1}^{nc} \frac{w_i^p x_i^2}{\left[(x_i^{(k-1)})^2 + \epsilon^2 / w_i^2 \right]^{1-p/2}}. \quad (4.47)$$

The important point to note is that the weights are now function of the p -value, while the threshold parameter ϵ depends on the cell-based weights. In most cases, the sensitivity weights increase with distance from the observation location, which in turn reduce the threshold parameter ϵ and increase the influence of sparse norms. This explains the increase in sparsity constraint with distance as observed in Figure 4.12. In order to preserve the character of cell-based weights, it is important to isolate their effect from the l_p -norm penalty. The same idea extends to any type of cell-based weights (i.e. volumetric, *a priori*). In the following chapter, various norm measures will be used simultaneously on the model and model gradient. The predictability and stability of the algorithm depends greatly on the scaling between the various components of the objective function.

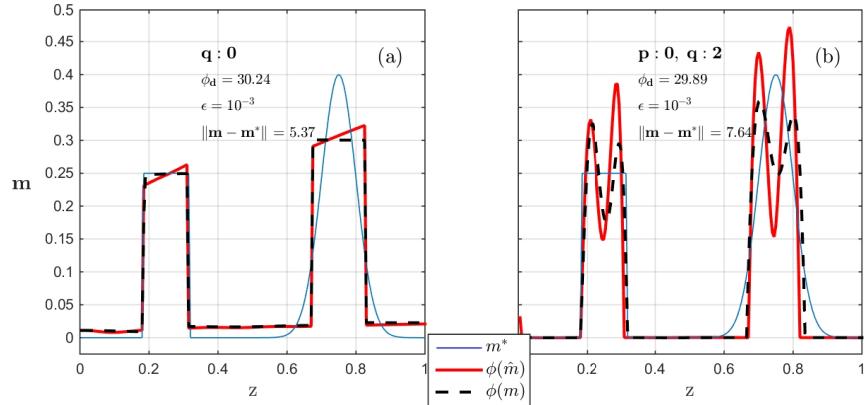


Figure 4.12: Recovered models for two different depth weighting formulations: (red) weighted sensitivity $\phi(\hat{m})$, (black) weighted regularization $\phi(m)$. (a) True and recovered models using the $\phi(\hat{m})$ and $\phi(m)$ formulations for penalty applied on the model gradients for $q = 0$ and (b) for $p = 0$, $q = 2$. The weighted sensitivity formulation $\phi(\hat{m})$ increases the influence the regularization function with distance along the x -axis, skewing the model towards the right.

4.5 Mixed l_p -norm regularization

I showcase the robustness and flexibility of the S-IRLS algorithm by inverting a total of 441 models, corresponding to all the combinations of norms on the interval $0 \leq p \leq 2$ and $0 \leq q \leq 2$, on a 0.1 increment. Figure 4.13(a) presents a contour map of the final model errors $\|\delta\mathbf{m}\|_1$ recovered after each inversion. The gradual variation in model error is a good indicator of the predictability of the algorithm, as small changes in the regularization yield equivalently small changes in the solution. For each inversion, the trade-off parameter β was adjusted to fit within $\pm 2\%$ of the target data misfit for appropriate comparison of the results, as shown in 4.13(b). In this case, the optimal model is found with a combined norm of $[p = 1.5, q = 0.4]$. In this particular case, the worst recovery was obtained with $p = 0, q = 2$, yielding a sparse solution with few large oscillations. Even though this type of analysis would be computationally prohibitive for large scale inverse problems, it illustrates the flexibility of the S-IRLS method in designing an objective function reflecting specific characteristics.

Nine of the inverted models are shown in Figure 4.14 for comparison. As expected from the property of the IRLS functions, small norms on the gradient promote blocky solutions. From right to left, the inversion result tends to favor right-angled anomalies. From bottom to top, smaller norms on the model parameter enforce sparse solutions with few non-zero values.

4.5.1 Localized S-IRLS

While I have so far implemented global regularization functions applied on the entire model, the S-IRLS method can easily be extended to problems made up of several sub-regions with varying norm penalties. Also based on the IRLS method, Sun and Li (2014) introduce an automated process to extract structural information from the data and impose either an l_1 or l_2 -norm penalties to specific regions of the model. While designing this objective function can be challenging and may still require direct input from the user, their approach has great value as it increases the flexibility over traditional global penalty functions. The S-IRLS formulation has the potential to further generalize the work of Sun and Li (2014) in allowing the use of any norm penalty in the range of $0 \leq p \leq 2$.

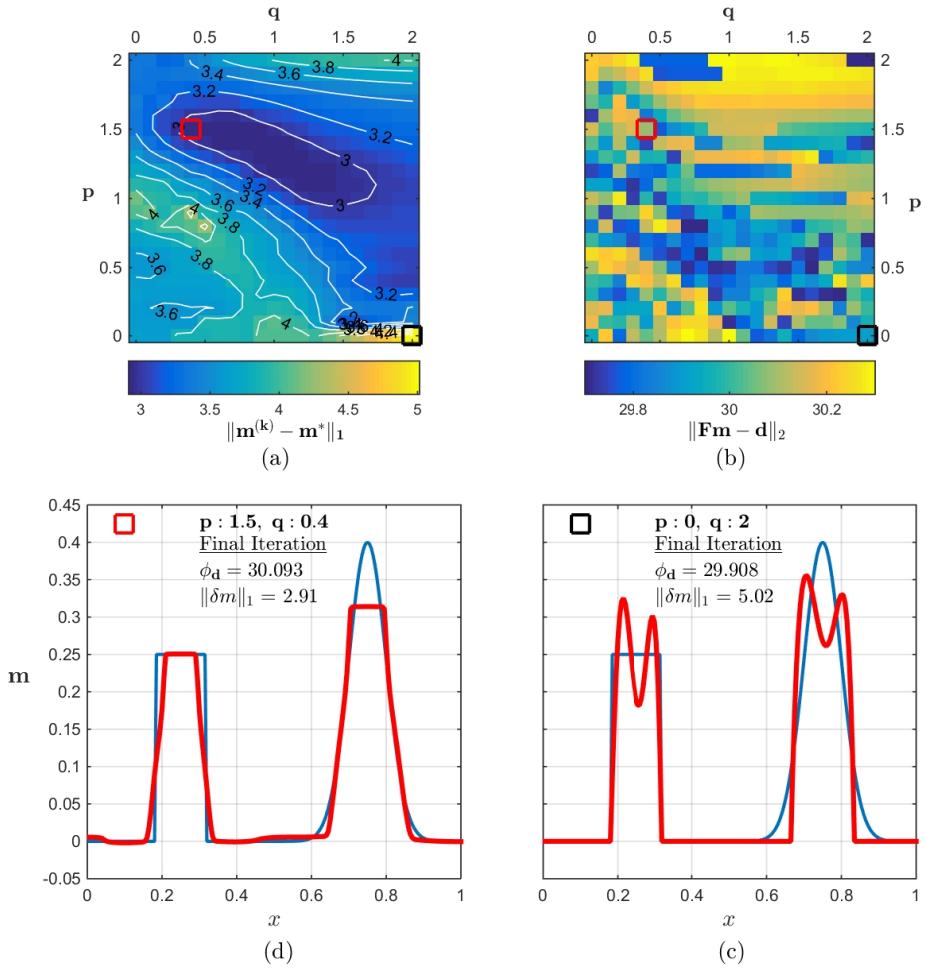


Figure 4.13: (a) Model error $\|m - m^*\|_1$ and (b) misfit function for the 441 inverted models using a range of regularization with mixed-norm penalty on the model for $0 \leq p \leq 2$ and on model gradients for $0 \leq q \leq 2$. (c) The largest model error ($\|\delta\mathbf{m}\|_1$) was obtained with the mixed-norm for $p = 0$, $q = 2$, compared to (d) the optimal solution found with $p = 1.5$ and $q = 0.4$.

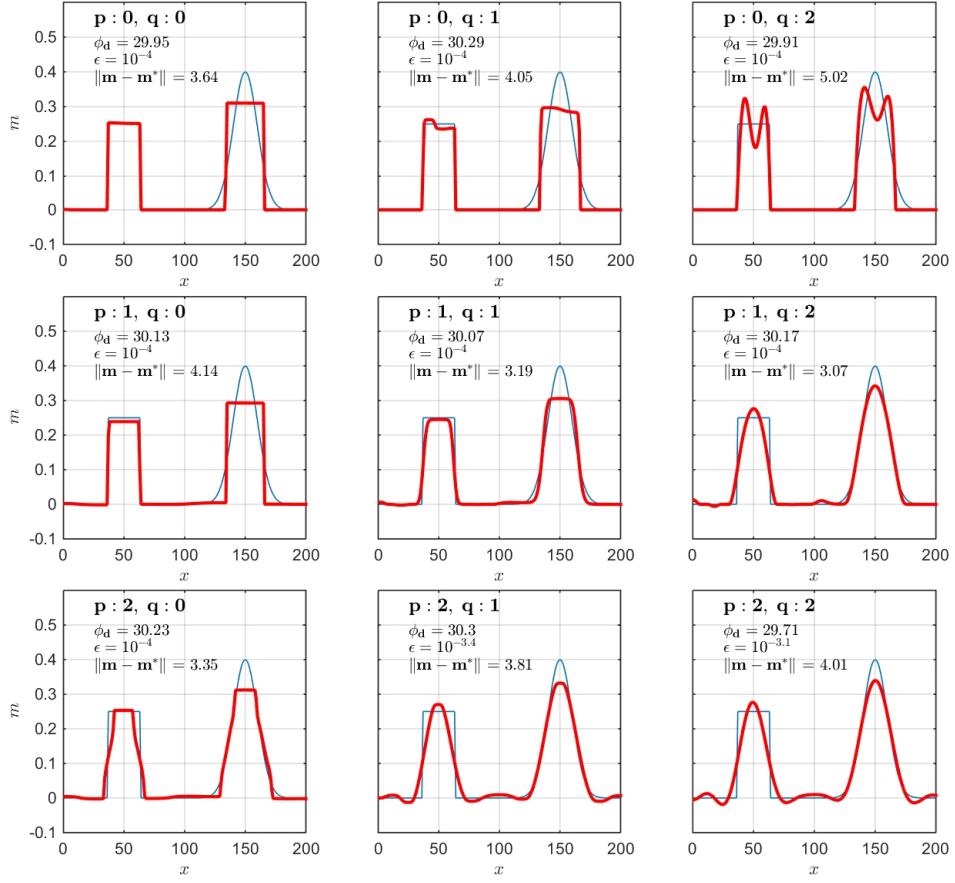


Figure 4.14: (a) Nine of the 441 inverted models for a range of mixed-norm penalties on the model and its gradient for $0 \leq p \leq 2$ and $0 \leq q \leq 2$.

Suppose an N -dimensional model space $\mathbf{m} \in \Omega$, divided in J sub-regions such that:

$$\begin{aligned} \mathbb{S} &= \{S_1, S_2, \dots, S_J \mid S_j \subset \Omega, j = 1, \dots, J\} \\ S_i \cap S_j &= \{\emptyset \mid i, j = 1, \dots, J, j \neq i\}. \end{aligned}$$

I can define a *model* of l_p -norms by the union of each sub-domain:

$$\mathbf{p} = \bigcup_{j=1}^J S_j p_j, \quad (4.48)$$

where the scalar p_j defines the l_p -norm to be applied to the j^{th} region of the model domain. Each sub-domain S_j can have a distinct value for p and q , penalizing the model values and model gradient differently than neighboring domains. The transition in l_p -norm penalty can be smoothed after applying a linear transition across neighboring domains as defined by:

$$\begin{aligned} \tilde{\mathbf{p}} &= \mathbf{A} \mathbf{p} \\ \mathbf{A} &= \begin{bmatrix} 1/2 & 1/2 & 0 & \dots & 0 \\ \vdots & \ddots & & & 0 \\ 0 & \dots & 0 & 1/2 & 1/2 \end{bmatrix}, \end{aligned} \quad (4.49)$$

where the transition between l_p -norm is extended over several cells by applying the averaging operator \mathbf{A} . Allowing for multiple sub-domains requires only a slight change to the S-IRLS method presented in Section 4.4. The general objective function for a 3-D discretized problem becomes:

$$\phi(m) = \phi_d + \beta \gamma^{(k)} \left[\|\mathbf{W}_s \hat{\mathbf{R}}_s (\mathbf{m} - \mathbf{m}^{\text{ref}})\|_2^2 + \sum_{i=x,y,z} \|\mathbf{W}_i \hat{\mathbf{R}}_i \mathbf{G}_i \mathbf{m}\|_2^2 \right], \quad (4.50)$$

such that the S-IRLS weights for the model and model gradients become:

$$\begin{aligned} \hat{R}_{sii} &= \sqrt{\tilde{\eta}_{pi}} \left[(m_i^{(k-1)})^2 + \epsilon_p^2 \right]^{(\tilde{p}_i/2-1)/2} \\ \hat{R}_{xii} &= \sqrt{\tilde{\eta}_{qxi}} \left[\left(\frac{\partial m_i^{(k-1)}}{\partial x} \right)^2 + \epsilon_q^2 \right]^{(\tilde{q}_{xi}/2-1)/2} \\ \hat{R}_{yii} &= \sqrt{\tilde{\eta}_{qyi}} \left[\left(\frac{\partial m_i^{(k-1)}}{\partial y} \right)^2 + \epsilon_q^2 \right]^{(\tilde{q}_{yi}/2-1)/2} \\ \hat{R}_{zii} &= \sqrt{\tilde{\eta}_{qzi}} \left[\left(\frac{\partial m_i^{(k-1)}}{\partial z} \right)^2 + \epsilon_q^2 \right]^{(\tilde{q}_{zi}/2-1)/2}. \end{aligned} \quad (4.51)$$

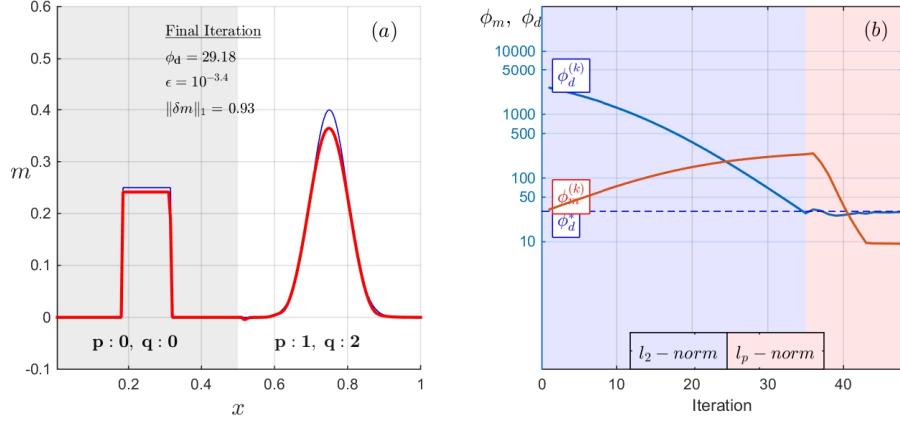


Figure 4.15: (Left) Improved solution for the 1-D problem after applying a localized mixed-norm penalty, where the regularization is divided into two regions with independent l_p -norm regularization: (left) $p = q = 0$, (right) $p = 1, q = 2$. (Right) Convergence curves for the mixed-norm S-IRLS inversion.

Figure 4.15 presents an optimal solution for the 1D problem, where I divide the model space into two regions. The left half uses a $p = 0, q = 0$, whereas the right half imposes a $p = 1$ and $q = 2$. The mixed-norm penalty function enforces the right characteristic to each portion of the model. The inversion recovers a sparse and blocky model over the rectangular pulse while at the same time managing to model the smooth Gaussian anomaly. Consequently, the model error $\|\delta m\|$ is much smaller than any of the previous attempts. The algorithm converges smoothly and rapidly to a stable solution.

4.5.2 2-D example

I test the S-IRLS algorithm on a 2-D synthetic model to further demonstrate the flexibility and robustness of the mixed-norm regularization. Analogous to the 1-D problem, the 2-D model consists of a square block and a smooth Gaussian anomaly, only this time the observations stations are located on both sides of the model domain. The kernel functions are exponentially decaying cosine functions of the form:

$$e^{-\omega r} \cdot \sin(2\pi\omega r), \quad (4.52)$$

as shown in Figure 4.16.

Since we are now dealing with a 2-D problem, the regularization function involves a measure of model gradients in the x and z -direction. Most methods proposed in the literature, and presented in 3.9, involve computing gradients with finite difference operators in orthogonal directions such that:

$$\begin{aligned} \frac{\partial m_i}{\partial x} &= \frac{m_{(i,j)} - m_{(i-1,j)}}{dx_{(i)}} \\ \frac{\partial m_i}{\partial z} &= \frac{m_{(i,j)} - m_{(i,j-1)}}{dz_{(j)}}, \end{aligned} \quad (4.53)$$

where the indexes (i, j) represent the cells ordering in the x and y -direction respectively. A second option is to penalize the absolute model gradients such that:

$$\nabla m_i = \left[\left(\frac{\partial m_i^{(k-1)}}{\partial x} \right)^2 + \left(\frac{\partial m_i^{(k-1)}}{\partial z} \right)^2 \right]^{1/2}. \quad (4.54)$$

The IRLS weights then become:

$$\begin{aligned} \hat{R}_{xii} &= \sqrt{\tilde{\eta}_{qxi}} \left[(\nabla m_i)^2 + \epsilon_q^2 \right]^{(\tilde{q}_{xi}/2-1)/2} \\ \hat{R}_{zii} &= \sqrt{\tilde{\eta}_{qzi}} \left[(\nabla m_i)^2 + \epsilon_q^2 \right]^{(\tilde{q}_{zi}/2-1)/2}. \end{aligned} \quad (4.55)$$

The same idea can easily be extended to three dimensions. I illustrate the difference between both measures of model gradients on this 2-D problem. Figure 4.17(a) presents the recovered model using an l_1 -norm penalty on the model gradi-

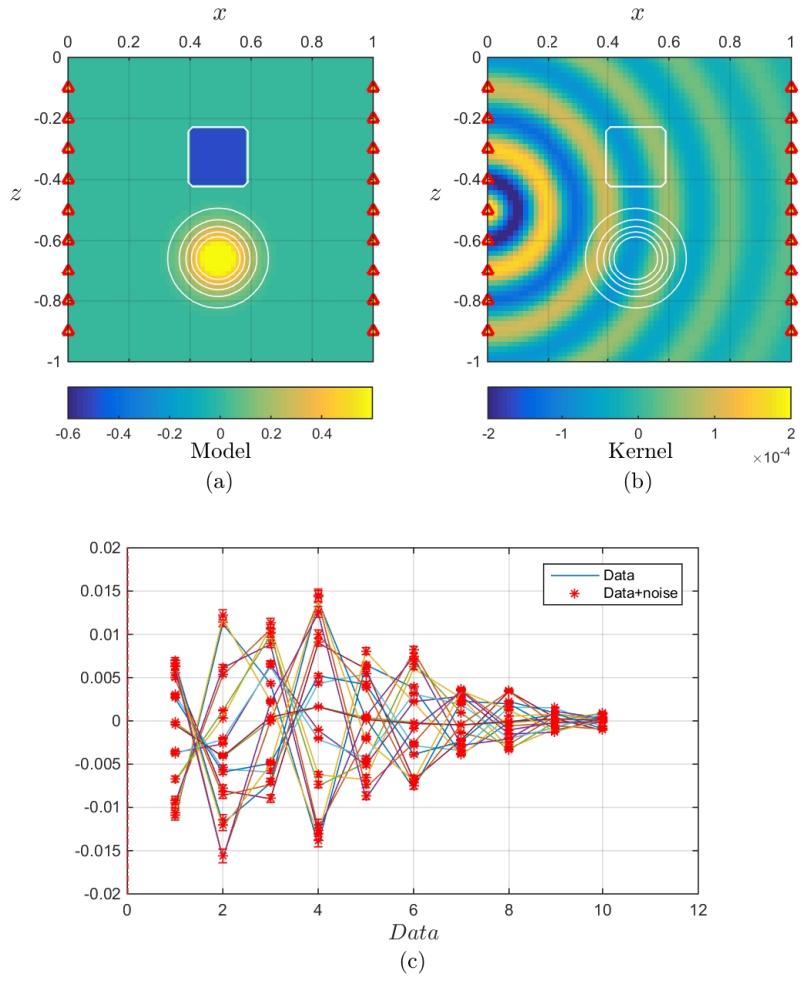


Figure 4.16: (a) Synthetic 2-D model made up of a square block and a smooth Gaussian function. (b) Example of a kernel function for $e^{-\omega r} \cdot \cos(2\pi\omega r)$ and (c) data generated from $\mathbf{d} = \mathbf{F} \mathbf{m}$. Five percent random Gaussian noise is added.

ents from the finite difference formulation. This yields blocky right-angled anomalies. Alternatively, penalizing the absolute model gradients from (4.54) allows for smooth corners as shown in 4.17(b). Both methods are valuable as they promote different solutions, which in some cases may better reflect the geometry of the true model.

To accommodate both the blocky and the smooth anomaly, I divide the model space into three zones. Figure 4.18 presents the different l_p -norm zones defined over the two anomalies. I extend the procedure of sub-regions presented in Section 4.5.1 to this 2-D problem. I impose sparse model values with the l_1 -norm over the background region in order to get a simple model. The norm on model gradient is variable and reflects the expected characteristics of the true model. I purposefully chose regions larger than the extent of the known anomalies in order to simulate a choice that could be made for a blind inversion. The goal is also to see if the transition from an l_2 -norm to a l_0 -norm can be done smoothly without creating artifacts. Figure 4.19(a) and (b) presents the recovered model after Phase 1 and Phase III of the S-IRLS algorithm previously introduced in Table 4.4. Both models fit the data within 2% of the target misfit ϕ_d^* . The final mixed-norm model nicely recovers the blocky and smooth Gaussian anomaly. The transitions between the different l_p -regions appear to be seamless.

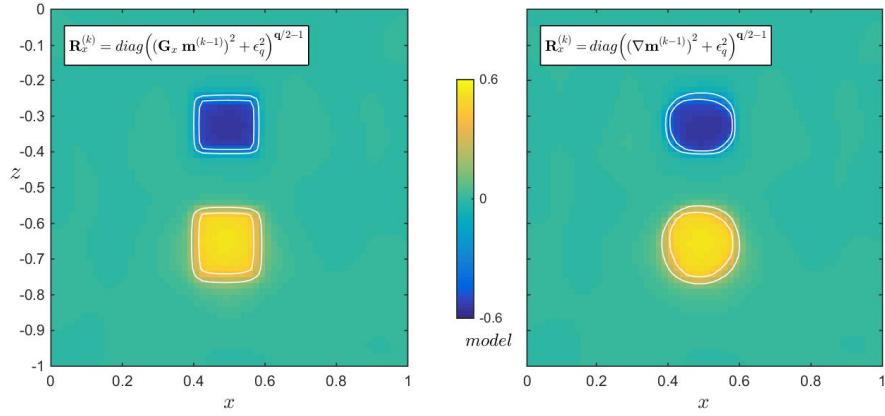


Figure 4.17: (a) Recovered model for $p = q_x = q_z = 1$ penalizing finite difference gradients in orthogonal directions, yielding right-angled anomalies. (b) Recovered model for the same norms but penalizing the absolute gradient of the model ($|\nabla \mathbf{m}|$) recovering round edges.

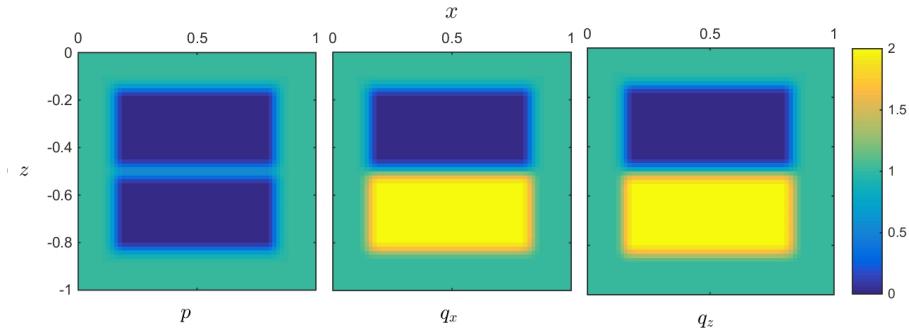


Figure 4.18: Distribution of l_p -norm on the model and model gradients over the 2-D model domain. The original boundary of each region was smoothed in order to get a slow transition and reduce visible artifacts. Regions were chosen to cover a larger area than the anomalies to simulate a blind-inversion.

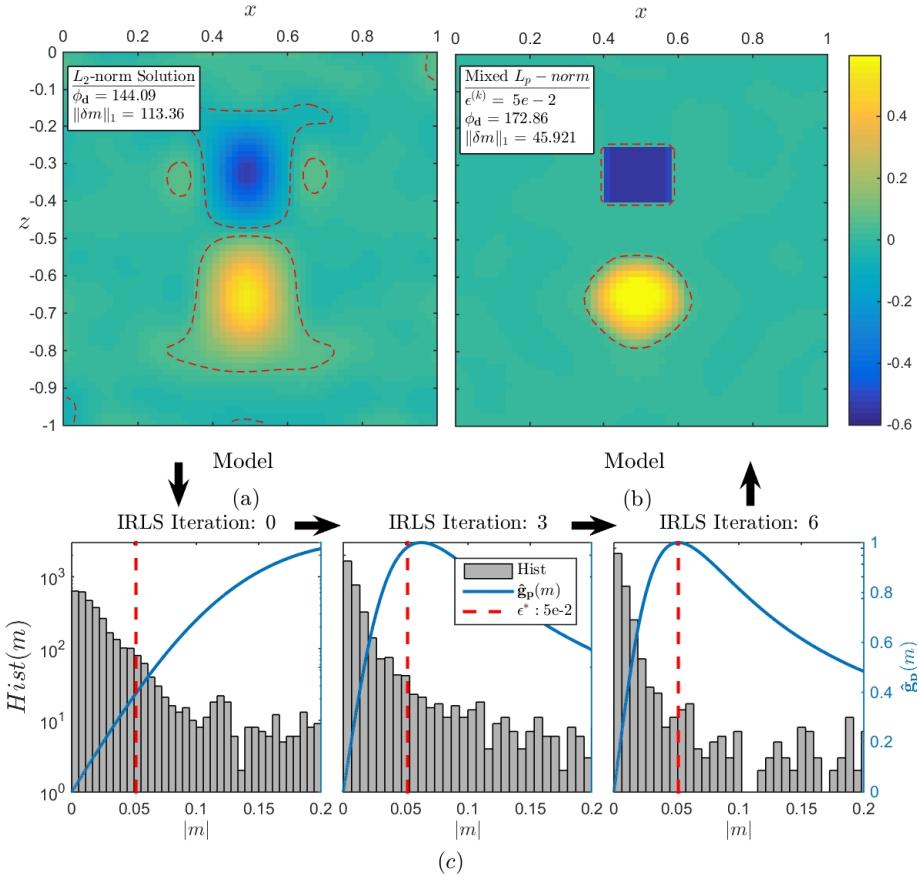


Figure 4.19: (a) Smooth L_2 -norm solution used to initiate the IRLS iterations. (b) Recovered model using the mixed-norm regularization after seven S-IRLS iterations. The contour line (red) marks the value of ϵ_p , judged to be the *effective zero* value of the model ($m_i \leq 5e-2$). Both models (a) and (b) fit the data within 2% of the target misfit ϕ_d^* . (c) Dual plots showing the distribution of model parameters and the gradient of the l_0 -norm penalty function $\hat{g}_p(m)$ as a function of S-IRLS iterations. High penalties are applied to progressively smaller model values. The final model nicely recovers both the blocky and smooth Gaussian anomaly.

4.5.3 3-D example

As a final test, I invert the same synthetic 3-D susceptibility example as presented in Section 3.2. The model consists of a folded anomaly with magnetic susceptibility of 0.075 SI, arching around a discrete block with susceptibility of 0.05 SI. The arc-shaped anomaly is dipping 20° towards the south.

I impose sparsity constraints on the model gradients such that ($p = 0$, $q_x = 2$, $q_y = 2$, $q_z = 2$). Figure 4.20 shows the recovered model after five IRLS iterations. The solution is remarkably closer to the true solution, both in shape and model values compared to the smooth l_2 -norm solution. The inversion also has an easier time reproducing the high frequency content from the data as shown in Figure 4.21. The inversion successfully recovers two distinct objects, no longer connected at depth. Notice that both arms are well defined and extend along the entire length of the arc.

As a second experiment, I impose a constraint on the model gradients in order to recover a blocky model. Model gradients are measured with the standard finite difference operators, expected to yield right-angled anomalies. Figures 4.22 and 4.23 show the recovered model and predicted data for ($p = 0$, $q = 1$), yielding a sparse and blocky solution. Once again, compared to the smooth l_2 -norm solution, the correlated data residuals are substantially reduced. The recovered susceptibility values within the block and arc shape are more uniform than with the smoothness constraint, ideal to get a bulk estimate in physical property.

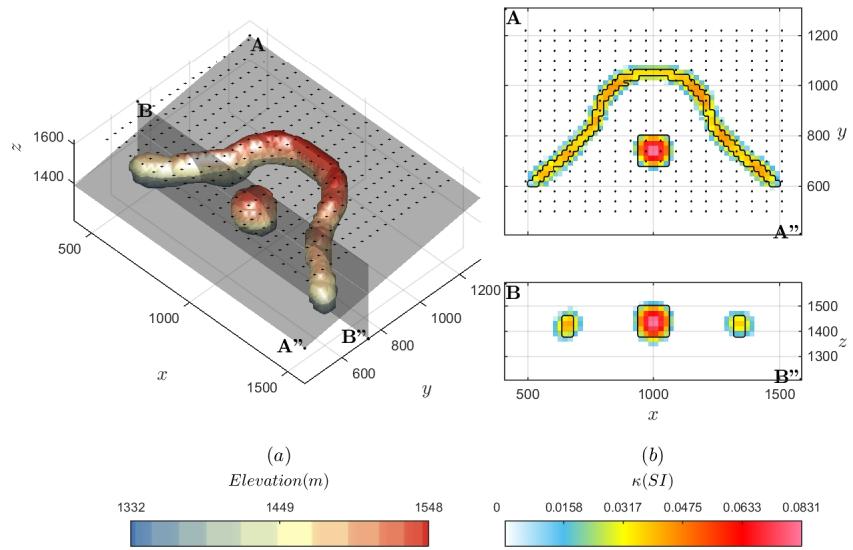


Figure 4.20: (a) Iso-surface (0.002 SI) and (b) sections through the recovered susceptibility model after five IRLS iterations for ($p = 0$, $q = 2$). The final model is substantially closer to the true solution.

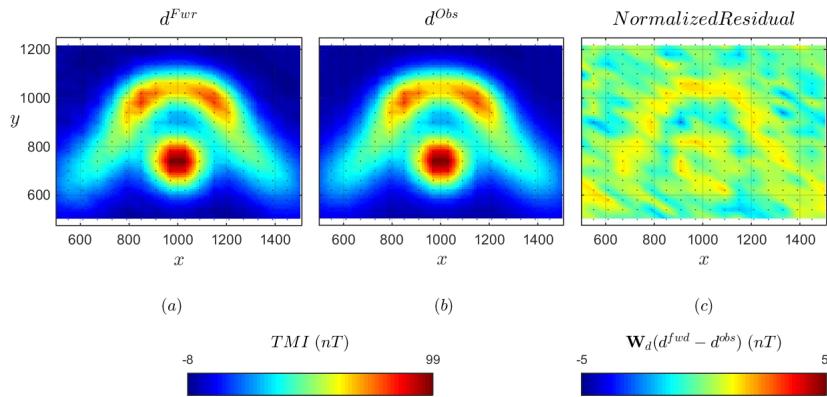


Figure 4.21: Comparison between (a) observed and (b) predicted data from the recovered susceptibility model using compact norms for ($p = 0$, $q = 2$). (c) Normalized data residuals are within two standard deviations.

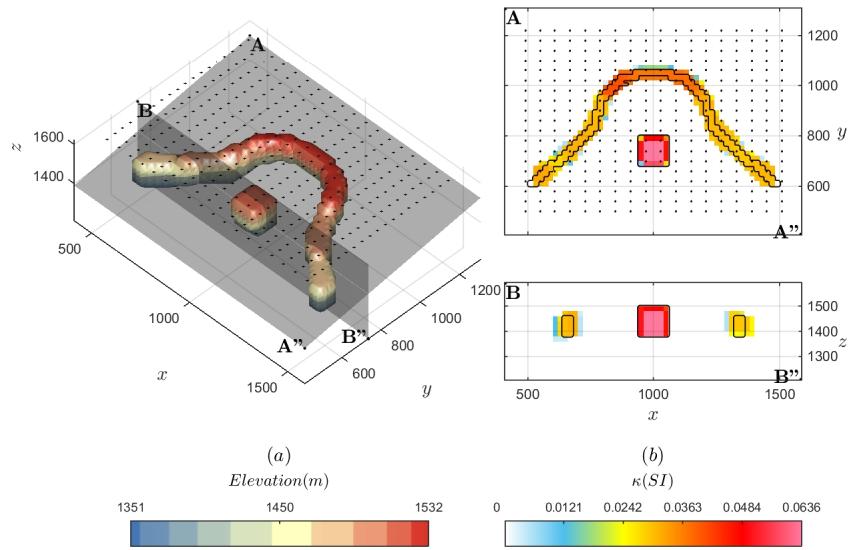


Figure 4.22: (a) Iso-surface (0.002 SI) and (b) sections through the recovered susceptibility model after nine IRLS iterations for ($p = 0, q = 1$) . Sparsity constraints on the model and model gradients yield a simple and blocky model.

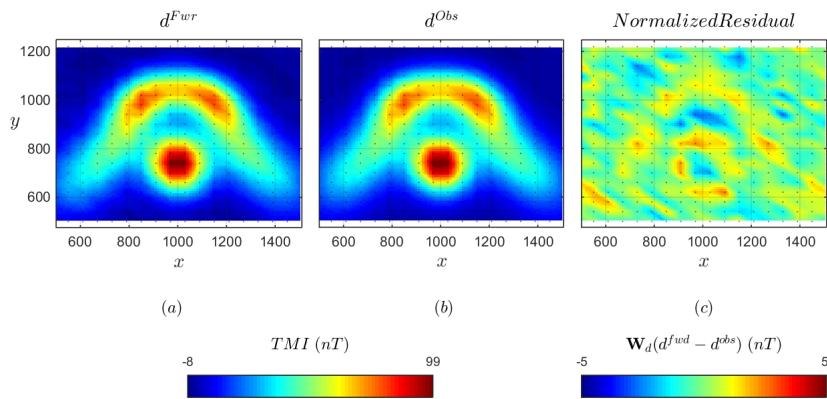


Figure 4.23: Comparison between (a) observed and (b) predicted data from the recovered susceptibility model using compact norms for ($p = 0, q = 1$). (c) Normalized data residuals are within two standard deviations.

4.6 Case study - Tli Kwi Cho kimberlite complex

As a final test, I apply the mixed-norm S-IRLS algorithm on a magnetic data from the Tli Kwi Cho (TKC) diamondiferous kimberlite complex. The property is located in the Lac de Gras region, approximately 350 km northeast of Yellowknife, Northwest Territories, Canada. The TKC deposit was originally discovered by an airborne DIGHEM survey in 1992, including frequency-domain EM and magnetic data. Two kimberlite pipes, dubbed DO27 and DO18, were later confirmed by drilling in 1993. Different volcaniclastic rock units suggests that the deposit was formed over several events (Doyle et al., 1999). The pipes are intruding into older granitic rocks of the Archean Slave Province, known to have little to no magnetite content. A magnetic susceptibility contrast between the kimberlite and the country rocks give rise to noticeable anomalies on the observed TMA data. Strong linear features associated with the Mackenzie dyke swarms can be observed on the eastern part of the survey. These dykes extend over the entire Lac de Gras region and generally strike 305°N (Wilkinson et al., 2001). They are known to be near vertical and are typically between 20-50 m wide. This geological setting is therefore ideal for implementing a mixed-norm algorithm. My goal is to apply *soft* constraints on the model values and model gradients in order to favor specific geometry.

Prior to the inversion, I rotate the regional dataset by 30° counterclockwise to a local coordinate system in order to align the principal axis of the dykes parallel to the grid. The declination of the inducing field are also rotated by 30° to preserve the geometry of the problem. The region of interest is discretized into 25 m cube cells. In preparation for Phase II of the S-IRLS algorithm, a smooth solution is found with the l_2 -norm regularization as shown in Figure 4.24(a). The inversion recovers two discrete bodies corresponding to the known location of DO18 and DO27. Two linear anomalies striking north-south correspond with the Mackenzie dyke swarms. A third anomaly intersects at right-angle running 90°N. As expected from the l_2 -norm regularization, the solution is smooth and edges are not clearly defined. Note also that the highest recovered susceptibilities are strongly correlated with the observation locations. Dykes appear to break up between each survey line, which is clearly an inversion artifact due to changes in sensitivities.

I want to modify the objective function in order to recover sharp and continuous

edges. In general terms, I want to enforce large gradients perpendicular to the strike of the dykes, while imposing a sparsity constraint over the kimberlite pipes. I divide the region of interest into four sub-regions as shown in Figure 4.25:

- S_1 : Sub-region covering the Mackenzie dyke swarms on the eastern edge of the survey. I impose an l_0 -norm on the model gradients in the \hat{x} -direction ($q_x = 0$) to get sharp edges along the strike of the dykes, but smooth in the other directions ($q_y, q_z, p = 2$).
- S_2 : Sub-region covering the dyke running perpendicular to the Mackenzie dyke swarms. Similar to S_1 , I impose an l_0 -norm on the model gradients in the \hat{y} -direction ($q_y = 0$).
- S_3 : Sub-region covering the kimberlite pipes DO27 and DO18. I impose an l_1 -norm on the model gradients ($q_x, q_y, q_z = 1$) and an l_0 -norm on model value ($p = 0$) in order to recover blocky and sparse anomalies.
- S_4 : Background region covering the rest of the survey area. I impose an l_2 -norm on the model gradients ($q_x, q_y, q_z = 2$) and an l_1 -norm on model value ($p = 1$) in order to recover a smooth and sparse model.

Figure 4.24(b) presents the recovered model obtained after five iterations of the S-IRLS algorithm. In comparison with the smooth inversion, the edges of all three dykes are better defined and extend vertically at depth. The inversion may suggest the presence of a fourth dyke parallel to the Mackenzie dyke swarm, on the eastern edge of the survey. Both DO18 and DO27 are recovered as compact bodies with higher susceptibility and well defined edges compared to the smooth l_2 -norm inversion. Because of the small l_0 -norm applied to model values, most of the near surface anomalies are removed, yielding a more compact anomaly at depth. This is also true away from the kimberlite pipe over the background region, where a simpler model is obtained. Predicted data for the smooth l_2 -norm model and the final S-IRLS solution are presented in Figure 4.26, both within 5% of the target misfit.

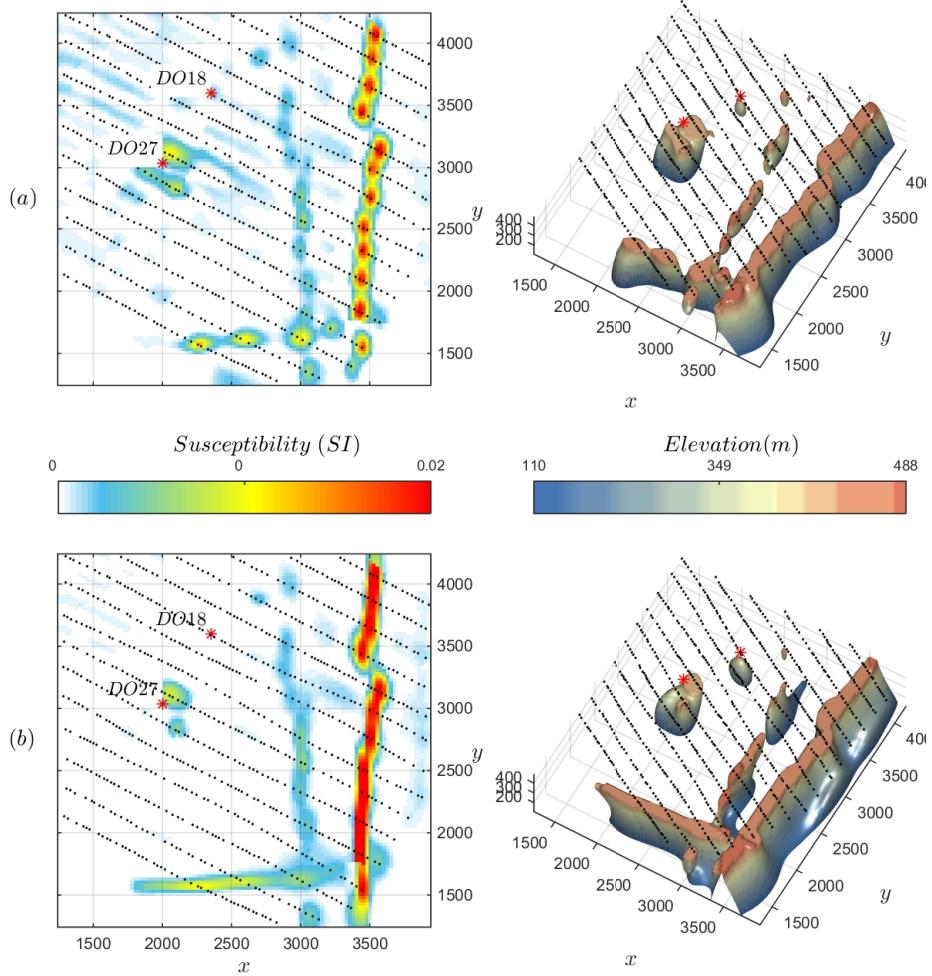


Figure 4.24: (a) (Left) Horizontal section through the recovered susceptibility model at 25 m depth below topography from the smooth l_2 -norm regularization. (Right) Iso-surface of susceptibility values around 0.002 SI. (b) Recovered model using the mixed-norm S-IRLS algorithm. Magnetic dykes are better recovered, imaged as continuous plates and extending vertically at depth. Susceptibility values for DO-27 and DO-18 have increased, showing as compact vertical pipes.

4.7 Summary

In this chapter, I reviewed and experimented with several algorithms for the approximation of l_p -norm measures on the model and model gradients to promote sparse and blocky solutions. I proposed a Scaled-IRLS algorithm in order to stabilize current inversion strategies, while reducing the computational cost. My algorithm offers a strategy to determine an optimal value for the stabilizing parameter ϵ , previously overlooked by other researchers. This robust method is further generalized, allowing for mixed-norm penalties on the model and model gradients. Scalings are applied to the individual terms of the regularization function in order to preserve their relative importance during the inverse process. A solution to the non-linear inverse problem is found iteratively using Gauss-Newton steps.

Varying the combination of l_p -norm of the model and its gradients allows me to shape the objective function to exhibit specific characteristics. For different geological settings, there is likely a specific combination of norms that can better recover the true geometry of causative bodies. Preliminary results show great promise at reducing the non-uniqueness of current potential fields inversion codes. Because my formulation is general, the algorithm can be applied to a wide range of inverse problems.

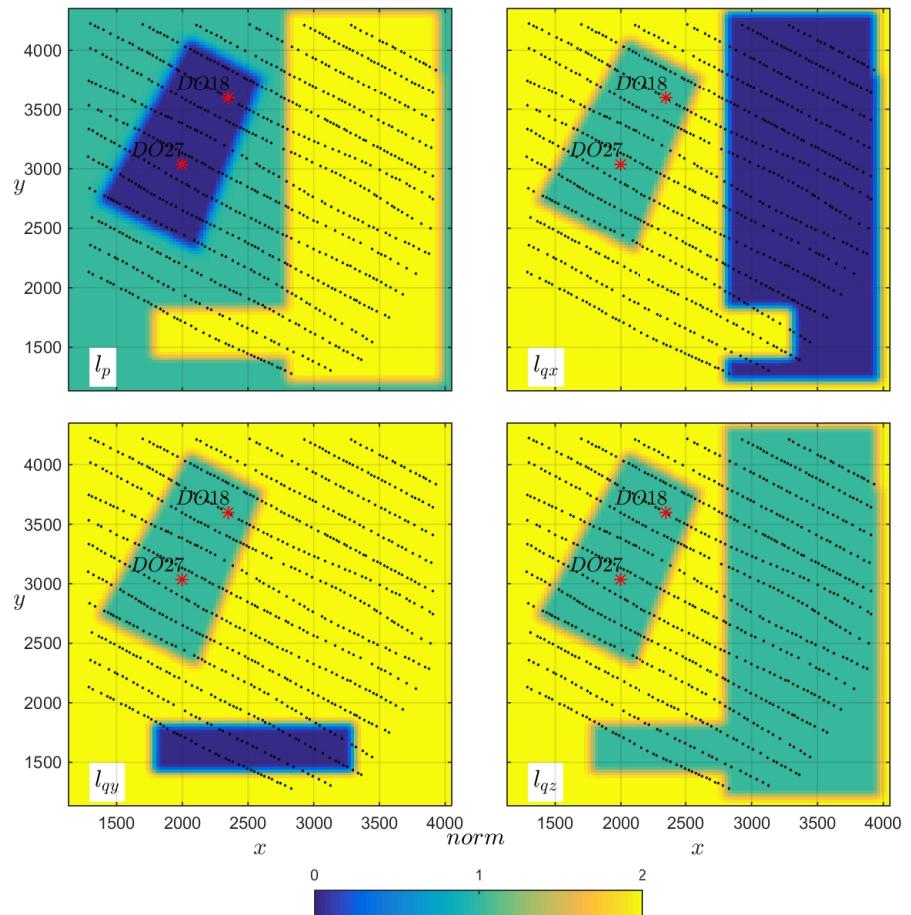


Figure 4.25: Horizontal section through the mixed-norm models applied to four sub-regions with smooth transition across zones.

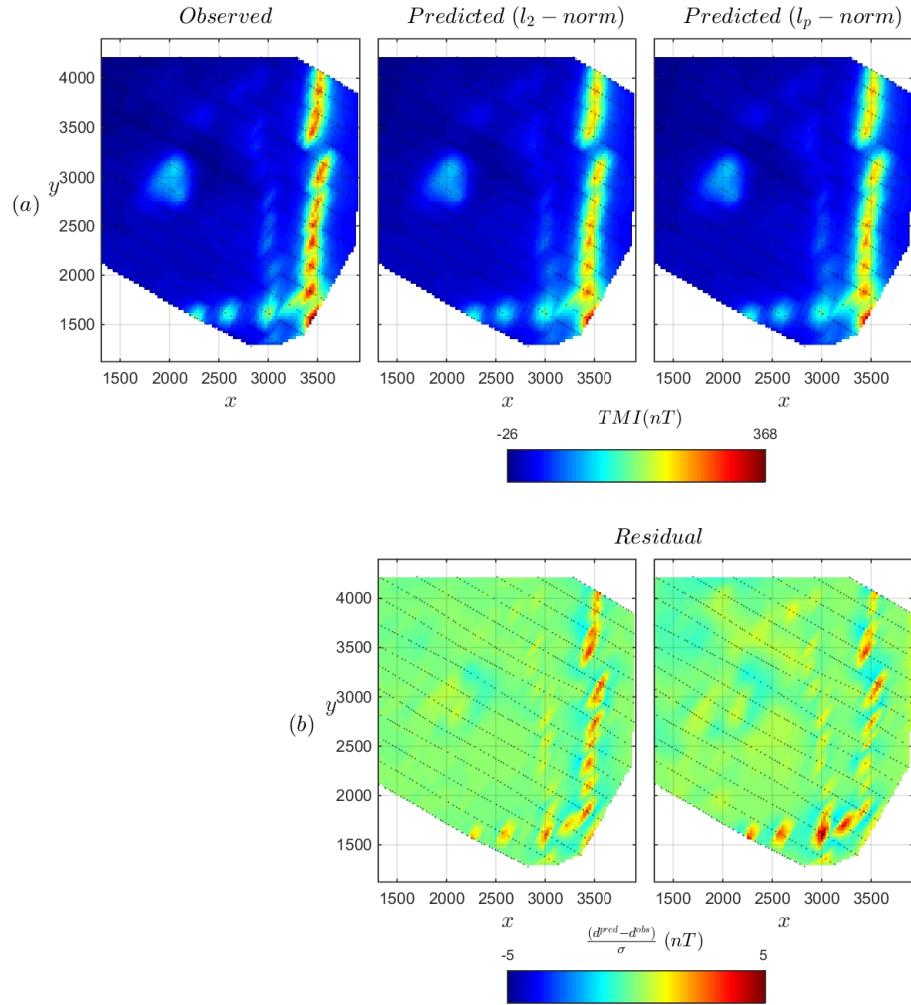


Figure 4.26: (a) Observed and predicted data over the TKC kimberlite complex. (b) Residuals between observed and predicted data normalized by the estimated uncertainties (10 nT). Both the smooth and mixed-norm inversions reproduce the data within four standard deviations.

Chapter 5

Cooperative Magnetic Inversion (CMI)

In Chapter 3, I reviewed three magnetic inversion algorithms introduced in the literature. The magnetic susceptibility inversion is robust and computationally cheap, but it runs into limitations whenever the direction of magnetization is not parallel to the inducing field. The more general Magnetic Vector Inversion (MVI) can handle any orientation of magnetization by solving directly for the magnetization vector. This comes at the cost of increasing the number of unknowns in our large under-determined system of equations. The l_2 -norm regularization yields smooth solutions that are often too complex for direct geological interpretation. Lastly, the inversion of magnetic amplitude data can provide a good estimate of the location and strength of magnetized bodies. The method is less sensitive to the orientation of magnetization than an inversion carried out with an assumed direction of magnetization. The main complication for amplitude inversion is in computing amplitude data from the observed TMI, either via Fourier transform or by the equivalent-source. The inverse problem is also non-linear with respect to the model, which increases the computational cost and is less stable.

While all of the above inversion methods have strength and weaknesses, they each bring complementary information. The susceptibility inversion can be used to rapidly compute an equivalent-source layer and to generate amplitude data. The amplitude data can then be fed into the amplitude inversion to locate magnetic

anomalies. In turn, imposing a geometric constraint on the MVI can greatly reduce the non-uniqueness of the problem. All three algorithms share variations of the same system of equations. This is beneficial as calculations for the sensitivity matrix can be stored and repurposed at each step.

In this chapter, I introduce a Cooperative Magnetic Inversion (CMI) method, which combines the equivalent-source, amplitude inversion and MVI method into a single inversion algorithm. Using the S-IRLS method from Chapter 4, I impose a sparsity constraint on the model and model gradients to further reduce the solution space. From an imaging stand point, compact and well defined magnetic anomalies can simplify the geological interpretation. From a practical aspect, having an algorithm that can manage all three codes reduces the number of manual steps that were previously required from the user.

5.1 Methodology

In Chapter 3, I inverted a synthetic model with complicated magnetization orientations. The model recovered from the MVI inversion was smooth with poor recovery of the arc anomaly (Fig. 3.10). Using the CMI algorithm, I re-invert the synthetic example using an l_2 -norm regularization. I will here breakdown each step of the CMI algorithm, but I want to emphasize that the entire process is automated and can all be done as a single workflow. I divide the algorithm in three parts as shown in Figure 5.1 in schematic form.

Stage I of the algorithm is mainly a data preparation step, ahead of the amplitude inversion and MVI. As in other inversion methods, the algorithm requires TMA data and a mesh. The algorithm computes the large system of equations needed for sensitivity calculations as defined by 2.10. Following the sensitivity calculations, the algorithm proceeds with an equivalent-source layer inversion as described in Section 3.5. Figure 5.2 presents the inverted effective susceptibility layer and predicted data at the station locations. Amplitude data are saved and passed on to Stage II.

In Stage II, the algorithm proceeds with the amplitude inversion from Section 3.4. Figure 5.3 shows the recovered smooth effective susceptibility model. The recovered effective susceptibility model is then converted to sensitivity weighting

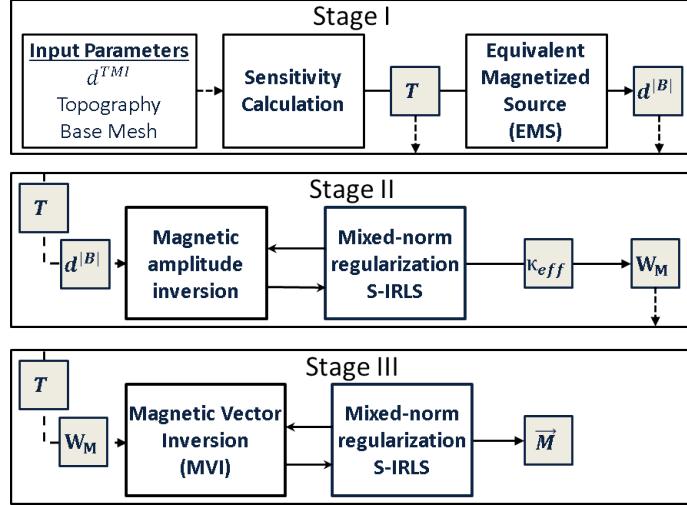


Figure 5.1: Schematic representation of the Cooperative Magnetic Inversion (CMI) algorithm. Input and output parameters are indicated by a dash arrow.

matrix such that:

$$W_{mii} = \left[\frac{\kappa_{ei} * 0.9}{\max(\kappa_e)} + 0.01 \right]^{-1}, \quad (5.1)$$

where $\mathbf{W}_m \in \mathbb{R}^{nc \times nc}$ is a diagonal matrix of the normalized effective susceptibilities κ_e . A small number (1e-2) is added to assure that all values are between [1 100]. This type of re-scaling was determined empirically to be robust. The sensitivity weighting matrix is saved and passed on to the MVI inversion.

In Stage III, the regularization function is pre-multiplied by the effective susceptibility weighting calculated in 5.1. The model objective function becomes:

$$\phi(m) = \phi_d + \beta \gamma^{(k)} \left[\|\mathbf{W}_m \mathbf{W}_s \hat{\mathbf{R}}_s (\mathbf{m} - \mathbf{m}^{\text{ref}})\|_2^2 + \sum_{i=x,y,z} \|\mathbf{W}_m \mathbf{W}_i \hat{\mathbf{R}}_i \mathbf{G}_i \mathbf{m}\|_2^2 \right]. \quad (5.2)$$

The model weighting matrix \mathbf{W}_m imposes a high penalty on cells that received low effective susceptibility from the amplitude inversion. Figure 5.5 presents the inverted model after reaching the target data misfit. The added information from the amplitude inversion greatly improves the solution over the MVI alone. Although still smoothly varying, the inversion manages to separate the arc and the block anomaly, while also recovering the orientation of magnetization.

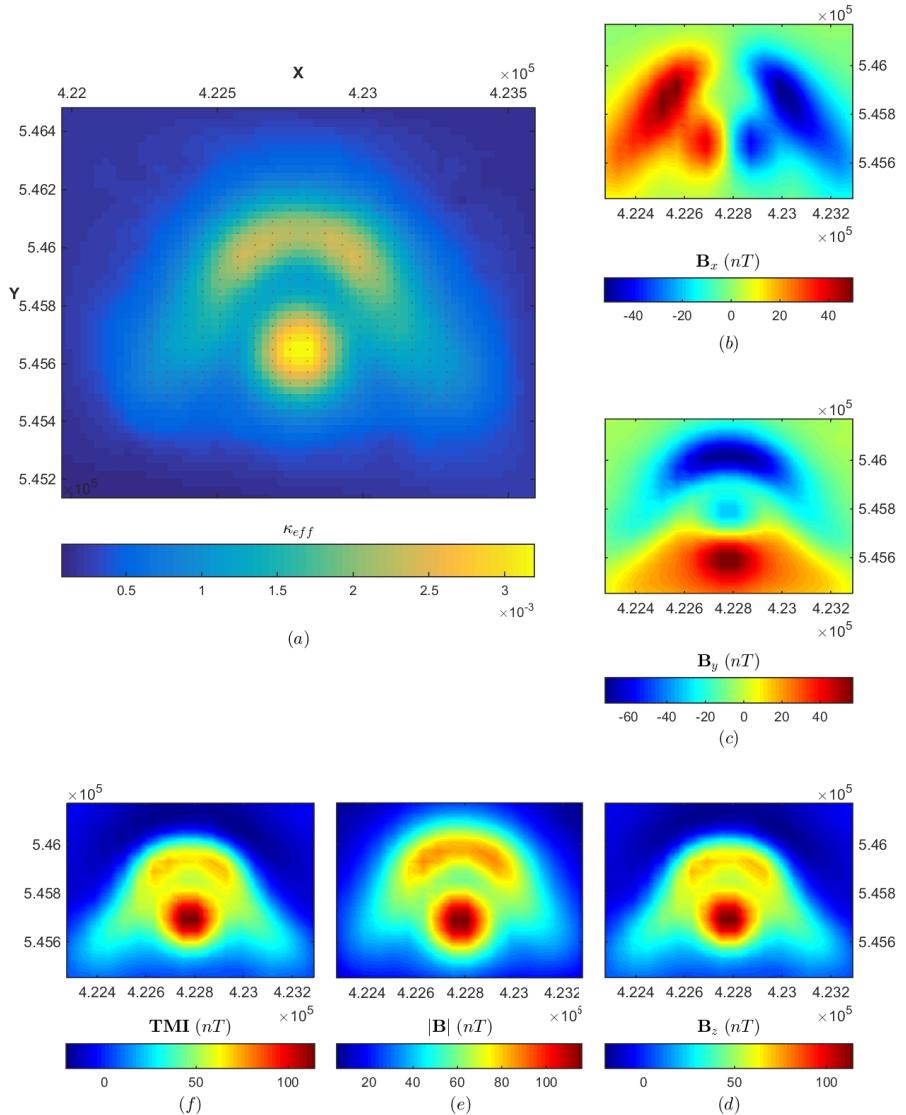


Figure 5.2: (a) Inverted equivalent-source layer and (b-f) predicted \hat{x} , \hat{y} , \hat{z} -component, TMI and magnetic amplitude data for the synthetic model.

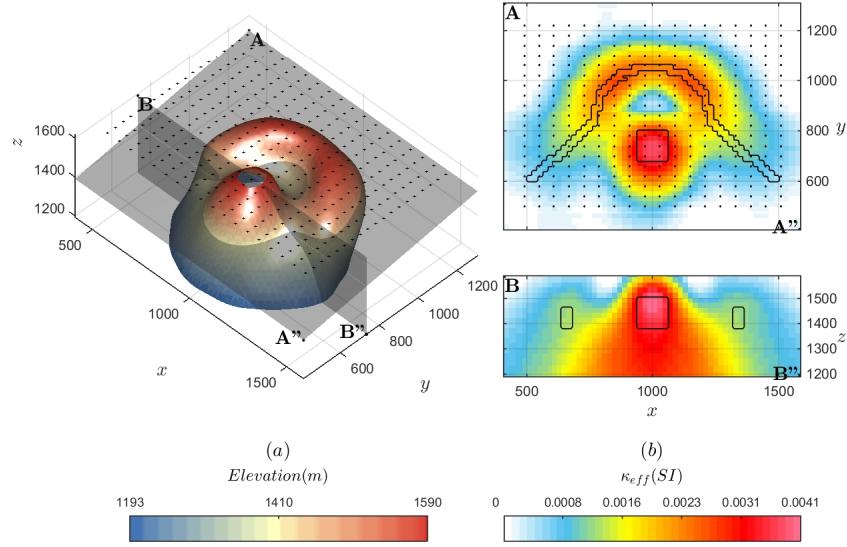


Figure 5.3: (a) Iso-surface (0.002 SI) and (b) sections through the recovered effective susceptibility model. This effective susceptibility model is used to construct a weighting matrix to constrain the MVI.

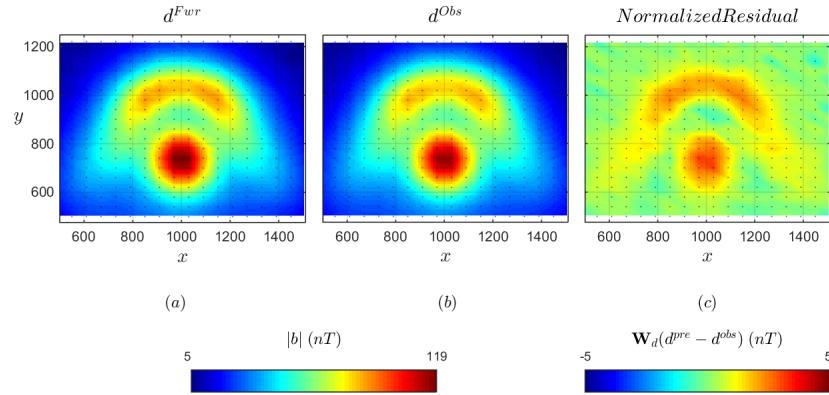


Figure 5.4: Comparison between (a) observed and (b) predicted data from the recovered effective susceptibility model. The inversion can predict most of the data within one standard deviation.

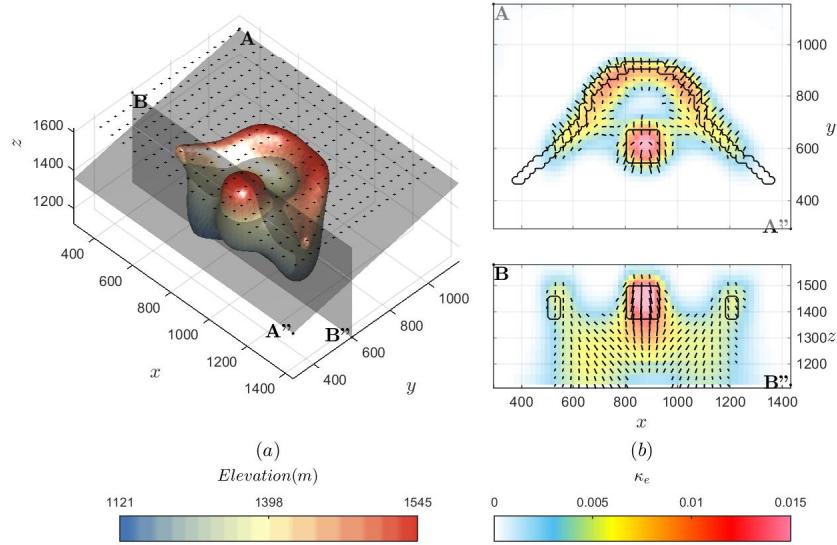


Figure 5.5: (a) Iso-surface (0.005 SI) and (b) sections through the recovered magnetization model from the CMI algorithm ($p = 2$, $q = 2$) . The inversion recovers both the arc and the block anomaly as distinct objects.

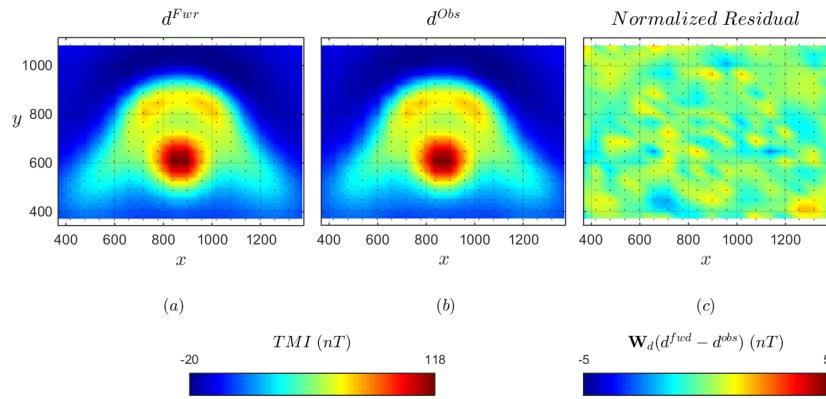


Figure 5.6: Comparison between (a) observed and (b) predicted data from the recovered susceptibility model. (c) Normalized data residuals are within two standard deviations.

5.2 CMI with S-IRLS regularization

As a final experiment I impose sparsity constraints on the amplitude inversion for ($p = 0$, $q = 1$), in order to simplify the distribution of effective susceptibility. The goal is to get a simpler effective susceptibility model to further reduce the active space of the MVI algorithm. It is a *soft* constraint on the solution as I do not impose *a priori* information to individual cells, but rather a general constraint on the behavior of the solution. Figure 5.7 presents the recovered model after convergence.

The l_p -norm constraint considerably reduces the complexity of the model, although some artifacts remain at depth from the amplitude inversion. In Stage III, the compact effective susceptibility model is used to constrain the MVI result. Figure 5.9 presents the final magnetization vector. The arc and the central block are recovered at the right location and effective susceptibility values are near the true model. From a mineral exploration perspective, the model showed in 5.9 would be easily interpreted. I note that most of the deep artifacts from the amplitude inversion have been removed. The final data residual is within one standard deviation, and no longer shows structure correlated with the magnetic anomaly.

5.3 Conclusion

In this chapter, I implemented a Cooperative Magnetic Inversion for the inversion of magnetization vectors. The algorithm ties together three inversion algorithms previously introduced in the literature: the magnetic susceptibility, magnetic amplitude and MVI. The recovered vector magnetization models are simpler and more representative of the true solution. The inversion process is automated, hence reducing the overall work required, both numerically and from the end-user.

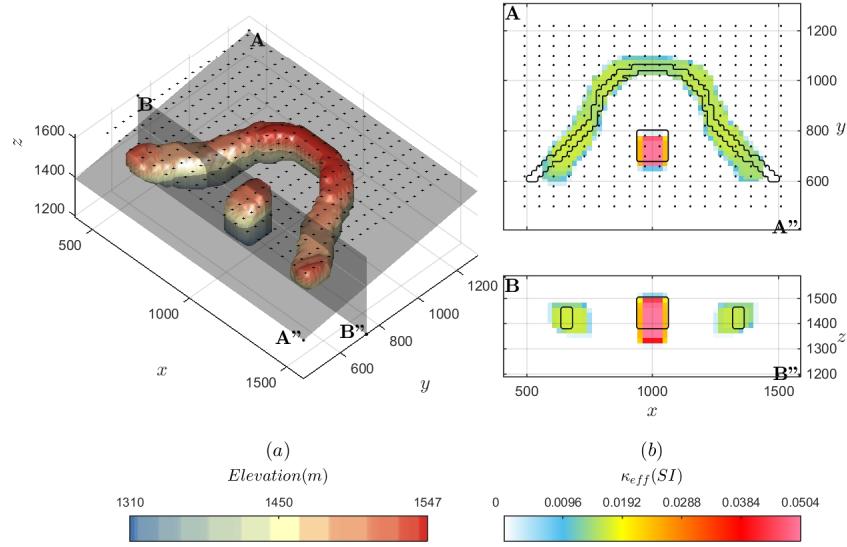


Figure 5.7: (a) Iso-surface (0.01 SI) and (b) sections through the recovered effective susceptibility model from the amplitude inversion with sparsity constraint applied ($p = 0, q = 1$). The l_p -norm constraint considerably reduces the complexity of the model, although the model is still stretched vertically.

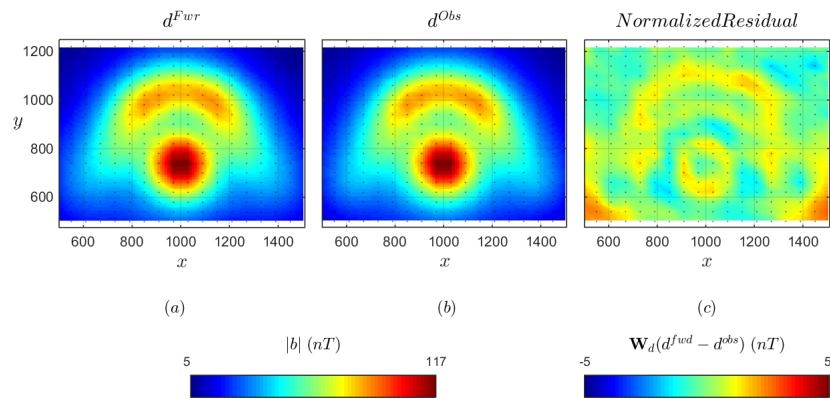


Figure 5.8: Comparison between (a) observed and (b) predicted amplitude data from the recovered compact effective susceptibility model ($p = 0, q = 1$). (c) Normalized data residuals are within two standard deviations.

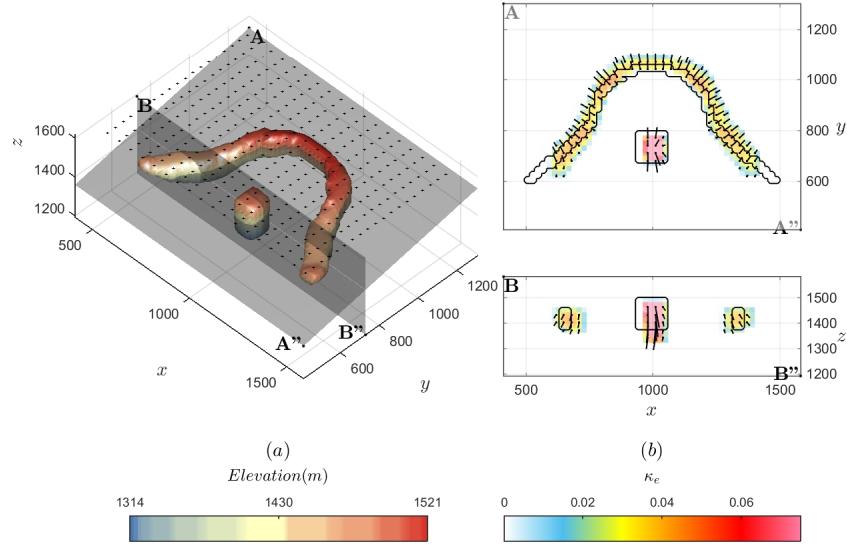


Figure 5.9: (a) Iso-surface (0.01 SI) and (b) sections through the recovered magnetization model from the CMI algorithm. Compact norms ($p = 0, q = 2$) were applied during the amplitude inversion. The l_p -norm constraint considerably reduces the complexity of the model.

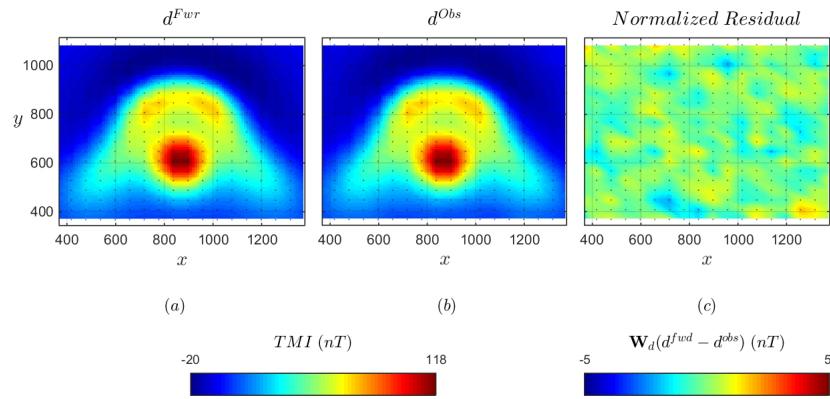


Figure 5.10: Comparison between (a) observed and (b) predicted data from the recovered magnetization CMI model. (c) Normalized data residuals are within two standard deviations. Correlated residuals are no longer seen.

Chapter 6

Case Study

In this chapter, I apply the CMI algorithm on a 1992 aeromagnetic data set collected over the Ekati Property, Northwest Territories. Airborne magnetic surveys have been an integral part of diamond exploration in the region since the early 1990s (Pell, 1997). The Lac de Gras region has been particularly productive, hosting two of the largest deposits found in Canada, the Ekati and Diavik mines. Over 150 kimberlites have been identified on the Ekati Property (Carlson et al., 2015), but diamond grades are highly variable. It is estimated that less than 10% of the 150 known kimberlites on the Ekati Property are of economic interest. While most kimberlite pipes can easily be identified by their geophysical signature, estimating the economic potential early in the exploration stage remains challenging (Coopersmith et al., 2006).

Kimberlite pipes are generally magnetic anomalies known to retain a strong remanent component. There have been several studies specifically dedicated to magnetic methods applied to kimberlite deposits. Previous geophysical studies relied primarily on data processing techniques to isolate exploration targets (Cowan et al., 2000). Direct inversion of magnetic data over remanently magnetized kimberlite pipes has long been challenging for reasons discussed in Chapter 3. In Shearer (2005), the magnetic amplitude inversion is implemented on a data set from the Galaxie trend, Nunavut. The location and geometry of negative magnetic anomalies were successfully inverted and attributed to hypabyssal kimberlite dykes, later confirmed by drilling. A few years later, Zhao (2012) inverted a subset of a large-

scale aeromagnetic survey over the Ekati Property. The inclination and amplitude of magnetization were computed for the Grizzly and Leslie pipes and compared to the 2D analysis done by Cheman (2006).

In this chapter, I invert the large data set that was partially analyzed by Zhao (2012), over the Ekati Property. Using the CMI algorithm introduced in Chapter 5, I extract regional information about the distribution of dyke swarms in relation to the known kimberlite deposits. I then proceed with deposit-scale inversions over 17 of the known pipes and recover a bulk estimate of the magnetization. The recovered orientations of magnetization are compared to physical property measurements from the literature. Finally, I compare the published radiometric dating on 11 of those pipes with the polarity predicted from the inversion and the global geomagnetic timescale.

6.1 Ekati Property

The Ekati Property is located north of the Lac de Gras region, approximately 300 km northeast of Yellowknife, Northwest Territories. The current mineral claim covers an area of over 262,000 ha. Since discovery of the first kimberlite in 1991, over 150 additional pipes were identified from geophysical surveys, till sampling and drilling programs. Figure 6.1 shows the topography and locations of known pipes over the region. The Panda and Beartooth pipes were actively mined between the period of 1998 and 2010 and are now fully depleted. Three are still under production: Misery, Koala and Koala North. Since the beginning of operations in 1999, over 58 million carats have been extracted from the various kimberlite pipes (Carlson et al., 2015). South of Ekati is the Diavik mine, which has also produced over 91 million carats between 2003 and 2014 (Yip and Thompson, 2015).

6.1.1 Geology

The Lac de Gras kimberlite field is located in the Slave craton, which consists mainly of granitoid and meta-sediment rocks dated between 2.66-2.58 Ga-old (Creaser et al., 2004). The region is intruded by several phases of diabase dyke swarms clearly visible on the airborne magnetic data. Table 6.1 summarizes the age and trend of five important groups of dykes (Buchan et al., 2009). It has been argued

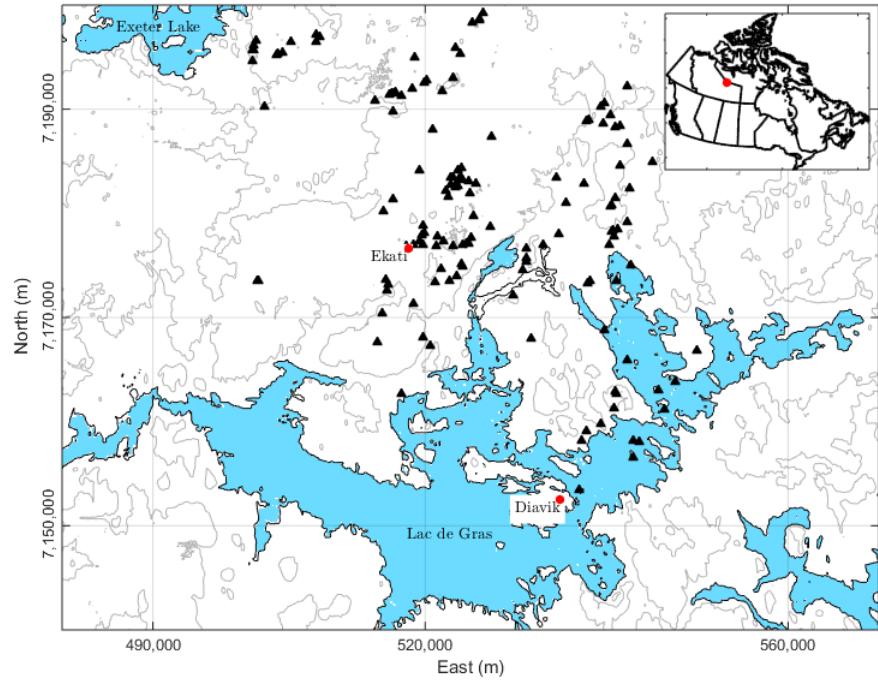


Figure 6.1: Topography and known kimberlites over the Lac de Gras region. The location of the main operations for the Ekati and Diavik mines are shown (red dot).

that the dykes may serve as structural control for kimberlites (Wright, 1999).

Kimberlite pipes in the Lac de Gras region are narrow, steeply sided inverted cones, with the majority of them covering less than 5 ha at the surface. The standard model for kimberlites in the Lac de Gras region is a mix of fragmented crater facies and non-fragmented hypabyssal (HK) facies. The crater facies in the upper portion of the type can be further subdivided in various types of volcaniclastic (VK) and pyroclastic (PK) units (Pell, 1997). Found at greater depths, the HK unit is a coherent olivine-rich rock, and is also found as intrusive dykes and sills. The composition and physical properties of kimberlites can vary greatly between pipes. Geophysically, it has been found that crater facies are generally associated with density and magnetic susceptibility lows. Weathering of the crater facies may also

Table 6.1: Dyke swarms of the Lac de Gras region listed in increasing age as published by Buchan et al. (2009)

Name	Orientation (°)	Age (Ma)
Mackenzie	315	1267
'305'	305	-
Lac de Gras	0	2027-2023
MacKay	90	2210
Malley	45	2230

appear as resistivity lows due to the abundance of clay minerals, although they can easily be confounded with fine grained glaciofluvial sediments in northern regions (Power and Hildes, 2007). The hypabyssal facies on the other hand are mostly associated with magnetic susceptibility highs and are prone to retaining remanent magnetization.

Fossil bearing xenoliths in some pipes are indicative of a sub-marine emplacement between the Late Cretaceous to Eocene time. Radiometric dating of 36 pipes indicates a wide range of emplacement times (Creaser et al., 2004), divided into five temporally discrete episodes (Lockhart et al., 2004). Three of the most productive kimberlites at Ekati (Koala, Koala North and Panda pipes) were emplaced 53 Ma ago during the Panda Age Array (PAA). Physical property measurements of the rocks were also acquired over the Lac de Gras region and are summarized in (Buchan et al., 2009; Enkin, 2003). It has been suggested that the orientation of magnetization could be used to determine the age of a kimberlite pipe, hence it could be used to estimate the economic potential of a deposit (Lockhart et al., 2004).

6.1.2 Airborne magnetic data

In 1993, a fixed-wing magnetic survey was flown over the Ekati Property as shown in Figure 6.2. The "*Paul's Lake*" aeromagnetic data were downloaded from the Natural Resources Canada's website. The data set consists of 91 survey lines of Total Magnetic Field (TMI) measurements, spaced 250 m apart, covering an area of 23×36 km. Observations were provided at a 50 m spacing along line, at an

average flight height of 120 m, as measured by radar altimeter. From historical geomagnetic data, the inducing field strength was 60,275 nT, [I: 27°, D: 84°].

A Digital Elevation Model (DEM) was downloaded from the Canadian Digital Elevation Data collection (1:50,000), also provided by Natural Resources Canada. The DEM was used to convert the radar height to absolute elevation referenced to mean sea level.

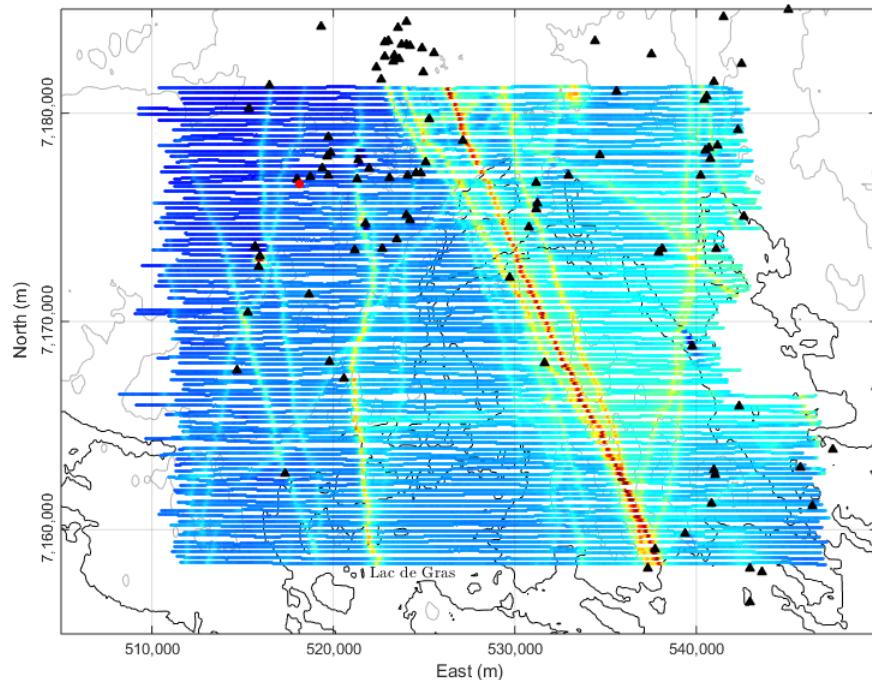


Figure 6.2: Topography, aeromagnetic survey and known kimberlites over the Lac de Gras region. The location of the main operations for the Ekati and Diavik mines is marked with a red dot.

6.2 Regional scale inversion

The inversion procedure follows the methodology presented in Chapter 5. The objective of a regional scale inversion is to first gain knowledge about the general

distribution of magnetic anomalies. Moreover, this step is used to estimate uncertainties and to identify regions of interest for a deposit-scale inversion.

Prior to the inversion, I remove a regional trend in the data in order to focus the inversion on local anomalies. Small regions 1 km^2 showing low variations in the field were manually chosen as shown in Figure 6.3(a). Within each region, the median data value and horizontal location are recorded. Those values are then used to create a first-order polynomial (Fig-6.3(b)). A regional trend of roughly 250 nT trending up towards east is subtracted as shown in Figure 6.3(c). I assign five percent uncertainties, plus 5 nT floor uncertainty to the data.

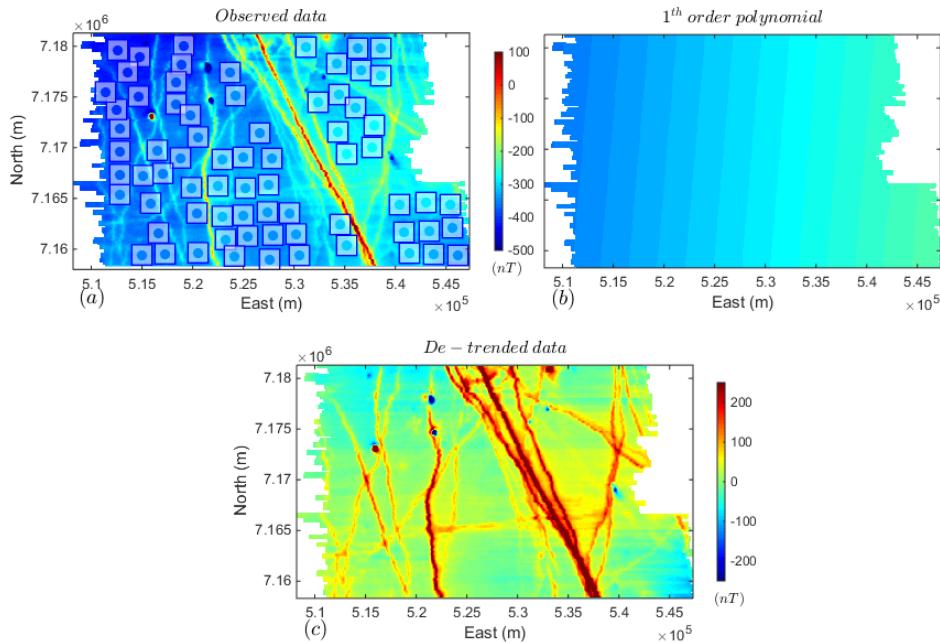


Figure 6.3: Regional data (a) before and (c) after removal of a regional trend. The median value within selected regions (boxes) were used to compute (b) a first-order polynomial trend.

Table 6.2 presents some of the parameters used for the inversion. The model space is discretized by a 50 m cell size resolution. The depth extent of the mesh covers the top 1 km below topography, for a base mesh of 6 M cells. Inversion of the entire data set, approximately 50,000 observations, at this resolution remains

Table 6.2: Regional-scale inversion parameters.

Core cell size	$75 \times 75 \times 50$ m
Base mesh size	3.3 M cells
Number of tiles	77
M / tile	$\approx 100k$ cells
N / tile	≈ 900 data
p, qx, qy, qz	1, 1, 1, 1
$\alpha_s, \alpha_x, \alpha_y, \alpha_z$	1.8e-4, 1, 1, 1
Uncertainties	0.05*d + 5 nT

computationally challenging for most computers. In order to reduce the computational cost, I designed a tiled inversion procedure. Figure 6.4 presents the tiling configuration with respect to the data set. Neighboring regions overlap by 500 m in order to share a minimum of two survey lines, guaranteeing good lateral continuity. The procedure is readily parallelized across multiple processors, each inverting a subset of the problem independently. The final magnetization model is interpolated onto the base mesh using a weighted average scheme such that:

$$m_i = \frac{\sum_{j=t}^T m_i^{(j)} w_i^{(j)}}{\sum_{j=t}^T w_i^{(j)}} \quad (6.1)$$

$$w_i^{(j)} = \left[(x_i - x_c^{(j)})^2 + (y_i - y_c^{(j)})^2 \right]^{-1/2},$$

where the weights $w_i^{(j)}$ are inverse distance between the i^{th} cell and the center location $[x_c; y_c]$ of the j^{th} tile. The number of tiles T overlapping at the i^{th} cell is variable depending on the horizontal location. The same weight is used for all cells vertically at the same $[x_i; y_i]$ location.

Figure 6.5 presents a horizontal section at ≈ 50 m depth through the recovered effective susceptibility model. As expected from the observed data, several long and narrow anomalies are recovered. Based on the analysis by Buchan et al. (2009), I interpret the linear trends as regional dyke swarms belonging to the Mackenzie, Lac de Gras, MacKay and Malley groups. A large number of kimberlite pipes show

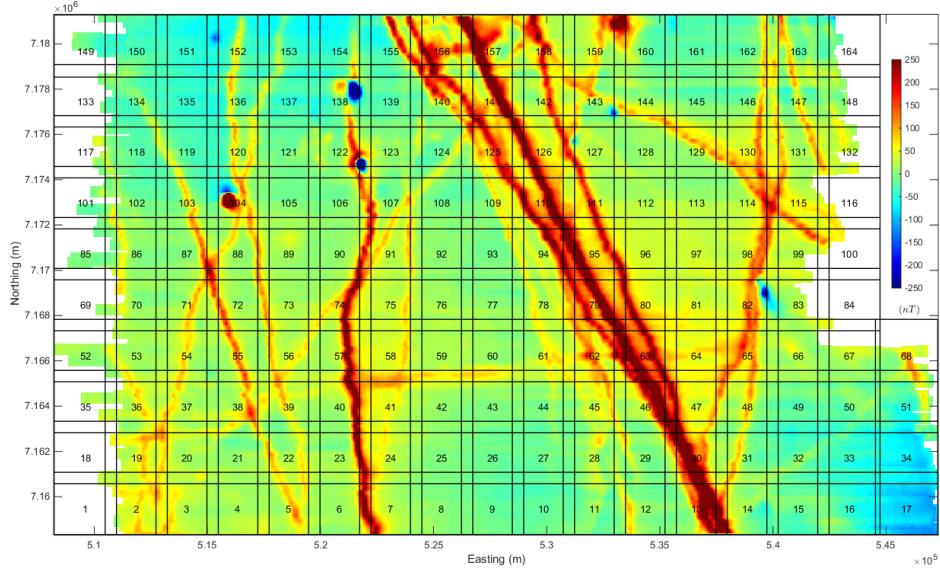


Figure 6.4: TMA data after removal of the regional signal and tiling configuration used for the regional inversion.

as effective susceptibility highs, with few exceptions. The three Koala pipes, for example, west of the Grizzly pipe, were not well recovered. Although important economically, the anomalous magnetic response from the pipe is below the noise level. Proper imaging of the deposits would require a magnetic survey at lower elevation to increase the noise to signal ratio.

I note the clustering of kimberlite pipes near the intersection and along some of the dyke swarms. The apparent correlation between kimberlite pipes and dykes seems to reinforce the idea of structural controls as put forward by Wright (1999). While the topic is beyond the scope of this Masters thesis, future work in structural geology over the region may benefit from a regional scale inversion as presented here.

Figure 6.6 compares the observed and predicted data from the tiled inversion. In the left panel, the predicted data from individual tiles were merged following 6.1. The normalized residual data show horizontal striations due to variations in magnetic data between adjacent lines, which could not be accounted for by the inversion. Overall, each individual inversion could predict the data well within

the assigned uncertainties. On the other hand, the right panel shows the predicted data from the final merged susceptibility model. I here notice that some of the low frequency content has been lost during the merging process, with higher residuals over the regions of low magnetic fields. This can be a problem when attempting to image regional features and will require further research. In this case, we are only interested in the dykes and kimberlite pipes, which for the most part were well recovered.

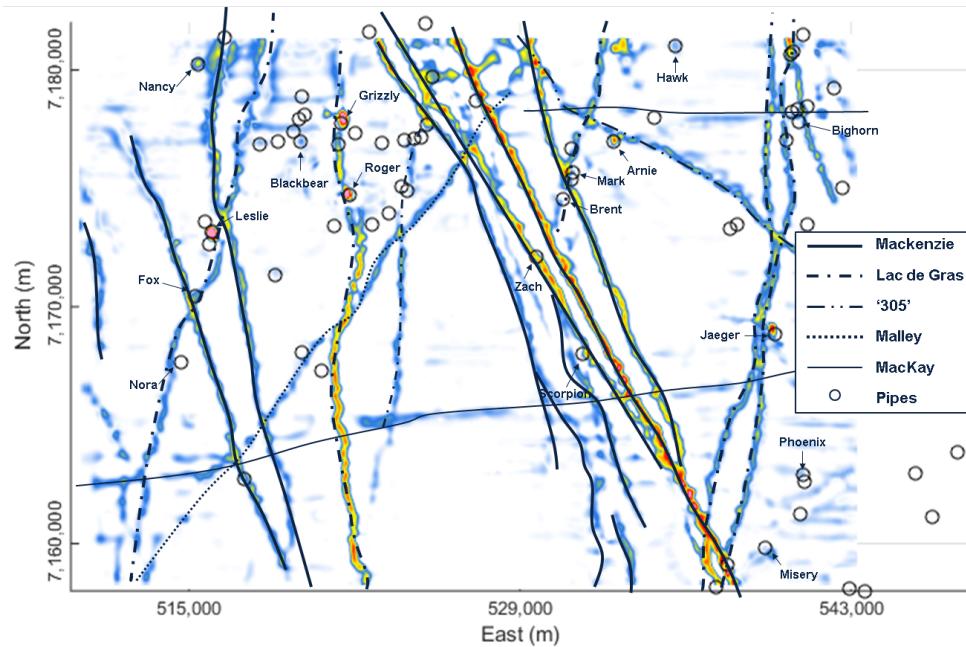


Figure 6.5: Interpreted dykes (line) from the property-scale magnetization inversion and known kimberlite pipes location (circle). The 16 pipes chosen for a deposit-scale inversion are labeled.

6.3 Deposit scale inversion

From the property scale inversion, 16 sites were chosen based on the strength of the magnetic anomaly for a high resolution inversion. The goal of this section is to gain knowledge about the bulk magnetization direction and compare those values to physical property measurements published in the literature.

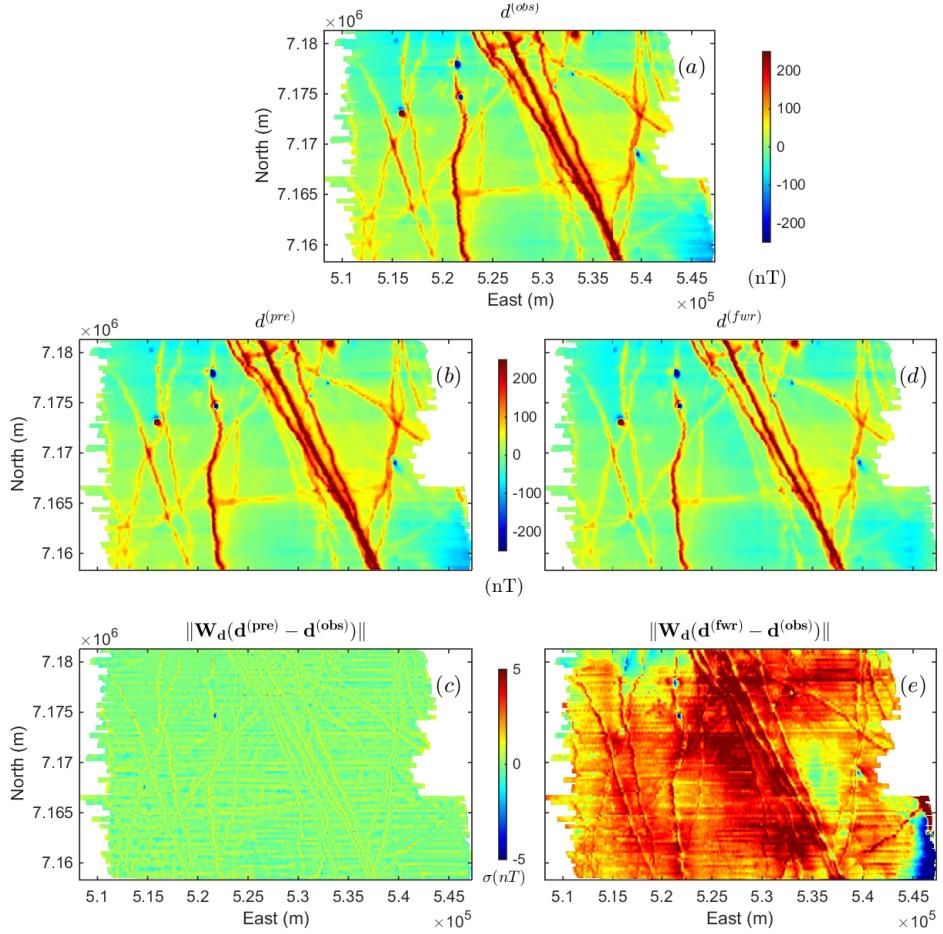


Figure 6.6: Comparison between (a) observed and (b) merged predicted data from the recovered magnetization model over the Ekati Property. (c) Normalized residual data show horizontal striations due to variations in magnetic data between adjacent lines, which could not be accounted for by the inversion. Most of the predicted data from the individual tiles can fit the observed data within one standard deviation. (d) Forward modeled data from the merged magnetization model and (e) normalized data residual. It appears that a large portion of the low frequency content has been lost during the merging step. More research is required in order to preserve the long wavelength information.

For each inversion, a regional field was removed in order to focus the inversion on the local anomalies. Description of the regional field removal method can be found in Li and Oldenburg (1998). In summary, the following steps were taken:

- Creation of a local mesh with large padding cells
- Inversion of the data assuming induced magnetization
- Removal of the local source within the core mesh. Only susceptible material outside the core region remain.
- Forward modeling of the regional data at the observation location
- Subtraction of the regional signal from the observed data

The regional-removal step is important in regions where the regional field is larger than the local anomaly. Figure 6.7 presents the magnetic data before and after removal of the regional field over the Misery pipe. Note that the original data are trending up towards the east, with the largest magnetic field data occurring near the edge of the dataset. Consequently, the magnetic amplitude inversion would be biased towards having a ring of effective susceptibility around the edge of the dataset, rather than the pipe itself. After removing the regional field, most of the anomalous fields come from a strong negative anomaly corresponding with the location of the Misery pipe.

Following the regional field removal, each local tile is inverted with the CMI algorithm on a 25 m cell size mesh. This resolution is required in order to properly model narrow anomalies in the range of 50 to 100 m in width. The process was repeated for the selected 16 kimberlite pipes. Table 6.3 summarizes the parameters used for the inversions. I imposed an l_0 -norm on the amplitude of magnetization in order to focus the inversion within the core of the different kimberlites pipes. The direction of magnetization was allowed to vary smoothly in order to get an estimate of the variability, and hence to get a measure of uncertainty. A 27 point average (3x3x3 cells) centered over the location of each pipe was used to compute an effective susceptibility and direction as shown in Table 6.4. In the following section, the results are compared to those published in the literature.

Table 6.3: Deposit-scale inversion parameters.

Core cell size	$25 \times 25 \times 25$ m
Base mesh size	50 M cells
Number of tiles	17
M / tile	$\approx 200k$ cells
N / tile	≈ 150 data
p, qx, qy, qz	0, 2, 2, 2
$\alpha_s, \alpha_x, \alpha_y, \alpha_z$	1.6e-3, 1, 1, 1
Uncertainties	0.02*d + 5 nT

Table 6.4: Average magnetization amplitude and direction from the local inversion over 16 known kimberlite pipes. Uncertainties were calculated from a three-cell cube standard deviation around the approximate location of each kimberlite pipe.

Pipe ID	X (m)	Y (m)	κ_e	Incl (°)	Decl (°)
Arnie	532980	7176930	$3.9\text{e-}2 \pm 1.0\text{e-}2$	-85 ± 4.2	355 ± 5.0
Blackbear	519700	7176980	$9.9\text{e-}4 \pm 1.3\text{e-}4$	-88 ± 1.9	211 ± 3.0
Brent	530855	7174206	$4.6\text{e-}3 \pm 2.0\text{e-}3$	-79 ± 3.1	299 ± 2.3
Fox	515091	7169946	$2.4\text{e-}2 \pm 8.0\text{e-}3$	66 ± 17.7	203 ± 14.9
Grizzly	521529	7177946	$8.7\text{e-}2 \pm 8.6\text{e-}3$	-82 ± 1.3	100 ± 1.7
Hawk	535550	7180934	$2.0\text{e-}3 \pm 1.6\text{e-}4$	-84 ± 0.2	244 ± 0.1
Jaeger	539575	7168970	$2.3\text{e-}2 \pm 6.6\text{e-}3$	-73 ± 2.6	53 ± 2.8
Leslie	515960	7173130	$3.6\text{e-}1 \pm 6.0\text{e-}2$	65 ± 3.7	346 ± 5.2
Mark	531202	7175715	$4.5\text{e-}2 \pm 7.0\text{e-}3$	-83 ± 4.0	159 ± 5.4
Nancy	515351	7180271	$1.0\text{e-}2 \pm 8.0\text{e-}4$	-84 ± 2.1	11 ± 2.0
Nora	514665	7167532	$2.0\text{e-}3 \pm 2.6\text{e-}4$	-87 ± 2.1	287 ± 4.3
Phoenix	540829	7162932	$1.5\text{e-}4 \pm 2.1\text{e-}5$	-85 ± 1.8	238 ± 1.3
Roger	521770	7174655	$1.3\text{e+}0 \pm 4.2\text{e-}1$	-82 ± 5.3	160 ± 10.9
Scorpion	531854	7167847	$2.5\text{e-}2 \pm 1.1\text{e-}2$	68 ± 13.0	258 ± 23.6
Zach	529812	7171900	$3.4\text{e-}2 \pm 8.8\text{e-}3$	82 ± 7.9	185 ± 9.4
Misery	539621	7159630	$7.3\text{e-}3 \pm 1.3\text{e-}3$	-77 ± 5.0	172 ± 10.1

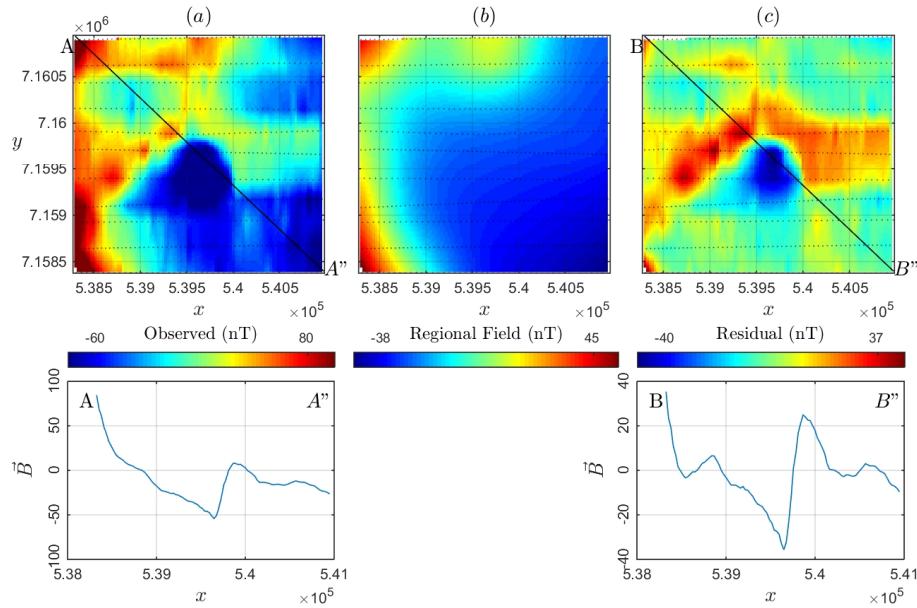


Figure 6.7: (a) Local data collected over the Misery pipe showing a local western trend. (b) Regional field data are computed from a local inversion. Most of the signal comes from a dyke running north-south along the western edge of the local tile. (c) Residual data after regional field removal, showing a clear reversely magnetized anomalous corresponding with the location of the Misery pipe.

6.4 Discussion

Magnetization inversion at a regional scale reveals interesting patterns. Figure 6.8 compares the induced and remanent components of magnetization over the property. The remanent component is noticeably stronger on most of the NE trending Mackenzie dykes, while the Lac de Gras and MacKay dykes seem to be mainly magnetized along the inducing field. Several points of intersection between the Lac de Gras, Mackenzie and Malley dykes in the south of survey [537,000 E ; 7,176,000 N] show a strong remanent component. This result seems to agree with the study of Buchan et al. (2009). The low resolution of this airborne survey does not support accurate modeling of narrow dykes on the order of 10 to 30 m width. It can however provide a relative time of emplacement when comparing the strength

and trend of effective susceptibility of intersecting dykes. Note the clear magnetic overprinting of a NW trending dyke onto a NE dyke near the Fox pipe [515,000 E ; 7,170,000 N] and Leslie pipe [516,600 E ; 7,174,000 N]. This result fits with the idea of a younger Mackenzie dyke swarm (1270 Ma) crosscutting an older Lac de Gras dyke (2020 Ma).

Note that the NS trending dyke intercepting the Grizzly pipe has been recovered on its entire length with a relatively strong induced component. This is despite the large footprint of a reversely magnetized body, which illustrates the strength of the CMI algorithm in its ability to distinguish between adjacent anomalies with varying magnetic orientations.

Several of the 16 pipes have been studied in the past, some of which have borehole samples measured in laboratory. Table 6.5 summarizes the information collected from the literature. In general terms, my results seem to agree more with Enkin (2003) and Cheman (2006) in terms of orientation compared to the results of Zhao (2012). It is important to point out that the inversion results offer a bulk estimate of magnetization within the pipe, which are known to be highly heterogeneous. Laboratory measurements provided by Enkin (2003) showed large variations in magnetic properties along individual boreholes. Moreover, larger uncertainties are expected from the recovered declination in cases where the magnetization vector is nearly vertical ($I > 80^\circ$).

Finally, I analyze the age of the kimberlite pipes in relation to the magnetization direction. Table 6.6 summarizes the age group as identified by Lockhart et al. (2004), as well as the predicted polarity from the geomagnetic timescale provided by Cande and Kent (1995). Figure 6.9 maps on a timescale the polarity of the various pipes with respect to Earth's polarity reversal. It is interesting to note that several of the predicted polarity directions from radiometric dating do not agree with the recovered polarity. Those differences could be a result of uncertainties in the Rb-Sr radiometric dating of ± 1 Ma (2σ).

It is important to note the frequent changes in Earth's polarity during the period of 46 and 69 Ma, with at least 14 reversals identified. My results seem to suggest that the age array MAA should be between 48 and 49 Ma in order to account for the recovered negative polarity. Similarly, the Leslie pipe should be either younger than 51.7 or older than 52.7 to be normally magnetized. The Roger pipe

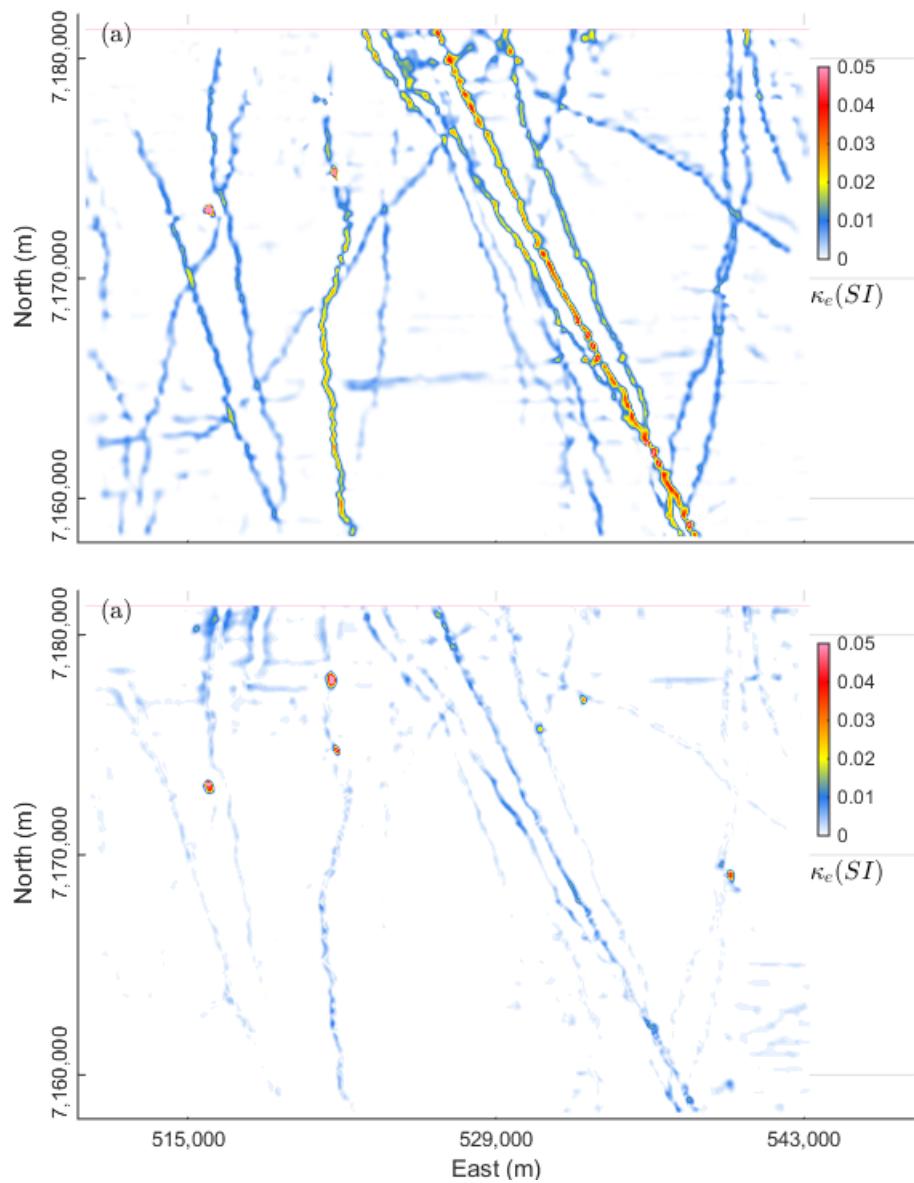


Figure 6.8: Horizontal sections through the recovered (a) induced and (b) remanent magnetization model. Several pipes with strong remanence can easily be identified as discrete circular anomalies.

Table 6.5: Summary of published results over selected kimberlite pipes

Source	Leslie $I^\circ D^\circ$		Misery $I^\circ D^\circ$		Grizzly $I^\circ D^\circ$	
CMI	65	346	-77	172	-82	100
Enkin (2003)	64	329	-74	168		***
Zhao (2012)	59	334		***	-58	316
Cheman (2006)	82	***		***	-73	***

is particularly interesting as it is clearly reversely magnetized, and could have only been formed during the short time window of 67.61 to 67.73 Ma.

As suggested by Lockhart et al. (2004), the microdiamond abundance of kimberlite pipes observed at Ekati may be correlated in time. It has been found however that the magmatism can vary greatly between each eruption episode, even if all episodes occur in close spatial proximity. Future research could look into correlations between magnetization strength and direction, as well as electrical properties of the different pipes. Assuming that diamond grade and abundance are related to the mantle source, it can be expected that rock magnetic susceptibility and electrical conductivity would be a better proxy for diamond content than age of emplacement. High resolution magnetic and airborne EM inversions could lead to interesting insights into the Ekati property.

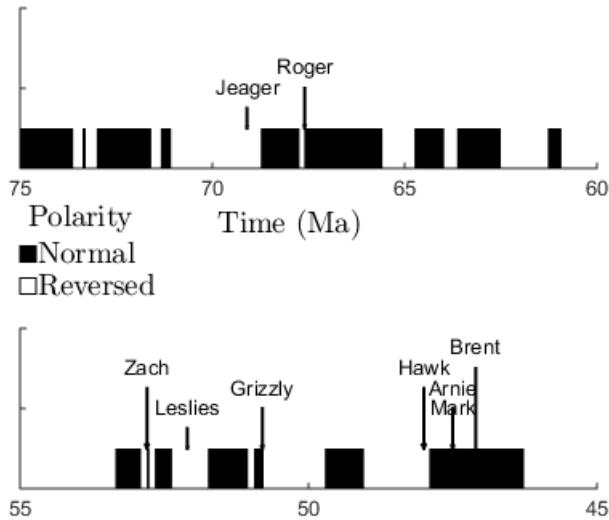


Figure 6.9: Age of 11 pipes from the Ekati region with respect to Earth's polarity reversal.

Table 6.6: Radiometric age and inverted magnetization direction for 11 pipes.

Pipe ID	Age Array Lockhart et al. (2004)	Age (Ma)	Polarity Chron (+, -)	Incl (°)	Decl (°)
Brent	MAA	47.1	C21n (+)	-79.0 ± 3.1	299.0 ± 2.3
Arnie		47.5	C21n (+)	-85.9 ± 4.2	355.5 ± 5.0
Mark		47.5	C21n (+)	-83.9 ± 4.0	159.1 ± 5.4
Hawk		48.0	(-)	-84.3 ± 0.2	244.4 ± 0.1
Grizzly		50.8	C23n.1n (+)	-82.2 ± 1.3	100.1 ± 1.7
Leslie	PAA	52.1	(-)	65.4 ± 3.7	346.3 ± 5.2
Zach		52.8	C24n.2n (+)	82.5 ± 7.9	185.8 ± 9.4
Roger	MSK	67.6	C30n (+)	-82.7 ± 5.3	160.1 ± 10.9
Jaeger		69.1	(-)	-73.2 ± 2.6	53.2 ± 2.8
Nora		2106	-	-87.4 ± 2.1	287.9 ± 4.3
Misery		2480	-	-77.6 ± 5.0	172.3 ± 10.1

Chapter 7

Conclusion

The inversion of magnetic data affected by remanent magnetization is an active field of research. Several inversion strategies have been proposed in the past. The magnetic amplitude inversion of Shearer (2005) inverts for the distribution of effective susceptibility from magnetic amplitude data. The algorithm is robust and can define regions of high effective susceptibility, but no information is provided regarding the orientation of magnetization. On the other hand, the Magnetic Vector Inversion (MVI) of Lelièvre and Oldenburg (2009) has the ability to recover the magnetization direction over variable and complicated geology, but the algorithm requires *a priori* information to constrain the solution. Moreover, most inversion algorithms use l_2 -norm measures of model structure for regularization, yielding smooth and small models.

In this thesis, I combined the magnetic amplitude inversion and the MVI algorithm into a Cooperative Magnetic Inversion (CMI) algorithm, in order to improve the robustness and versatility of current magnetic inversion codes used separately. I developed a flexible l_p -norm regularization that allows for sparse and blocky models. The regularization function can be applied on sub-regions of the model domain, allowing for a smooth transition in norm penalties. Magnetic amplitude data are computed by the Equivalent Source method adapted from Li and Oldenburg (2010). I tested the algorithm on a synthetic model, showing improvement over the MVI method alone, yielding a simpler and more compact magnetization model. This in turn can help differentiate between neighboring anomalies with variable

magnetization directions.

Finally, I implemented the CMI algorithm on a large aeromagnetic survey of the Ekati Property. Information regarding the polarity and orientation of magnetization for 16 kimberlite deposits are compared to values from the literature. Relative age of dyke swarms are inferred from the apparent overprinting of magnetization. The inverted magnetization model yielded information about possible distribution of reversed magnetization at a regional level.

7.1 Future work

This thesis opens up several avenues for future research. While the CMI algorithm has the potential to estimate the magnetization vector orientation, there is still a need to differentiate between the induced and NRM components. In cases where the NRM is parallel to the present inducing field, the current method cannot differentiate between the induced and remanent part of magnetization. One option would be to use frequency-domain electromagnetic (FEM) data to estimate the magnetic susceptibility of rocks. FEM surveys are routinely used in mineral exploration projects to identify conductive anomalies. Inversion of FEM data could serve the dual purpose of providing susceptibility and conductivity values, as well as additional geometrical constraints for the inversion. In the case of kimberlite exploration, FEM data could also be used to characterize the electrical conductivity of kimberlite pipes. Combined magnetization and conductivity information may help distinguish between different pipes and estimate the diamondiferous potential.

The CMI algorithm could be implemented on other types of mineral deposits, as well as on global geophysical problems. The classic problem of ocean floor magnetization would be an interesting subject to revisit with this inversion technique. Accurate modeling of the magnetization direction may provide important insights into the transient behavior of Earth's magnetic field during polarity changes.

From a general inverse problem standpoint, the mixed-norm regularization implemented by the S-IRLS method could be implemented in a broad variety of inverse problems in geophysics. More specifically, the regularization may be applied to constrain non-linear inverse problems such as encountered in electromagnetics.

The issue regarding the vertical stretching experienced with the amplitude in-

version needs to be addressed. The closer the recovered effective susceptibility is to the true model, the more accurate the MVI solution will be. There might be a need to iterate between the amplitude and MVI steps in order to progressively refine the magnetization model.

The Cartesian formulation for the MVI is not well suited for compact norm constraints. Sparsity constraints on the individual components of magnetization tend to yield polarized models along the $\{\hat{p}, \hat{s}, \hat{t}\}$ directions. Future work should investigate the spherical formulation of the MVI, which would allow application of sparsity constraints directly on the amplitude of magnetization. In the spherical formulation, the magnetization vector is parameterized as an amplitude and two angles (Lelièvre and Oldenburg, 2009). This will allow additional information to be incorporated through a reference model with the constraint information coming from the amplitude of the magnetization or the Koenigsberger ratio of rock samples.

The current CMI code suffers from memory limitations due to the large size of the sensitivity matrix. Incorporating wavelet compression, as seen in Li and Oldenburg (2003), may greatly reduce the computational time and size of the problem. Some of these limitations may also be overcome by working in the differential equation domain (Davis et al., 2013). While the tiled inversion procedure proposed in Chapter 6 was effective in reducing the overall size of the inverse problem, there were several issues with the method. The robustness and accuracy of the final model depends on the distance of overlap between the tiles, as well as on the merging strategy. A more efficient and rigorous approach would be to separate the inversion mesh from the forward modeling mesh. This approach was recently proposed to solve large-scale EM problems (Haber and Schwartzbach, 2014; Yang and Oldenburg, 2014). The inversion code could handle large data sets discretized on a fine mesh, would be highly parallelizable and could potentially eliminate the need for compression.

Bibliography

- Ajo-Franklin, J. B., B. J. Minsley, and T. M. Daley, 2007, Applying compact constraints to differential traveltome tomography: *Geophysics*, **72**, 67–75. → pages 56, 58, 59, 68
- Barbosa, V. C. F., and J. B. C. Silva, 1994, Generalized compact gravity inversion: *Geophysics*, **59**, 57–68. → pages 56, 68
- Bhattacharyya, B. K., 1964, Magnetic anomalies due to prism-shaped bodies with arbitrary magnetization: *Geophysics*, **29**, 517–531. → pages 43
- Blakely, R., 1996, Potential theory in gravity and magnetic applications: Cambridge University Press. → pages 3
- Buchan, K. L., A. N. LeCheminant, and O. van Breemen, 2009, Paleomagnetism and U-Pb geochronology of the Lac de Gras diabase dyke swarm, Slave Province Canada: Implications for relative drift of Slave and Superior provinces in the Paleoproterozoic: *Canadian Journal of Earth Sciences*, **46**, 361–379. (NRC Research Press). → pages viii, 31, 111, 113, 116, 122
- Campbell, W. H., 1997, Introduction to geomagnetic fields, second edition ed.: Cambridge University Press. → pages 2, 14
- Cande, S. C., and D. V. Kent, 1995, Revised calibration of the geomagnetic polarity timescale for the late cretaceous and cenozoic: *Journal of Geophysical Research*, **100**, 6093–6095. → pages ix, 4, 123
- Carlson, J. A., P. J. Ravenscroft, C. Lavoie, and J. Cumming, 2015, Ekati diamond mine, Northwest Territories, Canada: Ni 43-101 technical report, Dominion Diamond Corporation. → pages 110, 111
- Chartrand, R., 2007, Exact reconstruction of sparse signals via nonconvex minimization: *IEEE Signal Processing Letters*, **76**, 167–188. → pages 56, 58, 69
- Cheman, A., 2006, Inversion magnetique tridimensionnelle des anomalies kimberlitique isolees avec presence ou sans remanence: Msc a, Ecole Polytechnique de Montreal. → pages 111, 123, 125
- Clark, D. A., 1991, Magnetic petrology of igneous intrusions: implications for exploration and magnetic interpretation: *Exploration Geophysics*, **30**, no. 2,

5–26. → pages 3

- Coopersmith, H., J. Pell, and B. Scott Smith, 2006, The importance of kimberlite geology in diamond deposit evaluation: A case study from the DO27/DO18 kimberlite, NWT, Canada: Presented at the 2006 Kimberlite Emplacement Workshop, Saskatoon. → pages 110
- Cowan, D. R., L. A. Tompkins, and S. Cowan, 2000, Screening kimberlite magnetic anomalies in magnetic active areas: *Exploration Geophysics*, **31**, 66–72. → pages 5, 110
- Creaser, R. A., H. Grutter, J. Carlson, and B. Crawford, 2004, Macrocrystal phlogopite RbSr dates for the Ekati property kimberlites, Slave Province, Canada: Evidence for multiple intrusive episodes in the Paleocene and Eocene: *Lithos*, **76**, 399–414. → pages 111, 113
- Dampney, C. N. G., 1969, The equivalent source technique: *Geophysics*, **34**, 39–53. → pages 44
- Dannemiller, N., and Y. Li, 2006, A new method for estimation of magnetization direction: *Geophysics*, **71**, L69–L73. → pages 6
- Daubechies, I., R. DeVore, M. Fornasier, and C. S. Gunturk, 2010, Iteratively re-weighted least squares minimization for sparse recovery: *Communication on Pure and Applied Mathematics*, **63**, 1–35. → pages 53
- Davis, K., E. Haber, and D. W. Oldenburg, 2013, Large-scale magnetic inversion using differential equations and octrees: Presented at the 23rd Geophysical Conf. and Exhibition, Australian Soc. of Expl. Geophys. → pages 129
- Doyle, B. J., K. Kivi, and B. H. S. Smith, 1999, The Tli Kwi Cho (DO27 and DO18) Diamondiferous Kimberlite Complex, Northwest Territories, Canada: Proceedings of the VIIth International Kimberlite Conference, 194–204. → pages 95
- Dransfield, M., A. Christensen, and G. Liu, 2003, Airborne vector magnetic mapping of remanently magnetized banded iron formation at Rocklea, Western Australia: *Exploration Geophysics*, **34**, 93–96. → pages 5
- Ekblom, H., 1973, Calculation of linear best l_p -approximation: *BIT*, **13**, 292–300. → pages 54
- Ellis, R. G., B. de Wet, and I. N. Macleod, 2012, Inversion of magnetic data from remanent and induced sources: Presented at the 22th International Geophysical Conference and Exhibition, ASEG. → pages 7, 8
- Enkin, R., 2003, Paleomagnetic analysis of selected Ekati kimberlite. Unpublished report for BHP Billiton Diamonds. Geological Survey of Canada. → pages 113, 123, 125
- Enkin, R. J., 2014, The rock physical property database of British Columbia, and the distinct petrophysical signature of the Chilcotin basalts.: *Canadian Journal of Earth Sciences*, **51**, 327–338. (NRC Research Press). → pages 5, 31

- Farquharson, C. G., and D. W. Oldenburg, 1998, Non-linear inversion using general measures of data misfit and model structure: *Geophysical Journal International*, **134**, 213–227. → pages 53
- Gorodnitsky, I. F., and B. D. Rao, 1997, Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm: *IEEE Transactions on Signal Processing*, **45**, 600–616. → pages 68
- Haber, C., and C. Schwartzbach, 2014, Parallel inversion of large-scale airborne time-domain electromagnetic data with multiple OcTree meshes: *Inverse Problems*, **30**, 1–28. → pages 129
- Kubota, R., and A. Uchiyama, 2005, Three-dimensional magnetization vector inversion of a seamount: *Earth Planets Space*, 691–699. → pages 7
- Last, B. H., and K. Kubik, 1983, Compact gravity inversion: *Geophysics*, **48**, no. 8, 713–721. → pages 54, 56, 59, 68
- Lawson, C. L., 1961, Contribution to the theory of linear least maximum approximation: PhD thesis, University of California, Los Angeles. → pages 54
- Lelièvre, P., 2003, Forward modelling and inversion of geophysical magnetic data: Master's thesis, University of British Columbia. → pages 2, 24
- Lelièvre, P. G., 2009, Integrating geological and geophysical data through advanced constrained inversion: Phd thesis, The University of British Columbia. → pages 7, 8, 31, 34
- Lelièvre, P. G., and D. W. Oldenburg, 2009, A 3D total magnetization inversion applicable when significant, complicated remanence is present: *Geophysics*, **74**, 21–30. → pages 7, 8, 127, 129
- Li, Y., 2012, Recent advances in 3D generalized inversion of potential-field data: *SEG Technical Program Expanded Abstracts*, 1–7. → pages 6
- Li, Y., M. N. Nabighian, and D. Oldenburg, 2014, Using and equivalent source with positivity for low-latitude reduction to the pole without stration: *Geophysics*, **79**, J81–J90. → pages 44, 47
- Li, Y., and D. W. Oldenburg, 1996, 3-D inversion of magnetic data: *Geophysics*, **61**, 394–408. → pages 5, 8, 18, 25, 26, 53, 60, 79
- , 1998, Separation of regional and residual magnetic field data: *Geophysics*, **63**, 431–439. → pages 120
- , 2003, Fast inversion of large-scale magnetic data using wavelet transforms and a logarithmic barrier method: *Geophysics Journal International*, **152**, 251–265. → pages 24, 129
- , 2010, Rapid construction of equivalent sources using wavelets: *Geophysics*, **75**, no. 3, L51–L59. → pages 7, 8, 44, 127
- Li, Y., S. E. Shearer, M. M. Haney, and N. Dannemiller, 2010, Comprehensive approaches to 3D inversion of magnetic data affected by remanent magnetization: *Geophysics*, **75**, L1–L11. → pages 5, 38

- Liu, S., X. Hu, Y. Xi, T. Liu, and S. Xu, 2015, 2D sequential inversion of total magnitude and total magnetic anomaly data affected by remanent magnetization: *Geophysics*, **80**, K1–K12. → pages 8
- Lockhart, G., H. Grutter, and J. Carlson, 2004, Temporal, geomagnetic and related attributes of kimberlite magmatism at Ekati, Northwest Territories, Canada: *Lithos*, **77**, 665–682. → pages 113, 123, 125, 126
- Nabighian, M. N., 1972, The analytic signal of two-dimensional magnetic bodies with polygonal cross-section: *Geophysics*, **37**, 507–517. → pages 6, 7, 38
- Nabighian, M. N., V. J. S. Grauch, R. O. Hansen, T. R. LaFehr, Y. Li, J. W. Peirce, J. D. Phillips, and M. E. Ruder, 2005, The historical development of the magnetic method in exploration: *Geophysics*, **70**, no. 6, 33ND–61ND. → pages 1
- Pell, J. A., 1997, Kimberlites in the Slave Craton, Northwest Territories, Canada: *Geoscience Canada*, **24**, 77–90. → pages 110, 112
- Phillips, J. D., 2003, Can we estimate total magnetization directions from aeromagnetic data using Helbig's formulas: Abstracts Week B, IUGG 2003, B261. → pages 6
- Pilkington, M., 1997, 3-D magnetic imaging using conjugate gradients: *Geophysics*, **62**, 1132–1142. → pages 25
- Portniaguine, O., and M. S. Zhdanov, 2002, 3-D magnetic inversion with data compression and image focusing: *Geophysics*, **67**, 1532–1541. → pages 58, 69, 80
- Portniaguine, O. N., 1999, Image focussing and data compression in the solution of geophysical inverse problems: Phd thesis, Department of Geology and Geophysics, The University of Utah. → pages 54, 69
- Power, M., and D. Hildes, 2007, Geophysical strategies for kimberlite exploration in northern Canada: Proceedings of Exploration '07: Fifth Decennial International Conference on Mineral Exploration, 1025–1031. → pages 113
- Roest, W. R., J. Verhoef, and M. Pilkington, 1992, Magnetic interpretation using the 3-D analytic signal: *Geophysics*, **57**, 116–125. → pages 6
- Sharma, P. V., 1966, Rapid computation of magnetic anomalies and demagnetization effects caused by bodies of arbitrary shape: *Pure Appl. Geophys.*, **64**, 89–109. → pages 11
- Shearer, S., 2005, Three-dimensional inversion of magnetic data in the presence of remanent magnetization: Master's thesis, Colorado School of Mines, Golden, Colorado. → pages 7, 8, 38, 110, 127
- Shewchuk, J. R., 1994, An introduction to the Conjugate Gradient method without the agonizing pain: School of Computer Science. → pages 21
- Stocco, S., A. Godio, and L. Sambuelli, 2009, Modelling and compact inversion of magnetic data: A Matlab code: *Computer and Geosciences*, **35**, 2111–2118.

→ pages 56, 59, 68

Sun, J., and Y. Li, 2014, Adaptive l_p inversion for simultaneous recovery of both blocky and smooth feature in geophysical model: *Geophysical Journal International*, **197**, 882–899. → pages 53, 58, 68, 82

Tikhonov, A. N., and V. Y. Arsenin, 1977, *Solution of Ill-posed Problems*: Winston Press. → pages 17, 50

van der Vorst, H. A., 2003, *Iterative Krylov methods for large linear systems*: Cambridge University Press. → pages 23

Vine, F. J., and D. H. Matthews, 1963, Magnetic anomalies over the oceanic ridges: *Nature*, **199**, 947–949. → pages 1

Vogel, C. R., 2002, *Computational Methods for Inverse Problems* (Frontiers in Applied Mathematics): Society for Industrial Mathematics. → pages 24

Wilkinson, L., B. A. Kjarsgaard, A. N. LeCheminant, and J. Harris, 2001, Diabase dyke swarms in the Lac de Gras area, Northwest Territories, and their significance to kimberlite exploration: initial results: *Current Research 2001-C8*, Geological Survey of Canada. → pages 95

Wright, K. J., 1999, Possible structural controls of kimberlites in the Lac de Gras Region, Central Slave Province, Northwest Territories, Canada: Msc thesis, Queen's University. → pages 112, 117

Yang, D., and D. W. Oldenburg, 2014, 3-d inversion of airborne electromagnetic data parallelized and accelerated by local mesh and adaptive soundings: *Geophysical Journal International*, **196**, 1492–1507. → pages 129

Yip, C. G., and K. S. Thompson, 2015, Diavik diamond mine, Northwest Territories, Canada: Ni 43-101 technical report, Diavik Diamond Mines Inc. → pages 111

Zhao, P., 2012, Three dimensional inversion of magnetic survey data collected over kimberlite pipes in presence of remanent magnetization: Msc. a thesis, Ecole Polytechnique de Montreal. → pages 110, 111, 123, 125

Zhdanov, M., and E. Tolstaya, 2004, Minimum support nonlinear parametrization in the solution of a 3d magnetotelluric inverse problem: *Inverse Problems*, **20**, 937–952. → pages 72