

**JOINT INVERSION FOR CORRELATED MODELS IN LINEAR
INVERSE PROBLEMS**

By

Yidong Liu

B. Sc. (Mathematics) Shandong University, 1985

M. Sc. (Mathematics) Academia Sinica, 1988

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

in

THE FACULTY OF GRADUATE STUDIES
DEPARTMENT OF MATHEMATICS & INSTITUTE OF APPLIED MATHEMATICS

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

November 1994

© Yidong Liu, 1994

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Mathematics

The University of British Columbia
Vancouver, Canada

Date Dec 12, 1994

Abstract

Inverse problems arise in many fields. They are usually ill-posed since they often violate one or more of Hadmard's three conditions for well-posedness: existence, uniqueness and stability. In this thesis, we propose a new method for computing approximate solutions in certain linear inverse problems.

We consider linear inverse problems based on integral equations of the first kind. Analysis of Picard's condition reveals that such equations may lead to ill-posed problems which may have no solution satisfying the observed data exactly and stably, but may have infinitely many solutions satisfying the data approximately. To get a unique and stable solution to this kind of inverse problem, we use Tikhonov's Regularization Method.

To obtain the best possible approximation to the true model, we should use any and all available information regarding the true model, although we can not expect to get sufficient data. For example, it is standard practice to use the positivity of the model in inverting magnetic and IP data, and to use special weighting functions in solving magnetic problems. The key feature of the present work is a method that exploits the correlation between different model parameters in inverting the geophysical data.

To keep different parameters in suitable confidence regions, a new methodology, Combined Inversion, is developed. In combined inversion, different kinds of data are inverted simultaneously. The objective functional imposing the correlation requirement may be neither convex nor quadratic, so corresponding algorithm and code are developed. When the objective functional is not quadratic, we use an iterative method to solve it and approximate the functional with its second order Taylor approximation in each iteration. When the objective functional is not convex, it may have more than one local minimum.

To get the minimum which well approximates the true model we should begin with a good initial model. In our case we produce the initial model by solving the combined problem with no correlation requirement.

We introduce our method in the context of two practical geophysical inverse problems: the magnetic problem and the Induced Polarization (IP) problem. As we expect, regularization smooths the inverted models, so some model characteristics are lost in the recovered models. Our numerical examples confirm the smoothing effects of the regularization operators.

Since magnetic susceptibility and chargeability are negatively correlated, we introduce a nonquadratic, nonconvex "correlation function", whose sub-level sets define confidence regions for the vector of susceptibility and chargeability. Then we require our recovered models to be in the confidence region. The recovered models from combined inversion method are significantly better than those from independent inversion.

This method should be useful in practical prospecting when several kinds of data are available and there is some correlation among the parameters. This is the case in mining industry where several kinds of geophysical data are usually measured at the same time and the different parameters producing the data are known to be correlated. If we approximate the correlation with a reasonable functional, we may reconstruct models satisfying the corresponding correlation.

Table of Contents

Abstract	ii
List of Tables	vii
List of Figures	viii
Acknowledgement	x
1 Introduction	1
1.1 Linear Inverse Problems In Geophysics	1
1.2 Two Practical Problems	2
1.2.1 Magnetic Problem	2
1.2.2 Induced Polarization Problem	5
1.3 Combined Inversion	8
2 Linear Inverse Problems	12
2.1 Ill-posed Inverse Problems: Functional Theory	12
2.1.1 Theoretical Analysis of an Inverse Problem	13
2.1.2 Generalized Solutions and Regularization of Ill-posed Problems . .	15
2.1.3 Construction of Regularization	19
2.2 Solving Inverse Problems	22
2.2.1 Discretization of Inverse Problems	22
2.2.2 Direct Methods	25
2.2.3 Iterative Methods	27

2.3	Linearizing Nonlinear Inverse Problems	31
3	Two Practical Problems	34
3.1	Magnetic Methods	34
3.1.1	Physical Principles	34
3.1.2	Mathematical Relations	37
3.1.3	Forward Problem	38
3.1.4	Inverse Problem	40
3.1.5	A Numerical Example	44
3.2	Induced Polarization	50
3.2.1	Physical Principles	50
3.2.2	Mathematical Relationship	53
3.2.3	The Forward Problem	55
3.2.4	The Inverse Problem	57
3.2.5	A Numerical Example	61
4	Combined Inversion	65
4.1	Known Approaches	65
4.2	Correlation Between Parameters	66
4.3	Combined Inversion	68
4.3.1	Optimization Problem	68
4.3.2	Statistical Description	72
4.4	Computing Techniques	74
4.4.1	Optimization Problem	74
4.4.2	Approximating the Nonlinear Correlation	75
4.4.3	Producing the Initial Model	76
4.4.4	Choosing the Correlation Weighting Parameter	78

4.4.5	Balancing the Regularization Operator	79
4.4.6	Working with SLIM	80
4.5	A Numerical Example	81
4.6	Conclusion	83
Bibliography		92

List of Tables

4.1	Chargeability and susceptibility of some minerals.	67
4.2	Parameters of regularization operator.	81
4.3	Output with $\alpha_c = 0$	82
4.4	Output with $\alpha_c \neq 0$	83

List of Figures

1.1	Four-electrode DC model.	5
1.2	DC current for IP.	6
1.3	IP decay voltage.	6
3.4	Geomagnetic field.	35
3.5	Rectangle geometry	39
3.6	Base function model.	41
3.7	Exact magnetic data from the forward problem.	45
3.8	Noisy data simulating magnetic observations.. . . .	45
3.9	Magnetic Model.	47
3.10	Predicted data from the output model.	48
3.11	Difference between observed data and predicted data.	48
3.12	Four-electrode DC model.	51
3.13	IP decay voltage.	51
3.14	DC current for IP.	54
3.15	IP Model.	64
4.16	Reconstructed Models with $\phi_c = 0$	85
4.17	Regularization term for magnetic part, ϕ_κ	86
4.18	Regularization term for IP part, ϕ_η	86
4.19	Total Misfit, ϕ_d	87
4.20	Correlation Function, $\phi_c(m)$	87
4.21	Total sum of the objective function, ϕ	88

4.22	Total sum of the objective function, ϕ .	88
4.23	Correlation Function, $\phi_c(m)$.	89
4.24	Total misfit, ϕ_d .	89
4.25	Reconstructed Models with $\phi_c = \kappa^2 \eta^2$.	90
4.26	Reconstructed Models with ϕ_c =function product.	91

Acknowledgement

I would like to thank my supervisors, Dr. Philip Loewen and Dr. Doug Oldenburg, for their guidance and support throughout my work.

I also like to express my gratitude to the friends at the Department of Mathematics and the Department of Geophysics and Astronomy who provide me friendly and supportive environment.

I have no doubt that without the strong and continued support from my family, I would not have been able to complete this work as it stands. Their love outlasted the dark clouds. I owe much to them and so does this thesis.

Finally I like to thank my wife Rui Zhang who gives me invaluable encouragement and help in these years.

Chapter 1

Introduction

This thesis concerns Linear Inverse Problems and their applications in Geophysics. Its main contribution is a combined inversion method that exploits known correlations between the underlying physical properties in linear inverse problems. After introducing the basic principles of linear inverse problems in part one, we work on two practical geophysical inverse problems in part two: Magnetic susceptibility inversion and Induced Polarization (IP) chargeability inversion. In part three we develop the combined inversion method and demonstrate its usefulness by inverting the susceptibility and chargeability jointly, since there is a correlation between these two parameters.

1.1 Linear Inverse Problems In Geophysics

Geophysical Inverse Problems are to determine, on the basis of measured data, internal properties of the Earth which cannot be measured directly, such as the mass density or the electrical conductivity. They are usually ill-posed as mathematical problems, since they violate one or more of the three properties of existence, uniqueness and stability.

For a linear inverse problem, we have to estimate some physical parameter $m(x)$, i.e., a function of spatial coordinates, using only the measured data d and a known linear operator A . Usually m is in an infinite dimensional Hilbert space, so we can not expect to evaluate it exactly from an inaccurate, finite-dimensional vector of data d . This is an ill-posed problem.

In practical work, we discretize the function m first: physical laws then take the form

of a system of linear equations:

$$Gm = d. \quad (1.1)$$

Usually G is an $N \times M$ matrix and $N \ll M$. Since there are measurement errors in the observed data d , we may not find any solution vector m satisfying the system. Even if we can satisfy the observed data exactly, we should not solve the system in that way since that kind of solution is usually unstable with respect to measurement errors. Therefore, the discrete problem is still ill-posed. The theory of regularization suggests that we determine an approximate solution by solving the following optimization problem:

$$\begin{aligned} &\text{minimize} \quad \phi_m(m) := \|W_m(m - m_0)\|^2, \\ &\text{subject to} \quad \phi_d(m) := \|W_d(Gm - d^{obs})\|^2 = \phi_d^*. \end{aligned} \quad (1.2)$$

Here m_0 is called the base model, ϕ_d is called the misfit criterion, W_d is called the data weighting matrix, W_m is called the model weighting matrix and ϕ_d^* is called the target misfit.

1.2 Two Practical Problems

1.2.1 Magnetic Problem

A typical linear inverse problem in geophysics is to calculate the magnetic susceptibility in magnetic prospecting. Magnetic prospecting is one of the oldest methods of geophysical exploration and is used to explore for oil, minerals, and even archaeological artifacts.

A magnetizable body placed in an external magnetic field becomes magnetized by induction; the magnetization is due to the reorientation of atoms and molecules so that their spins line up. The alignment of internal dipoles produces a field \vec{M} , which is added to the magnetizing field \vec{H} . For low magnetic fields, \vec{M} is proportional to \vec{H} and is in the direction of \vec{H} . The degree to which a body is magnetized is determined by its *magnetic*

susceptibility κ , which is defined by

$$\vec{M} = \kappa \vec{H}. \quad (1.3)$$

Here \vec{M} , \vec{H} and κ may all vary with position. Magnetic susceptibility is the significant variable in magnetics. It plays the same role as the density does in the gravitational problem.

A volume of magnetic material can be considered as an assortment of magnetic dipoles that results from the magnetic moments of individual atoms and dipoles. They are aligned by induction in the presence of the Earth's magnetic field ("the main field") \vec{H}_e ($\mu_0 |\vec{H}_e| \approx 0.5 \times 10^{-4}$ tesla) to produce a local magnetic anomaly. In magnetic prospecting we estimate the magnetic susceptibility κ by measuring the local magnetic anomaly.

Magnetic anomalies are caused by magnetic minerals (mainly magnetite and pyrrhotite) contained in the rocks. In any case, we can regard a magnetic body as a continuous distribution of dipoles resulting in a vector dipole moment per unit volume, \vec{M} . Each dipole at the origin causes an anomalous field at \vec{r} according to

$$\begin{aligned} \vec{F}(\vec{r}) &= \nabla \{ \vec{M}(\vec{r}) \cdot \nabla (1/|\vec{r}|) \} \\ &= \nabla \{ \kappa \vec{H}_e \cdot \nabla (1/|\vec{r}|) \} \text{ when } \vec{M} = \kappa \vec{H}_e. \end{aligned} \quad (1.4)$$

Magnetic anomalies caused by intrusions, flows, or iron-rich sedimentary horizons are common features in regions favorable for mineral exploration, and there is frequently a contrast between the magnetic mineral content of such features and that of the host rock. Usually magnetic polarization is in the \vec{H}_e direction and $\vec{M} = \kappa \vec{H}_e$. To get the data about our model, we partition the area of interest into a series of cells and assume that the susceptibility in each cell is constant. Thus, we can calculate the projections of the magnetic anomaly field in the direction of Earth's magnetic field.

With the magnetic anomaly field \vec{F} as the measured data and the corresponding

physical law (1.4) as the linear relation we get a standard linear inverse problem. It can be solved for the unknown functional κ through general methods for inverse problems.

In practical magnetic prospecting, data are gathered using magnetometers [4]. One type of magnetometer is a magnetic balance which measures the anomaly field values relative to a reference value. This type of magnetometer measures the magnetic anomaly field \vec{F} directly.

The other type of magnetometer measures absolute values of the magnetic field. Since this type of magnetometers are robust and extensively used, we show how can we compute one component of the anomalous field from the absolute value of the magnetic field. Indeed, if \vec{H}_e is the main field and \vec{H}_a is the anomalous field, then we can determine ΔH experimentally, where

$$\begin{aligned}\Delta H &= |\vec{H}_e + \vec{H}_a| - |\vec{H}_e| \\ &= (\vec{H}_e + \vec{H}_a, \vec{H}_e + \vec{H}_a)^{1/2} - |\vec{H}_e| \\ &\approx [(\vec{H}_e, \vec{H}_e) + 2(\vec{H}_e, \vec{H}_a)]^{1/2} - |\vec{H}_e| \quad (\text{since } |\vec{H}_a| \ll |\vec{H}_e|) \\ &\approx (\vec{H}_e, \vec{H}_a) / |\vec{H}_e| \\ &= (\vec{H}_a, \frac{\vec{H}_e}{|\vec{H}_e|}).\end{aligned}$$

That is, ΔH can be used as the approximation of the projection of anomalous field in the \vec{H}_e direction. Now the measured data are actually the magnetic flux density \vec{B} : we calculate the magnetic field from

$$\vec{H} = \vec{B} / \mu_0,$$

where $\mu_0 = 4\pi \times 10^{-7}$ weber/ampere-meter is the permeability of free space.

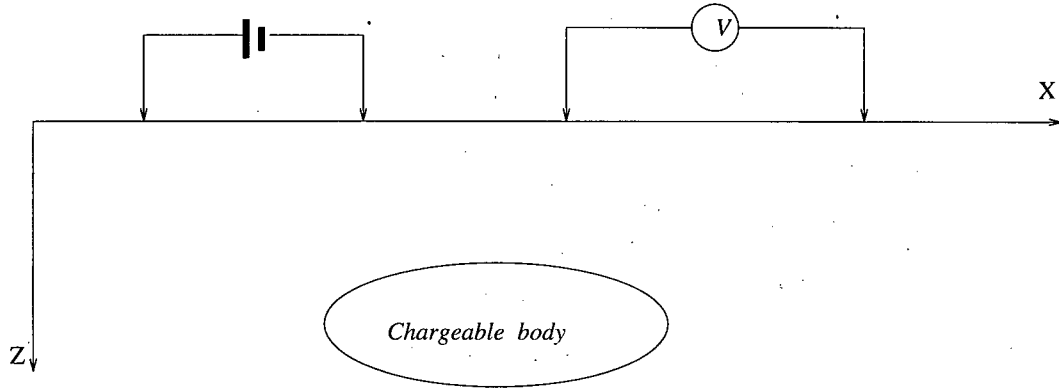


Figure 1.1: Four-electrode DC model.

1.2.2 Induced Polarization Problem

The other inverse problem we will consider in this thesis is to calculate the chargeability η in Induced Polarization (IP). Induced Polarization is a relatively new technique in geophysics; and has been employed mainly in base-metal exploration and to a minor extent in ground water search.

An illustration of induced polarization can be obtained with a standard four-electrode Direct Current (DC) resistivity spread by abruptly interrupting the current [20] (figure 1.1, 1.2). The voltage across the potential electrodes generally does not drop to zero instantaneously, but decays rather slowly, after an initial large decrease from the original steady-state value. This decay time is on the order of seconds. If the current is switched on again, the potential, after a sudden initial increase, builds up over a similar time interval to the original DC amplitude (figure 1.2, 1.3).

Measurements of IP data may be made in either the time or the frequency domain. Although the phenomenon is the same, different parameters are used to represent it. We use the chargeability defined by Siegel [21]:

$$\eta = \frac{\varphi_{\eta} - \varphi_{\sigma}}{\varphi_{\eta}}, \quad (1.5)$$

where φ_{σ} is the potential in the absence of the polarization effect and φ_{η} is the the

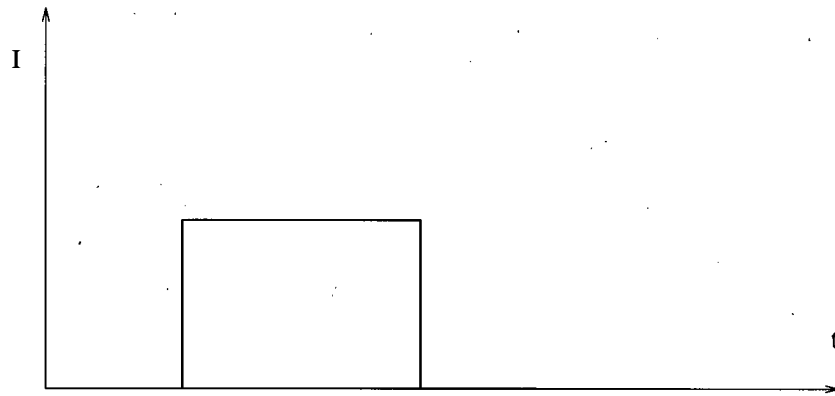


Figure 1.2: DC current for IP.

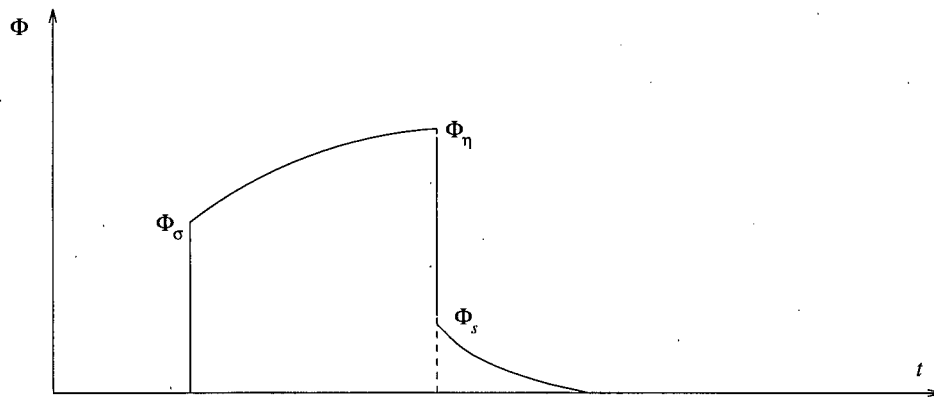


Figure 1.3: IP decay voltage.

potential with the polarization effect.

To estimate the chargeability η in the area of interest, we measure the potentials φ_σ and φ_η on the surface first and then compute the data η_a (apparent chargeability) using the equation

$$\eta_a = \frac{\varphi_\eta - \varphi_\sigma}{\varphi_\eta}. \quad (1.6)$$

For two-dimensional problems if we choose y in the strike direction, z downward and x perpendicular to the strike direction, then the chargeability is represented as $\eta(x, z)$ and we only measure φ_σ and φ_η along a line in the x direction.

The definition of apparent chargeability (1.6) shows that estimating the chargeability η from the data η_a is a nonlinear inverse problem. However, η values are usually small, so we can linearize equation (1.6) to produce a linear inverse problem that approximates the given problem.

Linearizing the potential φ_η about the conductivity model σ yields

$$\varphi_\eta = \varphi(\sigma - \eta\sigma) = \varphi(\sigma) - \varphi'(\sigma)\eta\sigma + o(\eta). \quad (1.7)$$

Substituting into equation (1.6) yields

$$\eta_a = \frac{\varphi_\eta - \varphi_\sigma}{\varphi_\eta} \approx \frac{-\varphi'(\sigma)\eta\sigma + o(\eta)}{\varphi(\sigma) - \varphi'(\sigma)\eta\sigma} \quad (1.8)$$

This can be approximately written as

$$\eta_a \approx -\frac{\sigma\varphi'(\sigma)}{\varphi(\sigma)}\eta = J\eta, \quad (1.9)$$

for a linear operator J independent of η . In the discrete version of the problem, J becomes a matrix—see Section 3.2.4.

So, instead of solving the nonlinear inverse problem (1.5) we can solve the linear inverse problem

$$J\eta = \eta_a. \quad (1.10)$$

with the same methods used for magnetic inverse problem.

Our inverse problems are always of the underdetermined type in which there are fewer observations than unknowns. Then the observations are not sufficient to determine the unknown parameters uniquely; additional information is needed. This lack of information cannot be remedied by any mathematical techniques. For geophysical inverse problems the physical parameters to be determined should be consistent with the geological structures and other information when it is available. Therefore, it is natural to construct optimization problems with proper objective functionals based on additional information.

1.3 Combined Inversion

In our case, although we can invert the IP data for chargeability and the magnetic data for susceptibility independently, we note that there is some correlation between the chargeability and the susceptibility of different kinds of rocks. Our main contribution is to use prior information about this correlation to invert these two sets of data simultaneously.

In fact, the tables in [20] shows that although there is a great variation in the κ -values, even for a particular rock type, and a wide overlap between different rock types, sedimentary rocks have the lowest average susceptibility and basic igneous rocks the highest. In every case the susceptibility depends only upon the amount of ferrimagnetic minerals present, mainly magnetite, sometimes ilmenite or pyrrhotite. The κ values for chalcopyrite and pyrite are typical of many sulphide minerals which are basically nonmagnetic.

At the same time, we find that the chargeabilities η of minerals increase with the sulphide concentration. In contrast to their magnetic susceptibilities, chalcopyrite and pyrite are typical minerals with high chargeabilities.

Therefore, the magnetic susceptibility and the chargeability of minerals are not independent. Although the correlation between susceptibility and chargeability may depend on the underlying rock type and an exact description of this correlation is not available, we know that the susceptibility and chargeability can not both be large at the same place. In this thesis we use this correlation to invert jointly the magnetic and IP data.

Suppose $\phi_c(\kappa, \eta)$ represents the correlation between κ and η , and the confidence region is determined as

$$\{(\kappa, \eta) : \phi_c(\kappa, \eta) \leq \phi_c^*\}. \quad (1.11)$$

If the discretized magnetic problem is

$$G\kappa = d_1^{obs}$$

and its corresponding regularized optimization problem is

$$\begin{aligned} & \text{minimize} \quad \phi_\kappa(\kappa) := \|W_\kappa(\kappa - \kappa_0)\|^2, \\ & \text{subject to} \quad \phi_d(\kappa) := \|W_{d_1}(G\kappa - d_1^{obs})\|^2 = \phi_{d_1}^*, \end{aligned} \quad (1.12)$$

and the discretized IP problem is

$$J\eta = d_2^{obs}$$

and its corresponding regularized optimization problem is

$$\begin{aligned} & \text{minimize} \quad \phi_\eta(\eta) := \|W_\eta(\eta - \eta_0)\|^2, \\ & \text{subject to} \quad \phi_d(\eta) := \|W_{d_2}(J\eta - d_2^{obs})\|^2 = \phi_{d_2}^*, \end{aligned} \quad (1.13)$$

then we can introduce a new optimization problem:

$$\begin{aligned} & \text{minimize} \quad \phi_m(m) := \|W_m(m - m_0)\|^2 + \alpha_c \phi_c(m) \\ & \text{subject to} \quad \phi_d(m) := \|W_d(Am - d^{obs})\|^2 = \phi_d^*, \end{aligned} \quad (1.14)$$

where

$$\begin{aligned}
 m &= (\kappa, \eta)^T, \\
 d^{obs} &= (d_1^{obs}, d_2^{obs})^T, \\
 \phi_d^* &= \phi_{d1}^* + \phi_{d2}^*, \\
 A &= \begin{pmatrix} G & 0 \\ 0 & J \end{pmatrix}, \\
 W_d &= \begin{pmatrix} W_{d1} & 0 \\ 0 & W_{d2} \end{pmatrix}, \\
 W_m &= \begin{pmatrix} W_\kappa & 0 \\ 0 & W_\eta \end{pmatrix}.
 \end{aligned}$$

If we treat η and κ as independent by setting $\alpha_c=0$, solving (1.14) is equivalent to solving for the two vectors η and κ independently except for the computing difficulty in keeping the two individual misfits at a reasonable level. But if we insist on a correlation between η and κ by choosing $\alpha_c > 0$, we can keep the parameters η and κ in a particular confidence region by choosing the proper correlation weighting factor α_c .

For a suitable Lagrange multiplier parameter μ , optimization problem (1.14) corresponds to the following unconstrained optimization:

$$\begin{aligned}
 \text{Minimize } \phi &= \phi_m(m) + \mu(\phi_d(m) - \phi_d^*) + \alpha_c \phi_c(m) \\
 &= \|W_m(m - m_0)\|^2 + \mu(\|W_d(Am - d^{obs})\|^2 - \phi_d^*) + \alpha_c \phi_c(m) \quad (1.15)
 \end{aligned}$$

Determining the correct value of μ is a separate issue that we discuss below. If $\phi_c(m)$ is a quadratic convex functional, optimization problem (1.15) can be solved as a general optimization problem of regularization for each α_c . In our case, $\phi_c(m)$ is the functional neither convex nor quadratic, so a corresponding algorithm must be developed to solve the optimization problem (1.15).

Since $\phi_c(m)$ is not a quadratic functional of m , we use an iterative method to solve the optimization problem. In each step, $\phi_c(m)$ is approximated with its second-order Taylor expansion

$$\phi_c(m) \approx \phi_c(m^{(n)}) + g^T(m - m^{(n)}) + 1/2(m - m^{(n)})^T H(m - m^{(n)}), \quad (1.16)$$

where g and H are the gradient and the Hessian matrix of ϕ_c at $m^{(n)}$ respectively. To realize this kind of approximation in each step, we modify the subspace method [15] which is designed specifically for linear inverse problems. To effectively reduce the correlation ϕ_c in each iteration, we include the gradient of ϕ_c in the subspace from which a new perturbation is chosen.

Since $\phi_c(m)$ is not a convex functional, a good initial approximation must be chosen. We use the solution to (1.15) when $\alpha_c = 0$ as the initial approximation. To obtain successively better reconstructed models, we must carefully choose $\alpha_c > 0$. In fact, the objective functional ϕ in (1.15) is not convex, so it may have different local minima. Furthermore, the regularization term ϕ_m and correlation ϕ_c in (1.15) produce contradictory effects on our model m , so α_c must be chosen such that these effects are balanced in the iteration proceeds. In our numerical experiments the correlation decreases as the iteration, so α_c must be increased to keep the correlation term and the regularization term compatible.

In the central numerical example of this thesis, we set up both a magnetic model and an induced polarization model in a common two dimensional domain in which κ and η satisfy a certain correlation ϕ_c . First, we invert κ and η independently according to the methods described above. Then we invert for the two parameters together as in problem (1.14). We compare the results produced by these two different methods and show that the joint inversion gives much better results.

Chapter 2

Linear Inverse Problems

2.1 Ill-posed Inverse Problems: Functional Theory

In 1923, the French mathematician J. Hadamard [7] introduced the concepts of well-posed and ill-posed problems. A mathematically well-posed problem should satisfy the following basic conditions:

1. A solution exists (existence).
2. The solution is uniquely determined by the data (uniqueness).
3. The solution depends continuously on the data (stability).

If a problem does not satisfy all three of conditions, it is called ill-posed. According to Hadamard's philosophy, ill-posed problems are actually ill posed, in the sense that their basic formulation is wrong. However, since Hadamard's time many applications in industry have led to the formulation of ill-posed problems where the solution has a well-defined physical meaning. Such applications arise quite naturally in science, technology and medicine when one is confronted with the need to interpret measurements. Such interpretation is particularly difficult in geophysics, medicine and astrophysics. The problem here is to determine, on the basis of the measured data, internal parameters such as the mass density $\rho(x)$ or the electrical conductivity $\sigma(x)$, which are difficult or impossible to measure directly. Such problems are briefly called inverse problems. In fact, studying such problems is the only way of completely analyzing experimental results.

Most inverse problems do not fulfil Hadamard's postulates of "well-posedness". They might not have a solution in the strict sense, their solutions might not be unique, or they might depend discontinuously on the data.

2.1.1 Theoretical Analysis of an Inverse Problem

Since many linear processes can be modeled by the integral equations of the first kind, it is interesting for us to show that such equations usually lead to ill-posed inverse problems.

An integral equation of the first kind has the form

$$(Tx)(s) = \int_0^1 k(s, t)x(t)dt = b(s), s \in [0, 1] \quad (2.17)$$

where $k(s, t) \in L^2[0, 1] \times [0, 1]$ is a known kernel, thus T is a bounded linear operator with respect to the usual L^2 topology induced by

$$(x, y) = \int_0^1 x(t)y(t)dt, \quad (2.18)$$

$$\|x\| = (x, x)^{1/2}, \quad x, y \in L^2[0, 1]. \quad (2.19)$$

The adjoint operator T^* of T is characterized by

$$(T^*y)(t) = \int_0^1 k(s, t)y(s)ds, \quad t \in [0, 1]. \quad (2.20)$$

Above all, T is a compact operator (see [2], [8] and [24]). Thus [2] there exist adjoint orthogonal functions $\{\phi_i(s)\}$, $\{\psi_i(t)\}$, and real scalars $\lambda_i \rightarrow 0$ such that

$$\int_0^1 k(s, t)\phi_i(s)ds = \lambda_i\psi_i(t), \quad t \in [0, 1] \quad (2.21)$$

and

$$\int_0^1 k(s, t)\psi_i(t)dt = \lambda_i\phi_i(s), \quad s \in [0, 1]. \quad (2.22)$$

Thus, if the right-hand side in (2.17) has the expansion

$$b(s) = \sum_{i=1}^{\infty} \beta_i \psi_i(s), \quad (2.23)$$

then the solution is given by

$$x(t) = \sum_{i=1}^{\infty} (\beta_i / \lambda_i) \phi_i(t); \quad (2.24)$$

however $x(t) \in L^2[0, 1]$ if only if

$$\sum_{i=1}^{\infty} (\beta_i / \lambda_i)^2 < \infty. \quad (2.25)$$

Equation (2.25) is called the *Picard condition*: it determines whether a square-integrable solution exists to the equation (2.17) with right-hand side $b(s)$.

If $b(s)$ of (2.23) does not belong to $R(T)$, we may still approximate $b(s)$ by truncating the expansion (2.23) to a sum of k terms,

$$b_k(s) = \sum_{j=1}^k \beta_j \psi_j(s); \quad (2.26)$$

each of these b_k clearly fulfils the Picard condition, and we have

$$T^{-1}b_k = \sum_{j=1}^k (\beta_j / \lambda_j) \phi_j. \quad (2.27)$$

We conclude that as $k \rightarrow \infty$ we have $b_k \rightarrow b$, but

$$\|T^{-1}b_k\| \rightarrow \infty, k \rightarrow \infty. \quad (2.28)$$

It is for this lack of stability that integral equations of the first kind often produce ill-posed inverse problems.

Unfortunately, in practical situations we will typically only have access to an approximate right-hand side b in equation (2.17), which is contaminated with measurement errors, approximation errors, and rounding errors:

$$b = b^{exact} + \epsilon, \quad b^{exact} \in R(T). \quad (2.29)$$

Here, b^{exact} denotes the unknown, exact right-hand side and ϵ denotes the perturbation. Ideally it is $x^{exact} = T^{-1}b^{exact}$ that we want to compute. Of course, we cannot expect ϵ to

satisfy the Picard condition, hence $b \notin R(T)$. As illustrated above, any naive approach which tries to compute $T^{-1}b$ instead of $T^{-1}b^{exact}$ will usually diverge or return a useless result with extremely large norm, no matter how small the norm of the perturbation ϵ is.

Furthermore, even if we know the exact values of $b(s)$ at some finite set of points s_1, \dots, s_N , from the expansion

$$b(s) = \sum_{i=1}^{\infty} \beta_i \psi_i(s), \quad (2.30)$$

we still can not determine $b(s)$ uniquely. In fact, usually there are infinitely many functions which have same values as $b(s)$ at the points s_1, \dots, s_N . This is the fundamental nonuniqueness for our inverse problems since our data are always finite.

2.1.2 Generalized Solutions and Regularization of Ill-posed Problems

As we have seen, inverse problems are usually ill-posed problems and may have no solution in the traditional sense, so we should introduce the definition of *Generalized Solutions* for inverse problems [6].

Suppose that H_1 and H_2 are Hilbert spaces over the same field of scalars. We consider the fundamental problem of solving for x in H_1 a general linear equation of the type

$$Tx = b \quad (2.31)$$

where $b \in H_2$ and T is a bounded linear operator from H_1 to H_2 . The most prevalent example of an equation of type (2.31) is the one which obtains when $H_1 = R^n$, $H_2 = R^m$ and T is an m by n matrix. If $H_1 = H_2 = L^2[0, 1]$, then the integral operator defined in (2.17) provides another example. In both these examples problem (2.31) may be ill-posed.

For the operator T in equation (2.31), the range $R(T)$ and null space $N(T)$ are defined as

$$R(T) = \{y \in H_2: y = Tx \text{ for some } x \in H_1\}, \quad (2.32)$$

$$N(T) = \{x \in H_1: Tx = 0 \in H_2\}. \quad (2.33)$$

If we let (x_1, x_2) and $\langle y_1, y_2 \rangle$ denote the inner products in the Hilbert spaces H_1 and H_2 respectively, then the adjoint operator T^* of T is a linear operator from H_2 to H_1 characterized by the following relation:

$$(T^*y, x) = \langle y, Tx \rangle, \quad \forall x \in H_1, \forall y \in H_2. \quad (2.34)$$

If the operator T is invertible then equation (2.31) has the unique solution $x = T^{-1}b$, and we have a well-posed problem since T^{-1} is also a bounded linear operator. But in general such a linear equation may have no solution at all when $b \notin R(T)$, and have more than one solution when $N(T) \neq \{0\}$.

When equation (2.31) has no solution in the traditional sense ($b \notin R(T)$), we may assign a “best possible” solution to the problem. In fact, if we assume $R(T)$ is closed and let P be the projection operator of H_2 onto $R(T)$, then Pb is the vector in $R(T)$ which is closest to b . It is possible to consider any $u \in H_1$ such that $Tu = Pb$ as a generalized solution of equation (2.31). This solution is often called the *least squares solution* of the equation $Tx = b$.

Usually there is more than one least squares solution for equation (2.31). To get a uniquely determined least squares solution $u \in H_1$, let us examine the set of least squares solutions. Groetsch [6] shows the set of least squares solutions of equation (2.31) is a closed convex set when $R(T)$ is closed, so there is a unique vector x^* with minimum norm in the set, that is,

$$\|x^*\| = \inf\{\|u\|: Tu = Pb\} = \inf\{\|u\|: \|Tu - b\| \leq \|Tx - b\|, x \in H_1\}. \quad (2.35)$$

Actually x^* is the solution to the following optimization problem

$$\begin{aligned} &\text{Minimize } \phi_x := \|x\|^2 \\ &\text{Subject to } \phi_d := \|Tx - Pb\|^2 = 0 \end{aligned} \quad (2.36)$$

Usually we call x^* the *Generalized solution* of equation (2.31). Sometimes it is also called “best approximate solution” of equation (2.31). It is designed to satisfy the right-hand side as closely as possible.

Now let us consider a finite-dimensional inverse problem with noisy data. Assume

$$Ax = d, \quad (2.37)$$

where A is an $N \times M$ matrix and d is a vector of N observed data with measurement errors. Usually $N \ll M$, and A is an ill-conditioned matrix when $N = M$.

Since the observed data $d = (d_1 \dots d_N)^T$ are contaminated with measurement errors, $d \notin R(A)$ and thus there is no solution to equation (2.37). Furthermore, even if there is a solution x to equation (2.37) which satisfies the observed data d , this solution is usually unstable since the matrix A is ill-conditioned.

If we write the observed data as

$$d = d^{exact} + \epsilon, \quad (2.38)$$

with d^{exact} being exact data from the true model and ϵ being the measurement errors, then equation (2.37) becomes

$$Ax = d^{exact} + \epsilon. \quad (2.39)$$

According to the analysis of generalized solutions, we should find the best approximate solution to linear equation system

$$Ax = d^{exact}. \quad (2.40)$$

In other words, we should solve the following optimization problem

$$\begin{aligned} &\text{Minimize } \phi_x := \|x\|^2 \\ &\text{Subject to } \phi_d := \|Ax - d^{exact}\|^2 = 0. \end{aligned} \quad (2.41)$$

However, the vector d^{exact} is not available in practical computing, so instead of calculating the generalized solution to equation (2.40) we look instead for the generalized solution to the equation

$$\|Ax - d\|^2 = \phi_d^*, \quad (2.42)$$

where ϕ_d^* is our best estimate of the expected error norm $E(\|\epsilon\|^2)$. This is to solve optimization equation

$$\begin{aligned} &\text{Minimize } \phi_x := \|x\|^2 \\ &\text{Subject to } \phi_d := \|Ax - d\|^2 = \phi_d^* \end{aligned} \quad (2.43)$$

Using the Lagrange multiplier method, problem (2.43) can be changed into the optimization problem

$$\begin{aligned} &\text{Minimize } \phi = \phi_x + \alpha(\phi_d - \phi_d^*) \\ &= \|x\|^2 + \alpha(\|Ax - d\|^2 - \phi_d^*). \end{aligned} \quad (2.44)$$

Now we show that this approach is consistent with widely accepted Tikhonov regularization method.

According to Tikhonov and Arsenin [22] a given algorithm is called a *regularization method* for (2.37) if—in the presence of perturbations ϵ as in (2.39)—the algorithm determines approximations $x(\epsilon)$ with the property that for some solution x^{exact} of (2.40),

$$x(\epsilon) \rightarrow x^{exact}, \quad \epsilon \rightarrow 0. \quad (2.45)$$

So, to find a generalized solution to equation (2.37) is also a regularization method.

Furthermore, in Tikhonov's method the approximate solution x to (2.37) is defined as the unique minimizer of the quadratic functional

$$\phi = \|d - Ax\|^2 + \alpha \|Bx\|^2 \quad (2.46)$$

in a finite space R^M . Here B is the discrete version of a *regularization operator* L which defines a (semi-)norm $\|L \cdot\|$. It is used to measure the “size” of the approximations. α is a *regularization parameter* which belongs to a suitable real parameter set. If we assume T is injective, then the minimizer x_α is unique.

Obviously, if $L = I$ in (2.46), optimization problem (2.46) is same with optimization problem (2.44). Therefore, the approach of generalized solution is consistent with Tikhonov regularization. In the future we use the phrase “regularization method” for both Tikhonov regularization and for any other regularization methods.

2.1.3 Construction of Regularization

As we have seen, to solve a general ill-posed problem

$$Tx = d, \quad (2.47)$$

we usually use some kind of regularization method. In other word, instead of solving equation (2.47) we usually solve the following optimization problem

$$\text{Minimize } \phi = \|d - Tx\|^2 + \alpha \|Lx\|^2. \quad (2.48)$$

Equivalently, the solution x_α to (2.48) may be introduced as the solution to the regularized normal equations

$$(T^*T + \alpha L^*L)x_\alpha = T^*b. \quad (2.49)$$

Note that L^*L is positive (semi-)definite so that the spectrum of the operator on the left-hand side of (2.49) is shifted in the positive direction; hence the solution of the regularized normal equations should be less susceptible to perturbations in the data d . In fact, L will typically be chosen such that Morozov’s complementation condition [14] is fulfilled, i.e.

$$\|Tx\| + \|Lx\| \geq \gamma \|x\| \quad (2.50)$$

for some $\gamma > 0$ and all x in the domain of L . In this case, $T^*T + \alpha L^*L$ is continuously invertible, hence (2.48) has a unique solution x_α depending continuously on d .

Instead of (2.48) one might consider the slightly more general functional

$$\text{Minimize } \|d - Tx\| + \alpha \|L(x - x_{guess})\|^2 \quad (2.51)$$

if one has available a good guess for the exact solution x^{exact} . However, (2.51) can easily be reduced to (2.48) by replacing d by $d - Tx_{guess}$ in (2.48).

The basic idea behind the regularization in (2.48) is to search for some x_α , that provides at the same time a small residual $\|d - Tx_\alpha\|$ and a moderate value of the penalty function $\|Lx_\alpha\|$. The keys to succeed are a proper regularization operator L and a suitable regularization parameter α .

Choosing a proper regularization parameter α in problem (2.48) is very important. If the parameter is chosen too small, equation (2.49) is too close to the original ill-posed problem and we must expect instabilities. In this case what we can expect to get is the best approximate solution of $Tx = d$. On the other hand, if the parameter is chosen too large, the problem we solve has very little similarity to the original equations since too much information from the data is lost. In this case, the solution we can expect to get is some $x \in N(L)$ or simply 0.

Finding the optimal regularization parameter α is important in solving ill-posed problems, but there are no general rule to choose it. One basic principle is to choose the regularization parameter α such that the norm of the misfit, $\|d - Tx_\alpha\|$, equals the norm of the error term [14], i.e.

$$\phi_d \equiv \|d - Tx_\alpha\|^2 = \|\epsilon\|^2. \quad (2.52)$$

For our examples in the following Chapters, $\|\epsilon\|^2$ has χ^2 distribution with expected value N , so we require the misfit ϕ_d to equal its expected value N . Obviously this rule

depends on the accuracy of the estimate for $\| \epsilon \|$. If no good estimate is available, it may be better to use an ϵ -free parameter choice strategy.

The role of L should not be underestimated. Numerical computing has shown that we could get completely different results with different choices of L . Varah [25] has shown that a wrong choice of L can lead to completely erroneous results.

Although a general rule for choosing L is unknown, it is most common to choose either the identity I or a linear differential operator $L \neq I$. This choice of L usually implies the validity of the complementation condition (2.50).

To choose a sensible L from appropriate *a priori* knowledge we use an example from the following chapters. Suppose we need to determine the unknown function $m = m(x, z)$ using

$$Tm = d \quad (2.53)$$

in a Cartesian coordinate system consisting of x -axis and z -axis. We can choose L by combining differential operators and the identity such that for suitable weight functions $w_s(x, z)$, $w_x(x, z)$ and $w_z(x, z)$

$$\begin{aligned} \| L(m - m_0) \|^2 = & \alpha_s \int_{area} w_s(x, z)(m - m_0)^2 dx dz \\ & + \alpha_x \int_{area} w_x(x, z) \left(\frac{\partial(m - m_0)}{\partial x} \right)^2 dx dz \\ & + \alpha_z \int_{area} w_z(x, z) \left(\frac{\partial(m - m_0)}{\partial z} \right)^2 dx dz. \end{aligned} \quad (2.54)$$

If we choose $\alpha_x = \alpha_z = 0$ and w_s as a constant, then $L = I$ and we have a regularization in *standard form* opposed to a regularization in *general form* if $L \neq I$. This choice is simple and satisfies the complementation condition, but it is useful only when we have no idea about what the model should be. This choice leads to a reconstruction called the “smallest model” method since it only concerns the size of the model.

If we choose $\alpha_x \neq 0$ and/or $\alpha_z \neq 0$, we are asking the model m to be smooth in some

sense. The reason for choosing differential operators $L \neq I$ is that noise components in the data usually lead to rough oscillations in the reconstructed model m which will provoke large L -(semi) norms, $\|Lm\|$ when $L \neq I$ while keeping standard norm $\|m\|$ to be moderate. This choice is reasonable for geophysical inverse problems since our true models are usually at least piecewise smooth, and not just integrable.

If we choose L as a differential operator with $\alpha_s = 0$, we can expect a very smooth reconstructed model. This is also called the “flattest model”.

2.2 Solving Inverse Problems

According to Anger [1], to solve an inverse problem we have to study the following issues:

1. The special process—both experimentally and theoretically,
2. Mathematical models of the process,
3. The direct problem—both theoretical and numerical,
4. The information content of the inverse problem,
5. Algorithms for the numerical solution of the inverse problem.

In this section we will work only on point 5 and leave the other four points to be studied in Chapter 3 for two specific practical problems.

2.2.1 Discretization of Inverse Problems

To compute an approximation to x^{exact} , we somehow have to discretize the original problem. One reason is that for inverse problems we usually know only finitely many contaminated data instead of a continuous function. The other reason is that it is difficult to find analytic solutions to practical inverse problems. Although there are various ways to

reduce the integral equation (2.17) to a discrete equation, we do not want to elaborate on the different advantages of the respective methods here since the discretization methods are usually determined by the problems.

No matter which method we use to discretize an integral equation, one problem we have to face is the choice of the proper mesh. If the discretization is too coarse, then the finite dimensional equation will be fairly well conditioned but its solution will suffer from large discretization errors; this is like using regularization with too large a regularization parameter α . If, on the other hand, the discretization level is too fine, then the influence of the small singular values of T will be too significant, and the solution of the discrete problem will resemble Tikhonov approximation with too small a regularization parameter α . Somewhere in between there is an optimal level of discretization with the resulting approximation being competitive to the one obtained from optimal regularization of the continuous problem.

For practical problems the optimal discretization is hardly ever known in advance. Even if we chose the optimal discretization by chance, it would still remain a good idea for the purpose of numerical stability to regularize the finite-dimensional problem so that its condition number is improved. In practice, we usually select a discretization so fine that the corresponding discretized inverse problem is ill-posed. This allows us to use prior knowledge and adjust the misfit by choosing the proper regularization parameter α .

There are two approaches to produce the system of linear equations: to discretize the integral equation (2.17) first, or to regularize the integral equation (2.17) first.

If we discretize the integral equation first, we obtain a linear system of equations

$$Ax = d, \tag{2.55}$$

where A is the coefficient matrix and d is the vector of observed data. Usually the

system of linear equations (2.55) is ill-posed since A will not be a square matrix when we discretize the integral equation with too fine a mesh. Even if A is a square matrix for some discretization methods, it may be singular or at least numerically singular. Applying regularization in standard form to the linear system of equations (2.55) amounts to solving the following regularized normal equation:

$$(A^T A + \alpha I)x = A^T d. \quad (2.56)$$

On the other hand, if we regularize the integral equation (2.17) first, the resulting discrete problem will involve in minimizing a quadratic functional of the form

$$\| Ax - d \|^2 + \alpha \| Bx \|^2, \quad (2.57)$$

where B is the discrete version of a differential operator L in the functional (2.46). Equivalently, its regularized normal equation is

$$(A^T A + \alpha B^T B)x = A^T d, \quad (2.58)$$

where B is the discrete form of the differential operator L .

In the examples in the following chapters, we regularize the integral equation first. To discretize the regularized problem we divide the domain of interest into M elements (e.g., intervals in 1-D, rectangles in 2-D) and assume the model x is constant in each element. If the number of observed data is N , then A is an $N \times M$ matrix. In our cases, $N \ll M$, but both numbers are large: we have $M = 560$ and $N = 310$.

After discretizing the inverse problem, we have to find an efficient algorithm to solve it. This process is especially important for practical problems since we often have several thousands of unknown variables. Here we review several direct methods for discrete ill-posed problems like equations (2.55), (2.56), (2.58).

2.2.2 Direct Methods

The Direct methods are simple and widely used, but usually they are not efficient for large scale inverse problems.

1. SVD with Truncation. Assume A is an $N \times M$ ($N \leq M$) matrix in equation $Ax = d$. To solve this system of linear equations, we suppose A has the following Singular Value Decomposition (SVD):

$$A = UDV^T, \quad (2.59)$$

where U is an $N \times N$ orthogonal matrix, V is an $M \times N$ matrix satisfying $V^T V = I^N$, and $D = \text{diag}(\sigma_1, \dots, \sigma_N)$. Here $\sigma_1 \geq \sigma_2 \dots \geq \sigma_N \geq 0$ are the singular values of matrix A .

Now for the linear system $Ax = d$, if $N = M$ and $\sigma_N > 0$, then the unique solution is

$$x = VD^{-1}U^T d. \quad (2.60)$$

Furthermore, if we assume

$$U = [u_1|u_2|\dots|u_N], \quad (2.61)$$

$$V = [v_1|v_2|\dots|v_N], \quad (2.62)$$

$$d = \sum_{i=1}^N \beta_i u_i = U\beta, \quad (2.63)$$

then

$$x = \sum_{i=1}^N \frac{\beta_i}{\sigma_i} v_i, \quad (2.64)$$

$$\|x\|^2 = \sum_{i=1}^N (\beta_i / \sigma_i)^2. \quad (2.65)$$

This is the solution to equation (2.55) when it is a well-posed problem.

If $N < M$ and $\sigma_N > 0$, then we can define an approximate solution to equation (2.55)

as

$$x^* = VD^{-1}U^T d \quad (2.66)$$

with

$$x^* = \sum_{i=1}^N \frac{\beta_i}{\sigma_i} v_i, \quad (2.67)$$

$$\|x^*\|^2 = \sum_{i=1}^N (\beta_i/\sigma_i)^2. \quad (2.68)$$

This solution to equation (2.55) is its generalized solution or best approximate solution as we introduced in formula (2.35). Sometimes the requirement that $\|x\|^2$ in (2.68) is moderate is called *discrete Picard condition*.

If $\sigma_i = 0$ or $\sigma_i \approx 0$ when $i < k \leq N$, the matrix D is either exactly or numerically singular. In this case, we introduce the truncated inverse of D with elements defined as

$$\frac{1}{\sigma_i} = \begin{cases} \frac{1}{\sigma_i} & i \geq k \\ 0 & i < k. \end{cases} \quad (2.69)$$

With D^{-1} defined in this way, we can obtain the truncated SVD solution to equation (2.55):

$$x^t = VD^{-1}U^T d, \quad (2.70)$$

with

$$x^t = \sum_{i=1}^k \frac{\beta_i}{\sigma_i} v_i, \quad (2.71)$$

$$\|x^t\|^2 = \sum_{i=1}^k (\beta_i/\sigma_i)^2, \quad (2.72)$$

$$\|d - Ax^t\|^2 = \|d - UDV^TVD^{-1}U^Td\|^2 = \sum_{i=k+1}^N \beta_i^2. \quad (2.73)$$

Theoretically we can choose proper k to get a reasonable solution for which the misfit $\|d - Ax^t\|^2$, is small and the the solution norm $\|x^t\|$ is not too large (see [24]). However, for many practical problems we can not expect a particular gap in the spectrum to separate large singular values σ_k from those that may be judged numerically zero (see [9]). To choose a proper k is completely subjective or depends on trial and error.

2. SVD with Regularization. Instead of truncating the small singular values of A as in (2.69), we can use a regularization method to truncate the small singular values automatically. This is to use SVD method to solve the regularized normal equation (2.56), that is,

$$(A^T A + \alpha I)x = A^T d. \quad (2.74)$$

As above, assume A has the singular value decomposition, $A = UDV^T$ with U , V and d in the form of (2.61), (2.62) and (2.63) respectively. Then we get a unique solution to (2.74),

$$x^\alpha = V(D^2 + \alpha I)^{-2}DU^T d, \quad (2.75)$$

with

$$x^\alpha = \sum_{i=1}^N \frac{\beta_i \sigma_i}{\sigma_i^2 + \alpha} v_i, \quad (2.76)$$

$$\|x^\alpha\|^2 = \sum_{i=1}^N \left(\frac{\beta_i \sigma_i}{\sigma_i^2 + \alpha} \right)^2, \quad (2.77)$$

$$\|d - Ax^\alpha\|^2 = \|d - UD(D^2 + \alpha I)^{-2}DU^T d\|^2 = \sum_{i=1}^N \frac{\alpha}{\alpha + \sigma_i^2} \beta_i^2. \quad (2.78)$$

Here we can see that as the regularization parameter α increases, the norm of the solution $\|x^\alpha\|$ decreases and the misfit $\|d - Ax^\alpha\|$ increases. This is consistent with our discussion above about the regularization parameter.

2.2.3 Iterative Methods

As we have seen, direct methods give us an inside look at finite-dimensional ill-posed problems. However, for large-scale applications a striking disadvantage of direct methods becomes obvious: the amount of work required to compute the regularized approximation is of the order $O(N^3)$ where N is the number of unknown variables.

Furthermore, 2-D and 3-D problems usually give rise to large structures or sparse matrices A and/or B . The direct methods for regularization become impractical because

they destroy the structure or sparsity of the matrices. Consequently, it is natural for us to turn to iterative methods that only require matrix-vector multiplications which can be implemented very efficiently and do not alter the matrices A and B .

Here we will show how to solve linear inverse problems in the general form (2.58) using iterative methods.

To find a minimum to the quadratic functional in (2.57) is to find a minimum to the following quadratic functional,

$$\phi = \phi_d + \alpha \phi_m = \|Ax - d\|^2 + \alpha \|Bx\|^2 \quad (2.79)$$

where $\phi_d = \|Ax - d\|^2$ is the misfit and $\phi_m = \|Bx\|^2$ is the size of the model. If we choose $\alpha = \mu^{-1}$, the optimization problem (2.79) becomes

$$\text{Minimize } \phi = \mu \phi_d + \phi_m = \mu \|Ax - d\|^2 + \|Bx\|^2. \quad (2.80)$$

Sometimes μ is called a Lagrange multiplier since (2.80) can be obtained from following optimization problem

$$\text{Minimize } \phi_m = \|Bx\|^2 \quad (2.81)$$

$$\text{Subject to } \phi_d \equiv \|Ax - d\|^2 = \phi_d^*. \quad (2.82)$$

1. Conjugate Gradient Method. Since $\phi(x)$ in (2.79) is a strictly convex functional, its unique minimum point is characterized by the regularized normal equation

$$Gx = b \quad (2.83)$$

where $G = A^T A + \alpha B^T B$ and $b = A^T d$. G is usually a positive definite symmetric matrix, so it is natural for us to turn to the famous Conjugate Gradient Method [5].

In contrast to the steepest descent method, which chooses the gradient of the objective functional $\phi(x)$ as its search direction in each iteration, the conjugate gradient method

chooses the search direction $p_k \neq 0$ for iteration k to satisfy

$$p_k^T A p_i = 0, \quad i = 1, \dots, k-1. \quad (2.84)$$

Then the vector $x_k = x_{k-1} + \alpha_k p_k$ is chosen to minimize $\phi(x)$ over $\text{span}\{p_1, \dots, p_k\}$. Since p_k is A -conjugate to p_1, \dots, p_{k-1} , the finite termination of the conjugate gradient method is guaranteed. In fact, when A is a positive definite symmetric matrix, any set of A -conjugate vectors is linearly independent.

Theoretically this procedure solves (2.83) in finitely many steps, but rounding errors lead to a loss of orthogonality among the residuals in the conjugate gradient method. As a consequence, the property of finite termination does not apply in practice. Furthermore, the loss of orthogonality may cause slow convergence in ill-conditioned problems. Pre-conditioning A may improve the converge rate, but we still have to exert a huge computing effort for each regularization parameter α until we find the proper one. This is the main reason we turn to the subspace iterative method.

2. Subspace Iterative Method. When we use direct methods to solve the normal equation (2.83), we search for the solution x in R^N and get its solution in one big step. However, when we use the conjugate gradient method to solve equation (2.83), in each iteration we search for an improved solution in only one direction. The subspace method of Oldenburg and Li [15] can be viewed as a compromise, in which we search for an improved solution in each iteration in a larger subspace of R^N .

Oldenburg and Li [15] give a detailed discussion of the subspace linear inverse method, so it suffices to review its basic ideas and steps here.

Solving the optimization problem (2.80)

$$\text{Minimize } \phi(x) = \mu \phi_d + \phi_m = \mu \|Ax - d\|^2 + \|Bx\|^2 \quad (2.85)$$

is equivalent to solving the regularized normal equation

$$(\mu A^T A + B^T B)x = \mu A^T d. \quad (2.86)$$

Suppose we know an approximate solution $x^{(k)}$ to (2.85) and want to compute δx such that $x^{(k+1)} = x^{(k)} + \delta x$ is a better approximate solution of (2.86). We can substitute $x^{(k+1)}$ into (2.85) and take derivatives to get

$$(\mu A^T A + B^T B)\delta x = \mu A^T d - (\mu A^T A + B^T B)x^{(k)}. \quad (2.87)$$

In contrast to direct methods, which solve (2.87) for δx in the whole space R^N , we solve for δx in the subspace of R^N spanned by some preassigned vectors $\{v_1, \dots, v_q\}$, which will be specified later.

Let $V = [v_1 | v_2 | \dots | v_q]$, $\beta = (\beta_1, \dots, \beta_q)^T$ and $\delta x = \sum_{i=1}^q v_i \beta_i = V\beta$. Then just as in (2.86) we can get

$$V^T(\mu A^T A + B^T B)V\beta = \mu V^T A^T d. \quad (2.88)$$

Here $V^T(\mu A^T A + B^T B)V$ is an $q \times q$ matrix and $q \ll N$, so it is easy to solve (2.88) using a direct method of the sort described above.

To choose proper basis vectors v_i is an important consideration in the subspace method. Usually the vectors v_i are chosen as the steepest descent directions of objective functionals or the steepest descent direction of specific components of the objective functionals [15].

Here are the basic steps of the subspace method. We will describe each step in detail in our numerical examples.

1. Set up the initial model $x^{(0)}$;
2. Compute the function values $\phi_m(x^{(k)})$, $\phi_d(x^{(k)})$ and the spanning vectors $V = \{v_i\}$, $i = 1, \dots, q$ from $x^{(k)}$, $k \geq 0$;

3. Linear search for proper μ such that ϕ_d decreases by a user-specified factor while ϕ_m does not increase too much.
4. Compute $\delta x = \sum_{i=1}^q \beta_i v_i$ by solving the linear system $\nabla_\beta \phi = 0$;
5. Update $x^{(k+1)} = x^{(k)} + \delta x$;
6. If $x^{(k+1)}$ satisfies the misfit requirement and $\phi_m(x^{(k+1)})$ differs very little from $\phi_m(x^{(k)})$, stop the iteration and print out the results; otherwise, return to 2.

The subspace algorithm is efficient, flexible and robust [15]. It chooses the proper regularization parameter μ automatically by controlling the misfit ϕ_d and searches for a perturbation from a subspace of moderate dimension exceeding one, so it is much more efficient than the conjugate gradient method. With proper transforms, it is possible for users to impose positivity constraints in practical problems. Therefore, we will use this algorithm to solve all our examples in the following chapters.

2.3 Linearizing Nonlinear Inverse Problems

In practical applications, many inverse problems are nonlinear. Although some nonlinear inverse problems can be transformed into linear inverse problems exactly, in many cases we have to make a linear approximation.

Although we only use an initial linearization in a numerical example, we review this Newton's method for general linearization in each step since their basic ideas are common.

Suppose $x \in R^M$ is a finite-dimensional unknown which produces data d_i under the vector-valued nonlinear transform F , that is,

$$F_i(x) = d_i, \quad i = 1, \dots, N. \quad (2.89)$$

The inverse problem is to find x , given the data $d = (d_1, \dots, d_N)^T$ and the nonlinear transforms $F_i, i = 1, \dots, N$. Usually this kind of inverse problem is ill-posed.

For problem (2.89), we usually seek a generalized solution since it may have no solution in the strict sense. To find a generalized solution, we have to solve the following optimal problem

$$\text{Minimize } \phi_m(x) \equiv \| Bx \|^2 \quad (2.90)$$

$$\text{Subject to } \phi_d(x) \equiv \| d - F(x) \|^2 = \phi_d^* \quad (2.91)$$

where B is an $M \times M$ matrix.

To solve equation (2.91), we use linear approximation and iteration. If $x^{(k)}$ is an approximate solution to the optimization problem (2.90)—(2.91), then for x near $x^{(k)}$, we have

$$F(x) = F(x^{(k)}) + J(x - x^{(k)}) + O(|x - x^{(k)}|^2). \quad (2.92)$$

Here

$$F(x) = (F_1(x), \dots, F_N(x))^T$$

and $J = (J_{ij})$ is the $N \times M$ sensitivity matrix defined by

$$J_{ij} = \frac{\partial F_i(x^{(k)})}{\partial x_j}, \quad i = 1, \dots, N, j = 1, \dots, M. \quad (2.93)$$

Dropping the quadratic error term in (2.92) and using the resulting approximation in (2.91) produces the following optimization problem:

$$\text{Minimize } \phi_m = \| Bx \|^2, \quad (2.94)$$

$$\text{Subject to } \phi_d = \| d^{(k)} - Jx \|^2 \quad (2.95)$$

where $d^{(k)} = d - F(x^{(k)}) + Jx^{(k)}$. This is a linear inverse problem and we can solve it as outlined above to produce an improved approximation solution $x^{(k+1)}$. If we iterate the above process, recomputing J from (2.93) at each step, we obtain a general Newton's method. For simple problems, however, a single computing of J and just one linear

inversion may provide sufficient accuracy. It is the latter method we have used in our IP inversions.

After we linearize the nonlinear inverse problems, we solve a linear inverse problem in each iteration. However, in some cases we can decrease the linear misfit efficiently, without decreasing the nonlinear misfit (2.91) at all. Therefore, when we solve linearized inverse problems in each iteration, it is necessary to choose the proper regularization parameter such that the nonlinear misfit (2.91) decreases in each iteration.

Another difficulty in linearizing nonlinear inverse problems is the computing of the sensitivity matrix J . If we cannot find an analytical form for J , we have to compute it numerically. This process requires large scale computing, so an efficient method is needed. To find efficient methods of computing J is quite problematic. We refer to [11] for some practical advice on computing sensitivity matrices using Frechet derivatives and adjoint differential operators.

Linearizing nonlinear inverse problems works well when the initial approximation is close to the true model of the original nonlinear problem. If the initial approximation is not close enough and/or the objective is not convex, it is possible for the method to diverge or to converge to a completely incorrect model.

Chapter 3

Two Practical Problems

In this chapter we use the regularization methods of Chapter 2 to solve two kinds of practical inverse problems in Geophysics: the Magnetic prospecting problem and the Induced Polarization (IP) problem.

3.1 Magnetic Methods

Magnetic prospecting is one of the oldest methods of geophysical exploration. It is used to explore for oil, minerals, and even archaeological artifacts. Collecting magnetic field measurements is easy, cheap and simple compared to most geophysical techniques. On the other hand, a precise interpretation of magnetic field data is very difficult. Therefore, it is important to find efficient algorithms to invert magnetic data.

3.1.1 Physical Principles

Before studying the magnetic effects associated with subsurface materials in the Earth we review the elementary physical concepts that are fundamental to magnetic prospecting.

Geomagnetic field. If a steel needle, not previously magnetized, could be hung at its center by a thread so that it were free to orient itself in any direction in space, at most points on the earth's surface it would assume a direction neither horizontal nor in line with the geographical meridian. Its orientation would rather be the direction of the Earth's total magnetic field \vec{H}_e at this point. The magnitude of this field, H_e , the inclination of the needle from the horizontal, I , and the angle it makes with geographic

north, D , completely define the magnetic field (figure 3.4). Here the units of \vec{H}_e are

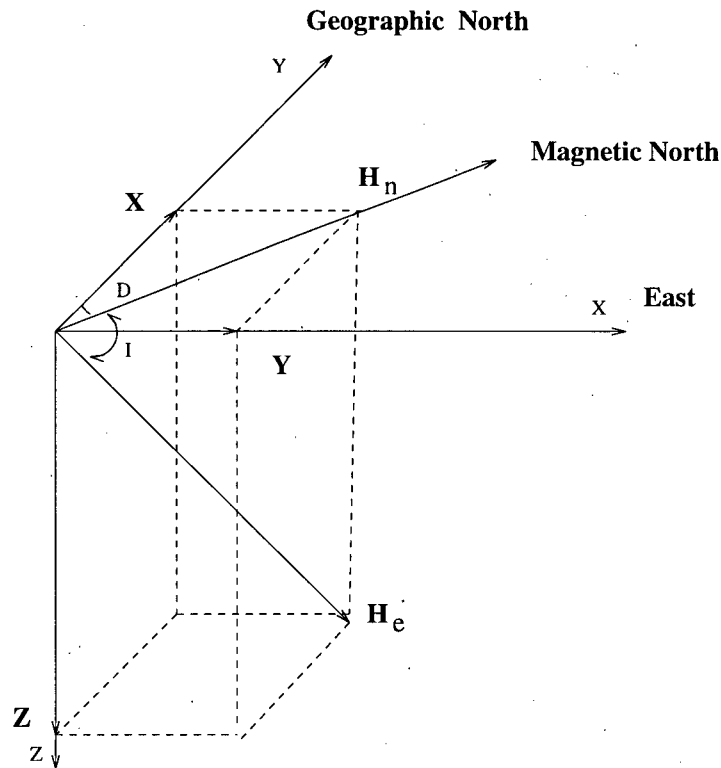


Figure 3.4: Geomagnetic field.

ampere/meter and the magnitude of $\mu_0 \vec{H}_e$ is of the order 0.5×10^{-4} tesla, where $\mu_0 = 4\pi 10^{-7}$ weber/ampere-meter is the permeability of free space.

Susceptibility. In the atoms or molecules of any magnetic material small loopcurrents (including rotating charges) occur, all having themselves directions but a random orientation. Each loopcurrent corresponds to a magnetic dipole. When the magnetic material is placed in a magnetic field, these magnetic dipoles tend to line up in the direction of the field. The magnetization \vec{M} (also called “Polarization”) is defined as the magnetic moment per unit volume of magnetized material at a point. The magnitude of magnetization depends on the strength of the external field through the relation

$$\vec{M} = \kappa \vec{H} \quad (3.96)$$

where κ , the proportionality constant, is called the *susceptibility*. Since \vec{M} and \vec{H} are both measured in *ampere/meter*, susceptibility is dimensionless in the SI system. Susceptibility is the fundamental parameter in magnetic prospecting, since the magnetic response of rocks and minerals is determined by the amount of magnetic material in them and the latter have κ values much larger than the rocks and minerals themselves.

For a vacuum, $\kappa = 0$. *Diamagnetic* substances are weakly magnetic, with negative susceptibilities. *Paramagnetic* materials are weakly magnetic, characterized by small positive susceptibilities. *Ferromagnetic* materials have positive and very high susceptibilities, but rarely occur naturally in the Earth's crust. *Ferrimagnetic* materials have positive and high susceptibilities. Ferrimagnetic materials contained in rocks are the most important in magnetic prospecting.

Magnetic anomaly. The magnetic anomaly is the difference between observed and background earth's magnetic field values. In magnetic prospecting magnetic anomalies are usually caused by the magnetization of ferrimagnetic material under the earth's surface, so they are our measured data for the magnetic inverse problem.

Magnetic flux density. If a current is passed through a coil consisting of several turns of wire, a magnetic flux flows through and around the coil annulus which arises from a magnetizing force \vec{H} . The magnitude of \vec{H} is proportional to the number of turns in the coil and the strength of the current, so that \vec{H} is expressed in *ampere/meter*. The density of the magnetic flux, measured over an area perpendicular to the direction of flow, is known as the *magnetic flux density* \vec{B} . \vec{B} is proportional to \vec{H} ; i.e.

$$\vec{B} = \mu \vec{H}, \quad (3.97)$$

where the constant of proportionality μ

$$\mu = \mu_0(1 + \kappa), \quad (3.98)$$

is known as the magnetic permeability. The units of permeability are *weber/ampere-meter* and in free space $\mu = \mu_0 = 4\pi \times 10^{-7}$ *weber/ampere-meter*. The units of \vec{B} are *tesla* (*weber/meter²*) and sometime \vec{B} is also used to express magnetic field. Although we mainly use \vec{H} to represent the strength of magnetic field, it is easy to transform between \vec{H} and \vec{B} using equation (3.97) once μ is known.

3.1.2 Mathematical Relations

The strength of magnetic field \vec{F} at the origin caused by a magnetic dipole \vec{M} at point P with position \vec{r} satisfies [10], [20]

$$\vec{F}(\vec{r}) = \nabla \times \left(\frac{1}{4\pi} \frac{\vec{M} \times \vec{r}}{|\vec{r}|^3} \right) = \frac{1}{4\pi} \nabla [\vec{M} \cdot \nabla \left(\frac{1}{|\vec{r}|} \right)]. \quad (3.99)$$

Generally, a volume of magnetic material can be considered as an assortment of magnetic dipoles that results from the magnetic moments of individual atoms and dipoles. They are aligned by induction in the presence of a magnetizing field. In any case, we may regard the body as a continuous distribution of dipoles resulting in a magnetization \vec{M} , of magnitude M . The strength of the field caused by the whole body measured at a point \vec{r}_0 outside the body is

$$\vec{F}(\vec{r}_0) = \frac{1}{4\pi} \int_V \nabla [\vec{M}(\vec{r}) \cdot \nabla \left(\frac{1}{|\vec{r} - \vec{r}_0|} \right)] dV, \quad (3.100)$$

where $\vec{r}_0 = (x_0, y_0, z_0)$ and $|\vec{r} - \vec{r}_0| = [(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2]^{1/2}$.

If \vec{M} is a constant vector with direction $\alpha = (l, m, n)$ and magnitude M then the operator

$$\vec{M} \cdot \nabla = M \frac{\partial}{\partial \alpha} = M \left(l \frac{\partial}{\partial x} + m \frac{\partial}{\partial y} + n \frac{\partial}{\partial z} \right). \quad (3.101)$$

Thus, we have:

$$\vec{F}(\vec{r}_0) = \frac{1}{4\pi} M \nabla \frac{\partial}{\partial \alpha} \int_V \left(\frac{dV}{|\vec{r}_0 - \vec{r}|} \right) \quad (3.102)$$

and we can get the projection of \vec{F} in any direction.

3.1.3 Forward Problem

The data for magnetic prospecting are the magnetic field anomalies caused by underground bodies with high magnetic susceptibility. Magnetic anomalies caused by intrusions, flows, or iron-rich sedimentary horizons are common features in regions favorable for mineral exploration, and there is frequently a difference between the magnetic mineral content of such features and that of the host rocks. Such features may often be simulated by a two dimensional dipping dike.

Here we set up a 2-dimensional model to produce magnetic data. We assume a dike with dip ξ and strike β ; see figure 3.5. If we take the y -axis along the strike direction, the z -axis downward and the x -axis perpendicular to the strike direction, then the magnetic anomalies are independent of y . For this two-dimensional model we can assume the susceptibility is $\kappa = \kappa(x, z)$ and the magnetic polarization is in the direction of the Earth's field \vec{H}_e , that is, $\vec{M} = \kappa(x, z)\vec{H}_e$. Then according to (3.100) the corresponding magnetic field anomaly $\vec{d}(\vec{r}_0)$ satisfies

$$\vec{d}(\vec{r}_0) = \frac{1}{4\pi} \nabla \left[\int_V (\kappa(x, z)\vec{H}_e) \cdot \nabla \left(\frac{1}{|\vec{r} - \vec{r}_0|} \right) dV \right]. \quad (3.103)$$

Here $\vec{r}_0 = (x_0, 0, z_0)^T$, $\vec{r} = (x, y, z)^T$ and $|\vec{r} - \vec{r}_0| = [(x - x_0)^2 + y^2 + (z - z_0)^2]^{1/2}$. Since \vec{H}_e is a constant vector, this is an integral equation of first kind when data, $\vec{d}(\vec{r}_0)$ are known. When we integrate the right-hand side with respect to y from $-\infty$ to ∞ we reduce (3.103) to a completely 2D problem.

To get the data about our two-dimensional model, we partition the area of interest into a series of rectangles. We assume that the susceptibility in each rectangle is constant, and use (3.102) to compute the magnetic field caused by each rectangle.

For the geometry illustrated in figure 3.5, we have the following relations [20]:

$$r_1^2 = d^2 + (x + \cot \xi)^2,$$

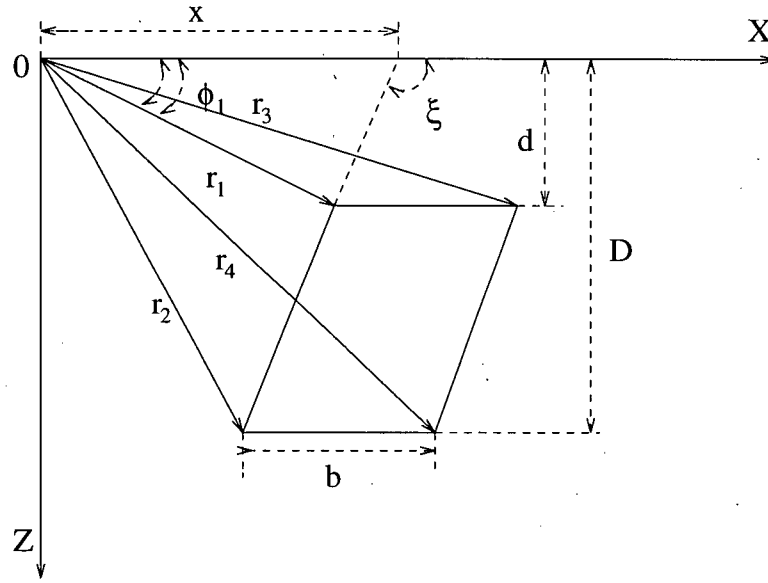


Figure 3.5: Rectangle geometry

$$r_2^2 = D^2 + (x + D \cot \xi)^2,$$

$$r_3^2 = d^2 + (x + b + d \cot \xi)^2,$$

$$r_4^2 = D^2 + (x + b + D \cot \xi)^2.$$

If we assume the angles between x -axis and \vec{r}_i are ϕ_i , then

$$\phi_1 = \tan^{-1}\{d/(x + d \cot \xi)\}, \quad (3.104)$$

and we can compute ϕ_2 , ϕ_3 and ϕ_4 similarly.

Usually we compute the projections of the magnetic anomaly field in z -direction or/and the Earth's magnetic field direction. Assuming that $\xi=90^\circ$, we get the projections of \vec{F} in \vec{H}_e direction and z -direction from (3.103):

$$F_e = \frac{\kappa |\vec{H}_e|}{2\pi} \{ \sin(2I) \sin \beta \times \ln(r_2 r_3 / r_4 r_1) \\ + (\phi_1 - \phi_2 - \phi_3 + \phi_4) (\cos^2 I \sin^2 \beta - \sin^2 I) \}, \quad (3.105)$$

$$F_z = \frac{\kappa |\vec{H}_e|}{2\pi} \{ \cos I \sin \beta \times \ln(r_2 r_3 / r_4 r_1) - (\phi_1 - \phi_2 - \phi_3 + \phi_4) \sin I \}. \quad (3.106)$$

At each observation point for the magnetic problem, the measured anomaly is a vector whose two components are the sums of F_e and F_z over all rectangles.

3.1.4 Inverse Problem

To solve the magnetic inverse problem, we must solve an integral equation of the first kind. In contrast to the standard integral equation of the first kind, we only know $\vec{d}(\vec{r}_0)$ for a constant z_0 , that is, along a line, instead of knowing $\vec{d}(\vec{r}_0)$ for all $(x_0, 0, z_0)$ in the domain of definition. In fact, in practical experiments, we only measure the magnetic anomaly on the surface. For two-dimensional problems we just measure the anomaly along a line on the surface. From theoretical analysis, we know that the magnetic field satisfies Laplace's equation above the surface, so the values of $\vec{d}(\vec{r}_0)$ along lines with different z_0 are not independent. Therefore, it is not helpful to measure $\vec{d}(\vec{r}_0)$ along different lines with different z_0 above the surface. That is, we are required to solve for $\kappa(x, z)$ in the system of equations

$$\vec{d}(x_i, z_0) = \frac{1}{4\pi} \nabla \left[\int_V (\kappa(x, z) \vec{H}_e) \cdot \nabla \left(\frac{1}{|\vec{r}(x, y, z) - \vec{r}_0|} \right) dV \right], \quad (3.107)$$

where $d(x_i, z_0)$ are the data which are measured along a line on the surface, $(x_i, z_0), i = 1, \dots, N$.

To find a numerical solution to equation (3.107), we have to discretize (3.107) first. We divide the area of interest into a series of M rectangles (see figure 3.6) and assume the susceptibility $\kappa(x, z)$ is constant, say κ_i , in rectangle $i, i = 1, \dots, M$.

Choosing basis functions

$$\phi_i(x, z) = \begin{cases} 1, & \text{if } (x, z) \text{ is in the rectangle } i, \\ 0, & \text{otherwise,} \end{cases} \quad (3.108)$$

we get

$$\kappa(x, z) \approx \sum_{i=1}^M \kappa_i \phi_i(x, z). \quad (3.109)$$

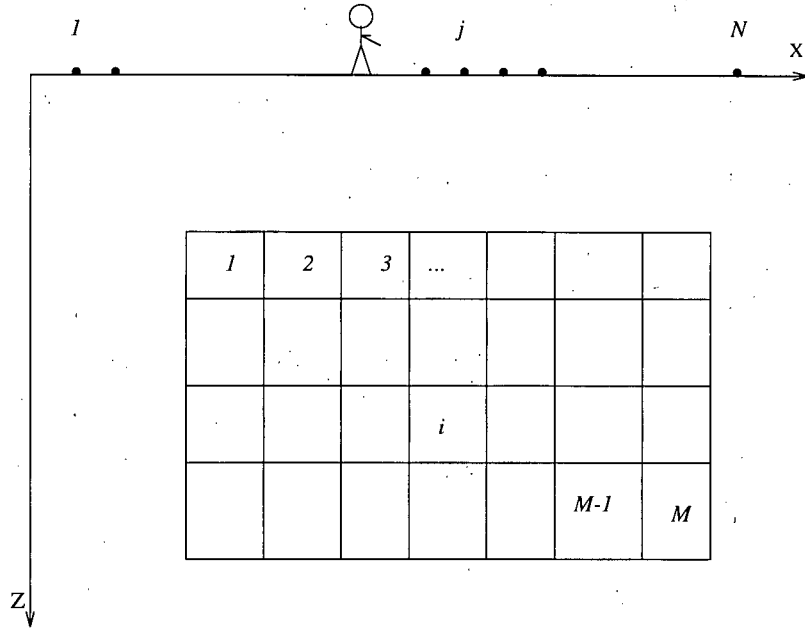


Figure 3.6: Base function model.

Substituting $\kappa(x, z)$ from (3.109) into (3.107) and using formulas (3.105) and/or (3.106), we arrive at the finite-dimensional linear problem

$$A\kappa = d \quad (3.110)$$

where A is an $N \times M$ matrix, $d = (d_1, \dots, d_N)^T = (d(x_1, z_0), \dots, d(x_N, z_0))^T$ are the observed data and $\kappa = (\kappa_1, \dots, \kappa_M)^T$ is the model to be determined. This is the discretized ill-posed inverse problem to be solved.

To use Tikhonov's regularization method for the integral equation (3.107), we introduce the following regularization operator

$$\begin{aligned} \|L(\kappa - \kappa_0)\|^2 = & \alpha_s \int_{area} w_s(\kappa - \kappa_0)^2 dx dz + \alpha_x \int_{area} w_x \left(\frac{\partial(\kappa - \kappa_0)}{\partial x} \right)^2 dx dz \\ & + \alpha_z \int_{area} w_z \left(\frac{\partial(\kappa - \kappa_0)}{\partial z} \right)^2 dx dz. \end{aligned} \quad (3.111)$$

Here κ_0 is an initial guess about the model, $w_x(x, z)$ and $w_z(x, z)$ are the weighting functions, and α_s , α_x and α_z are chosen to adjust the final computed model. If we choose $\alpha_x = 0$ and $\alpha_z = 0$, we call the final output “the smallest model”; this is similar to choosing $L = I$. If we choose $\alpha_s = 0$, we call the final output “the flattest model”; this is similar to choosing L as a differential operator with a nontrivial null space. Choosing the right parameters α_s , α_x and α_z depends on experience and prior knowledge about the true model to be determined.

For magnetic inverse problems, one important step is to choose proper weighting functions, $w_s(x, z)$, $w_x(x, z)$ and $w_z(x, z)$. It is obvious that the strength of the magnetic field on the surface depends on the depth of the magnetic body underground. Let us find the functional form of this relationship. For two-dimensional problems, if the rectangles are sufficiently small, we can consider each cell as a linear source which consists of dipoles along an infinitely long line. For a dipole with constant moment $\vec{M} = (\alpha, \beta, \gamma)$, the field at the origin is given by (3.100):

$$\vec{F} = \frac{1}{4\pi} \nabla(\vec{M} \cdot \nabla(1/r)) = \frac{1}{4\pi} \nabla(\alpha \frac{x}{r^3} + \beta \frac{y}{r^3} + \gamma \frac{z}{r^3}). \quad (3.112)$$

For a linear source, the total field at the origin is

$$\begin{aligned} \vec{F}_t &= \int \vec{F} dy \\ &= \int_{-\infty}^{\infty} \left(\frac{1}{4\pi} \nabla(\alpha \frac{x}{r^3} + \beta \frac{y}{r^3} + \gamma \frac{z}{r^3}) \right) dy \\ &= \frac{1}{4\pi} \nabla \left(\frac{2\alpha x + 2\gamma z}{x^2 + z^2} \right). \end{aligned}$$

and its projection in the direction $\vec{n} = (\alpha_0, \beta_0, \gamma_0)$ of any unit vector is

$$\begin{aligned} F_n &= \vec{n} \cdot \vec{F}_t \\ &= \frac{1}{4\pi} \alpha_0 \left[\frac{2\alpha}{x^2 + z^2} - \frac{2x(\alpha x + \gamma z)}{(x^2 + z^2)^2} \right] + \frac{1}{4\pi} \gamma_0 \left[\frac{2\gamma}{x^2 + z^2} - \frac{2z(\alpha x + \gamma z)}{(x^2 + z^2)^2} \right] \\ &\approx O\left(\frac{1}{z^2}\right) \text{ if } z \rightarrow \infty. \end{aligned}$$

So, we choose $1/z^2$ for each of the weighting functions $w_s(x, z)$, $w_x(x, z)$ and $w_z(x, z)$ to compensate for the decay of the magnetic field with depth.

After choosing the weighting functions, we can substitute $\kappa(x, z)$ from (3.109) into the regularization operator (3.111) and get, after discretization,

$$\phi_\kappa(\kappa) \equiv (\kappa - \kappa_0)^T (W_s^T W_s + W_x^T W_x + W_z^T W_z) (\kappa - \kappa_0) \quad (3.113)$$

or in general form,

$$\phi_\kappa(\kappa) \equiv (\kappa - \kappa_0)^T W_\kappa^T W_\kappa (\kappa - \kappa_0) \equiv \| W_\kappa (\kappa - \kappa_0) \|^2. \quad (3.114)$$

Combining equations (3.110) and (3.114), the ill-posed magnetic inverse problem is transformed to the following well-posed optimization problem:

$$\text{Minimize } \phi_\kappa \equiv \| W_\kappa (\kappa - \kappa_0) \|^2 \quad (3.115)$$

$$\text{Subject to } \phi_d \equiv \| W_d (G\kappa - d^{obs}) \|^2 = \phi_d^*, \quad (3.116)$$

where ϕ_d is our misfit criterion, W_d is an $N \times N$ data weighting matrix and ϕ_d^* is the target misfit. Here we choose ϕ_d^* according to the Morozov's discrepancy principle [14], that is, $\phi_d^* = E(\| W_d (d - d^{obs}) \|^2)$. If the observed data $d_i, i = 1, \dots, N$ have independent normal distributions $N(0, \sigma_i)$, then $\sum_{i=1}^N (\frac{d_i - d_i^{obs}}{\sigma_i})^2$ has a χ^2 distribution with expected value N . So, we may choose $\phi_d^* = N$ here when W_d is chosen as a diagonal matrix with diagonal elements $1/\sigma_i, i = 1, \dots, N$.

In practical computing we treat the constrained optimization problem (3.115), (3.116) by solving the following unconstrained problem:

$$\text{Minimize } \phi(\kappa) \equiv \phi_\kappa(\kappa) + \mu^{-1} (\phi_d(d) - \phi_d^*). \quad (3.117)$$

Here μ is the Lagrange multiplier. Differentiating with respect to the unknown vector κ produces the following regularized normal equation

$$(A^T W_d^T W_d A + \mu W_\kappa^T W_\kappa) \kappa = b \quad (3.118)$$

where $b = \mu W_\kappa^T W_\kappa \kappa_0 + A^T W_d^T W_d d$.

Equation (3.118) is a standard regularized normal equation and we can use the subspace method (*SLIM*) described in Chapter 2 to solve it.

3.1.5 A Numerical Example

Here we analyze a simple two dimensional magnetic problem. The forward problem is to produce the data using a known model. The inverse problem is to input the data from the forward problem and use them to reconstruct the model. For this magnetic problem, the forward problem is to produce the data from a known magnetic susceptibility distribution function, $\kappa = \kappa(x, y, z)$. The data here are measurements of the anomaly field along a line 2 meters above the surface. We use only the projection of the anomaly field onto the main field (geomagnetic field) direction as the data.

In this example we consider a rectangular area under ground. We set up a Cartesian coordinate system with y in the horizontal strike direction, z in downward direction and a corresponding x axis. Then the magnetic anomaly field is symmetric in y direction and we get a 2D problem.

Here we consider a 240 meter by 56 meter rectangle below the surface ($0 \leq x \leq 240$, $0 \leq z \leq 56$). We partition the rectangle into 280 cells with equal sizes: 12 meters in the x direction and 4 meters in the z direction. After adding bigger cells around the boundary of the rectangle, we get 560 cells all together. The magnetic susceptibility distribution we postulate is illustrated in Figure 3.9: it involves a high susceptibility part ($\kappa = 0.5$) inside the background part ($\kappa = 0.001$). Assuming the magnetic susceptibility is constant in each cell, we can compute the magnetic anomaly from each cell using the formula (3.105). The total magnetic anomaly can be produced by taking the vector sum of the magnetic anomalies of all the cells. In this example we compute the magnetic

anomaly at 200 equally spaced points along a line from $x=0$ to $x=240$.

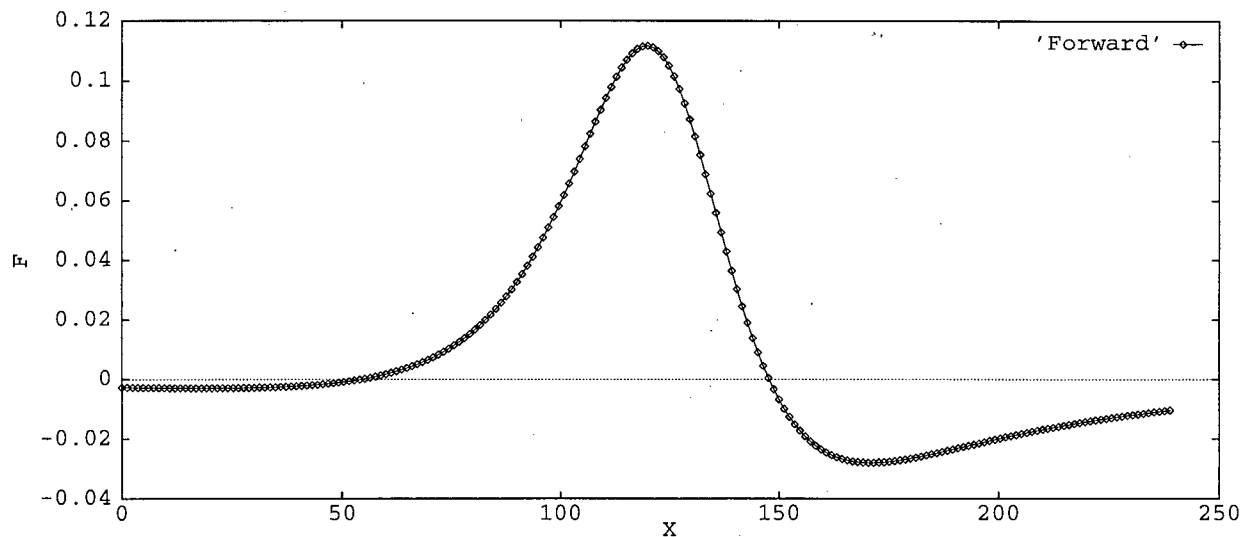


Figure 3.7: Exact magnetic data from the forward problem.

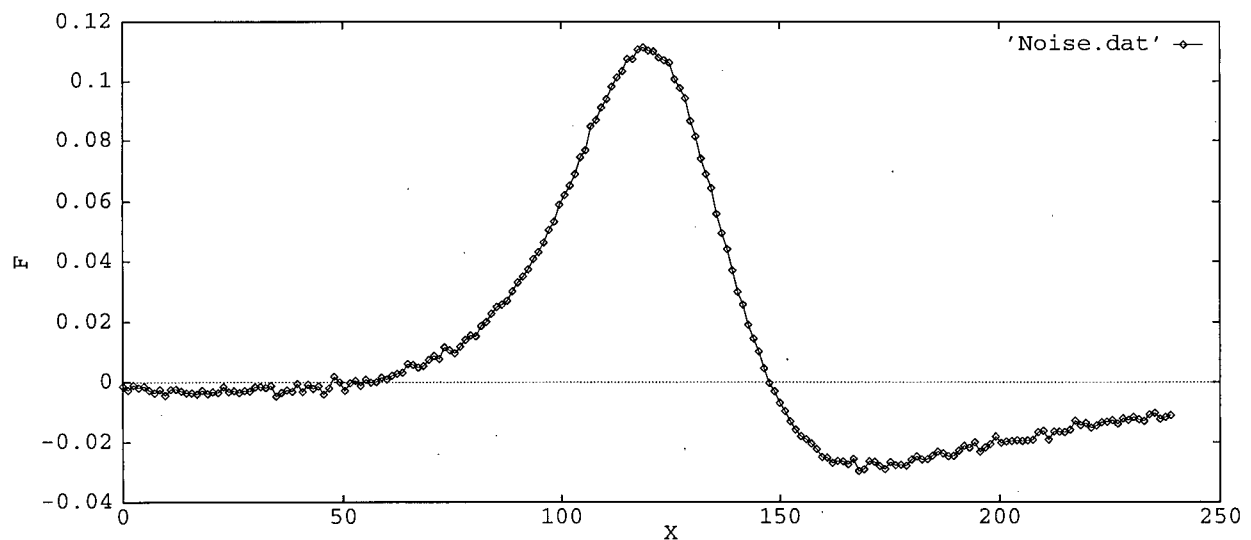


Figure 3.8: Noisy data simulating magnetic observations..

When we produce the data, we can calculate the coefficient matrix A for the inverse problem at the same time. In fact, from the formula (3.105) about the rectangular cell

we can produce each row of the matrix A before we calculate the data at each point. After we compute the data (figure 3.7) and the coefficient matrix A , we add normally distributed noise to the data, i.e.

$$d_i = \bar{d}_i + N(0, \sigma_i), \quad i = 1 \dots N \quad (3.119)$$

where $\sigma_i = \sigma$ equals 1% of the maximum of absolute values of the data. The contaminated data is our observed data d and the elements of the diagonal matrix W_d are the reciprocals of the standard deviations ($1/\sigma=10^3$ here).

In the regularization term,

$$\begin{aligned} \|L(\kappa - \kappa_0)\|^2 = & \alpha_s \int_{area} w_s (\kappa - \kappa_0)^2 dx dz + \alpha_x \int_{area} w_x \left(\frac{\partial(\kappa - \kappa_0)}{\partial x} \right)^2 dx dz \\ & + \alpha_z \int_{area} w_z \left(\frac{\partial(\kappa - \kappa_0)}{\partial z} \right)^2 dx dz, \end{aligned}$$

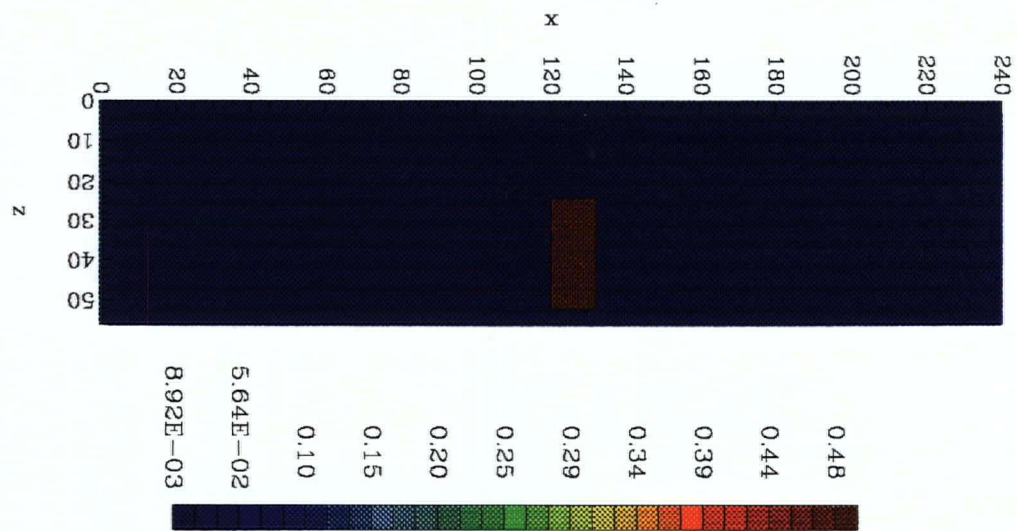
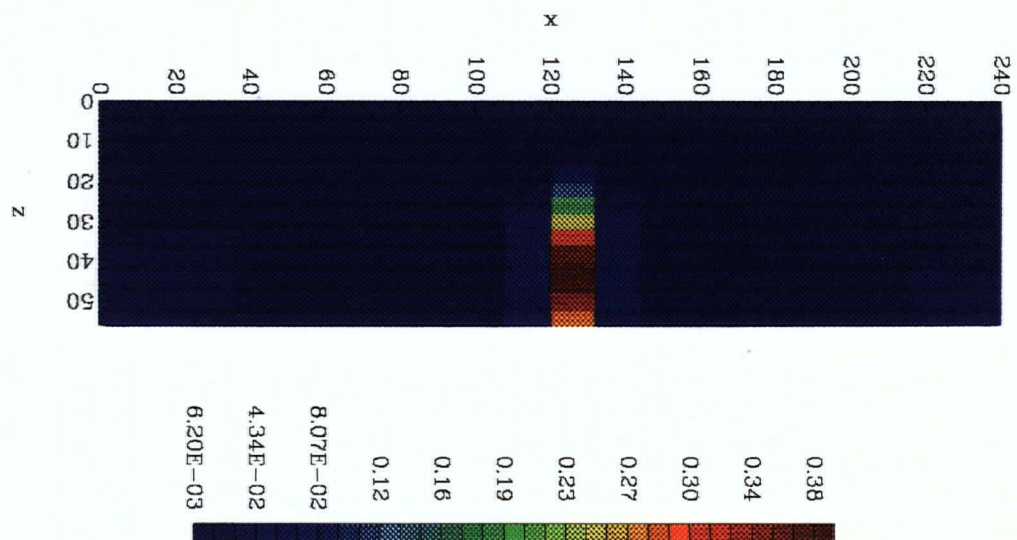
we choose $\kappa_0 = 0$, $w_s = w_x = w_z = 1/z^2$, $\alpha_s = 16$, $\alpha_x = 26$ and $\alpha_z = 70$, and produce the corresponding matrices, WSW , WXW and WZW with the relation

$$\|L\kappa\|^2 \approx \phi_\kappa = \kappa^T W_\kappa^T W_\kappa \kappa = \kappa^T W_s^T W_s \kappa + \kappa^T W_x^T W_x \kappa + \kappa^T W_z^T W_z \kappa. \quad (3.120)$$

With coefficient matrix A , contaminated data and regularization matrices we can set up the optimization problem like (3.117), and use *SLIM* (Subspace Linear Inversion Method) [15] to solve its regularized normal equation.

Since we mainly consider ferrimagnetic materials in magnetic prospecting, we require the reconstructed model to be nonnegative in the computing. The number of subspace vectors in this example is 5 (one from ϕ_d , one from ϕ_κ and three from the three parts of ϕ_κ) and the *SLIM* converges with about 80 steps. We plot the true model and numerically reconstructed model in figure (3.9).

Our inversion gives us the correct position and shape of the magnetic body, but the susceptibility of our calculated model is a little lower than that of the true model. The

**True Magnetic Susceptibility****Reconstructed Magnetic Susceptibility****Figure 3.9: Magnetic Model.**

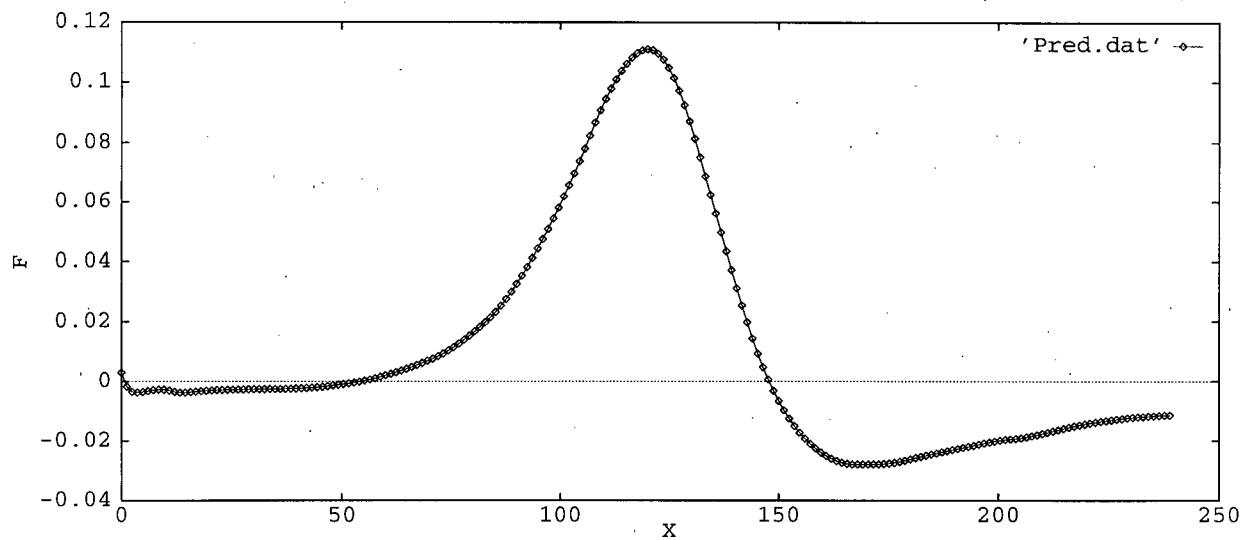


Figure 3.10: Predicted data from the output model.

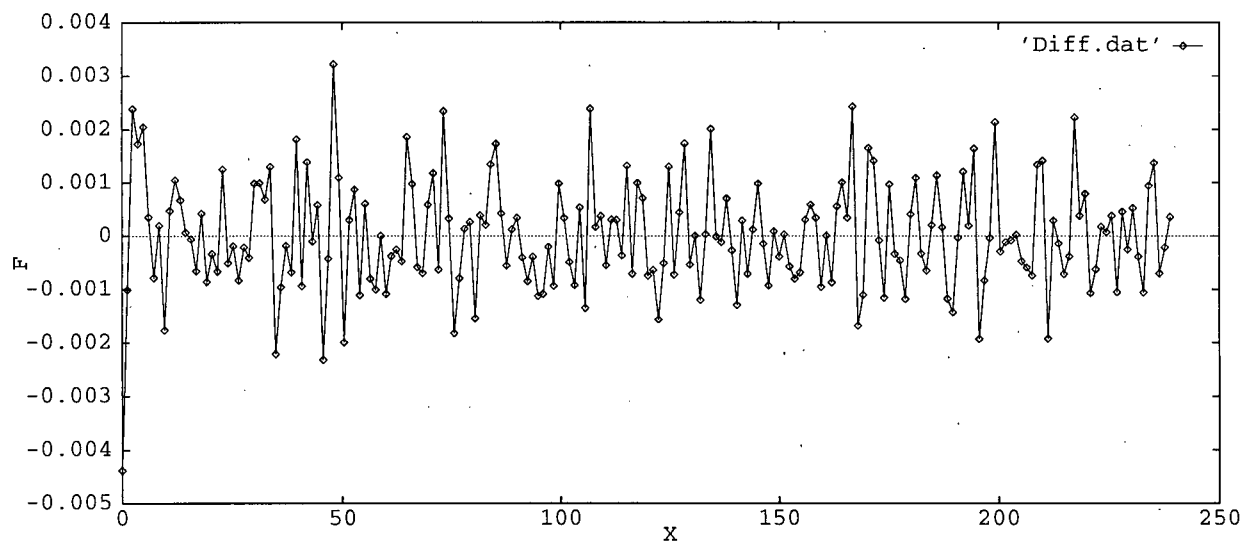


Figure 3.11: Difference between observed data and predicted data.

reason for this is that the solution of the magnetic inverse problem is not unique and the calculated model is smoothed by the regularization operator. So, even though we fit the data very well (Figures 3.7, 3.8, 3.10), our calculated model is still a little different from the true model (figure 3.9). Here we also notice that the difference between the observed data and the predicted data is small at each point and is independent of the observed points (figure 3.11), so it is reasonable to consider them as identical independent distribution random variables. Furthermore, our fitted data from the reconstructed model approximates the exact data better than our observed data with noise does (figure 3.7, 3.8, 3.11).

3.2 Induced Polarization

Induced polarization (IP) is a relatively new technique in geophysics, and has been employed mainly in base-metal exploration and to a minor extent in ground water search. As we have done for the magnetic problem, we treat only two dimensional situations here using a similar Cartesian coordinate system.

3.2.1 Physical Principles

A typical IP experiment involves putting a current into the ground and measuring the resulting potential some distance from the source. The simplest experimental equipment is a standard four-electrode DC (Direct Current) resistivity spread (figure 3.12). When the input current is suddenly interrupted the voltage across the potential electrodes generally does not drop to zero instantaneously, but decays rather slowly, after an initial large decrease from the original steady state value. This decay time is of the order of seconds or even minutes. If the current is switched on again, the potential, after a sudden initial increase, builds up over a similar time interval to the original DC amplitude (figure 3.13).

In one type of IP detector the decay voltage is measured as a function of time in various ways; this method is known *time-domain* IP. Since the build-up time is also finite, it is clear that the apparent resistivity (actually a complex impedance) must vary with the frequency, decreasing as the latter increases. Thus the measurement of apparent resistivity, ρ_a at two or more AC frequencies, generally below 10 Hz, constitutes another method of detection. This is known as *frequency-domain* IP. Measurements of IP may be made either in the time or the frequency domain. The former are known as pulse transient measurements, the latter as frequency variations. Several units of measurement which have been used in the two methods are defined below.

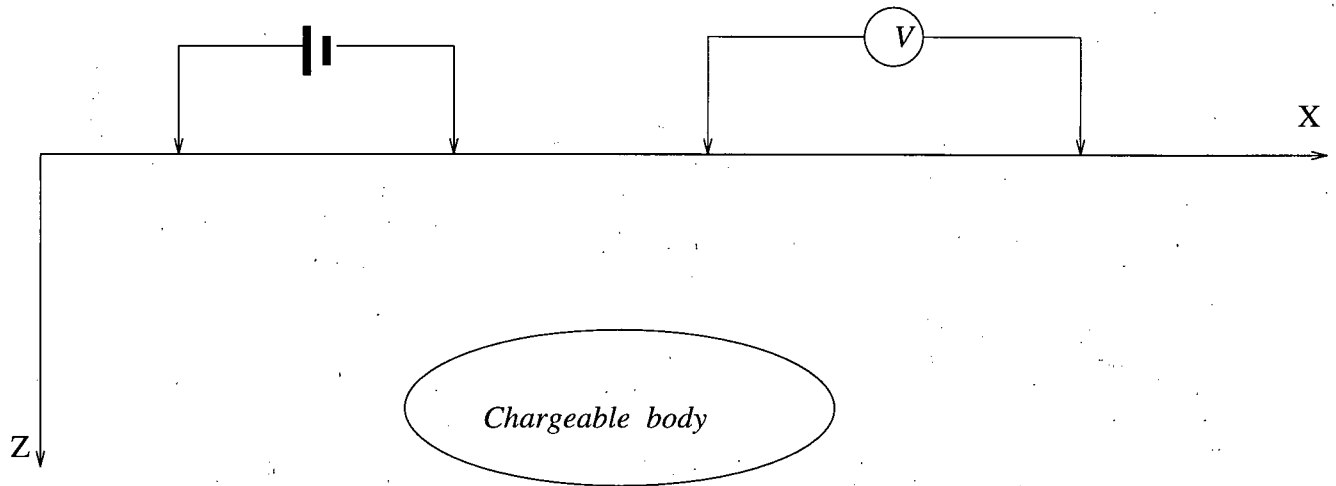


Figure 3.12: Four-electrode DC model.

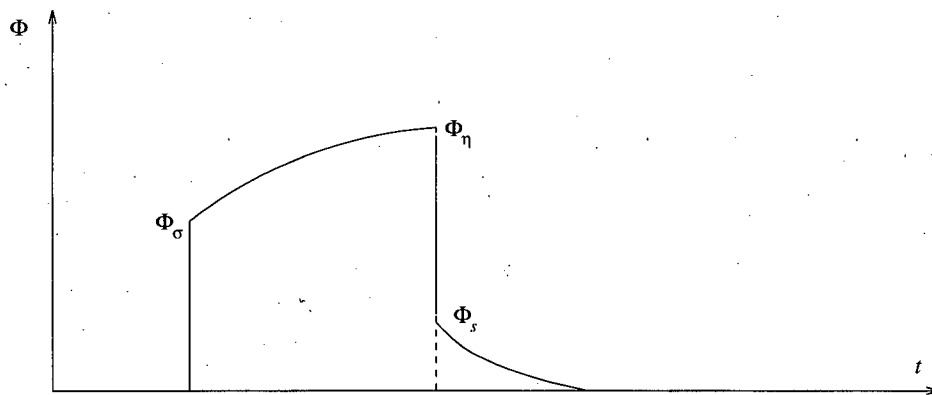


Figure 3.13: IP decay voltage.

Millivolts per volt (IP percent). The simplest way to measure IP effects with time domain equipment is to compare the residual voltage $\varphi(t)$ existing at a time t after the current is cut off with the final voltage φ_η during the current flow interval. Since $\varphi(t)$ is much smaller than φ_η , the ratio $\varphi(t)/\varphi_\eta$ is expressed as millivolts/volts, or as a percent, $100\varphi(t)/\varphi_\eta$, where both are in millivolts. The time interval t may vary between 0.1 and 10 seconds. The difficulty with this method is that it is not possible to measure potential at the instant of cut-off because of large transients caused by breaking the current circuit. On the other hand $\varphi(t)$ must be measured before the residual has decayed to noise level.

Decay-time integral. Commercial IP sets generally measure potential integrated over a definite time interval of transient decay. If the integration time is short and if the decay curve is sampled at several points, the values of the integral are effectively measurements of the potential existing at different times, i.e., $\varphi(t_1), \varphi(t_2), \dots, \varphi(t_n)$. This is an extension of the measurement above from which one also obtains the decay curve shape. Combining this method with the following definition of the chargeability we can calculate the integral interval by interval.

Chargeability. This is defined as

$$M = \frac{1}{\varphi_\eta} \int_{t_1}^{t_2} \varphi(t) dt \quad (3.121)$$

and is the most commonly used quantity in time domain IP measurement. When $\varphi(t)$ and φ_η have the same units, the chargeability M has the dimensions of time, and is usually reported in milliseconds.

Frequency effect. In frequency domain IP, one measures the apparent resistivity at two or more frequencies. The frequency effect is usually defined as

$$fe = \frac{\rho_{dc} - \rho_{ac}}{\rho_{ac}}, \quad (3.122)$$

where ρ_{dc} , ρ_{ac} are apparent resistivities measured at direct current and very high frequency. The per cent frequency effect is given by $PFE = 100fe$.

In theory, since both IP methods measure the same phenomenon, their results ought to be the same; practically, however, the conversion of time domain to frequency domain and vice versa is quite difficult.

Siegel [21] defines the chargeability as

$$M = \{ \lim_{t \rightarrow \infty} \varphi(t) - \lim_{t \rightarrow 0} \varphi(t) \} / \lim_{t \rightarrow \infty} \varphi(t) \quad (3.123)$$

By Laplace transform theory, it can be shown that

$$\lim_{t \rightarrow \infty} \varphi(t) = J\rho_{dc} \text{ and } \lim_{t \rightarrow 0} \varphi(t) = J\rho_{\infty}, \quad (3.124)$$

where ρ_∞ is the apparent resistivity at very high frequency and J is the current density. Consequently, assuming that $\rho_{ac} = \rho_\infty$, we can write for the chargeability

$$M = \frac{\rho_{dc} - \rho_\infty}{\rho_{dc}} = 1 - \frac{\rho_{ac}}{\rho_{dc}} = 1 - \frac{1}{1 + fe} = \frac{fe}{1 + fe} \approx fe \quad (3.125)$$

when $fe \ll 1$.

In practical situations this simple relation is not valid, partly because an exact theoretical analysis of the IP effect is not available (that is, the basic premises of the two systems of measurement are only approximately valid), and partly because the measurements are not made at DC and VHF (Very High Frequency) in either IP system. This complicates the conclusion from one result to the other.

3.2.2 Mathematical Relationship

According to the macroscopic representation of the physical property governing the IP response that was put forth by Siegel [21], we can use a macroscopic physical parameter called **chargeability** (a little different from that defined in (3.121)) to represent IP response. Here our earth model is described by two quantities: conductivity $\sigma(x, y, z)$ and chargeability $\eta(x, y, z)$. Both are positive, but while conductivity varies over many orders of magnitude, chargeability is confined to the region $[0,1]$.

A typical IP experiment involves putting a current I into the ground at the source position (x_s, y_s, z_s) and measuring the resulting potential at various points some distance from the source. In a time domain system the current has a duty cycle which alternates the direction of the current and has off-times between the current pulses at which the IP voltages are measured. Let φ_σ denote the potential that is measured in the absence of chargeability effects. This is the “instantaneous” potential measured when the current is turned on (see figure 3.14). In mathematical terms

$$\varphi_\sigma = F_{dc}[\sigma], \quad (3.126)$$

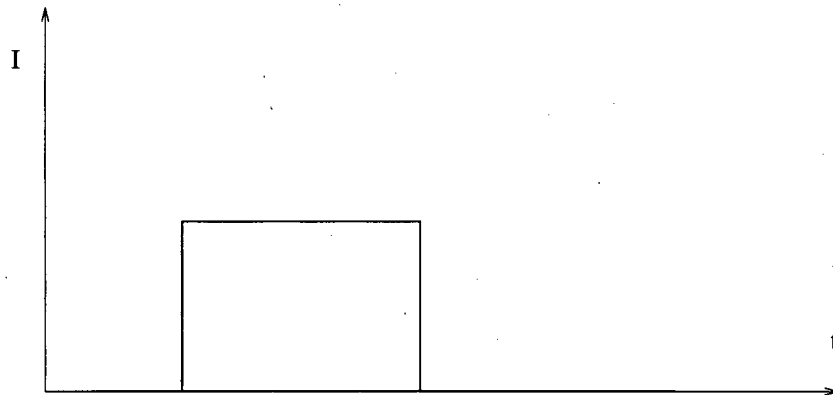


Figure 3.14: DC current for IP.

where the forward mapping operator F_{dc} is defined by the DC equation and its boundary conditions:

$$-\nabla \cdot (\sigma(x, z) \nabla \varphi_\sigma) = I \delta(x - x_s) \delta(y - y_s) \delta(z - z_s), \quad (3.127)$$

$$\frac{\partial}{\partial z} \varphi_\sigma(x, y, 0) = 0, \quad (3.128)$$

$$\lim_{R_s \rightarrow \infty} \varphi_\sigma(x, y, z) = 0 \quad (3.129)$$

where $R_s = \sqrt{(x - x_s)^2 + (y - y_s)^2 + (z - z_s)^2}$.

If the ground is chargeable then the potential φ_η , recorded well after the current is applied, will differ from φ_σ . According to Siegel's formulation, the effect of the chargeability of the ground is modelled by using the DC resistivity forward mapping F_{dc} but with the conductivity in (3.127) replaced by $\hat{\sigma} = \sigma(1 - \eta)$. That is,

$$\varphi_\eta = F_{dc}[\sigma(1 - \eta)] \quad (3.130)$$

with operator F_{dc} being defined as above (3.127)—(3.129).

3.2.3 The Forward Problem

The basic IP datum, which we refer to as **apparent chargeability**, is defined by (see figure 3.14)

$$\eta_a := \frac{\varphi_s}{\varphi_\eta} = \frac{\varphi_\eta - \varphi_\sigma}{\varphi_\eta} \quad (3.131)$$

or

$$\eta_a := \frac{F_{dc}[\sigma(1 - \eta)] - F_{dc}[\sigma]}{F_{dc}[\sigma(1 - \eta)]} \quad (3.132)$$

Equation (3.132) shows that the apparent chargeability can be computed by carrying out two DC resistivity forward modelling steps with the conductivities σ and $\sigma(1 - \eta)$. Here the equation (3.132) also defines the forward mapping for the IP data.

From the definition of chargeability (3.132), we see that here IP data depend on conductivity σ and chargeability η through differential operator F_{dc} . To produce the data, apparent chargeability η_a , we have to compute φ_σ and φ_η from known distributions of conductivity $\sigma(x, z)$ and chargeability $\eta(x, z)$. This requires solving the differential equation (3.127) with its boundary conditions two times. This is a forward Direct Current (DC) problem, an important subject in Geophysics. Although we refer to [12] for details regarding forward DC problems, it is possible to summarize the basic steps here.

1. Although a two-dimensional conductivity structure has been assumed in this work, the potential field that results from injecting current into the ground at a single point is still three-dimensional. The effective dimensionality of the problem can be reduced by considering the **Fourier Cosine** transform of the potential. This is possible since the potential can be thought as an even function of y . The transform is given by

$$\tilde{\varphi}(x, k_y, z) = \int_0^\infty \varphi(x, y, z) \cos(k_y y) dy \quad (3.133)$$

where k_y is a spatial wavenumber in the y direction. The transformed potential is

then found to satisfy the boundary value problem

$$-\frac{\partial}{\partial x}[\sigma(x, z)\frac{\partial \tilde{\varphi}}{\partial x}] + k_y^2 \sigma(x, z)\tilde{\varphi} - \frac{\partial}{\partial z}[\sigma(x, z)\frac{\partial \tilde{\varphi}}{\partial z}] = \frac{I}{2}\delta(x - x_s)\delta(z - z_s) \quad (3.134)$$

$$\frac{\partial}{\partial z}\tilde{\varphi}(x, k_y, 0) = 0 \quad (3.135)$$

$$\lim_{r_s \rightarrow \infty} \tilde{\varphi}(x, k_y, z) = 0 \quad (3.136)$$

where $r_s = \sqrt{(x - x_s)^2 + (z - z_s)^2}$.

2. Before using standard numerical methods (finite difference, finite elements and so on) to solve equation (3.134), we remove its singularity in the right-hand side by means of asymptotic approximation.
3. When discretizing the differential equation (3.134), we need to produce a proper approximation for the boundary condition (3.136) based on wide grids around the boundary.
4. After solving equation (3.134) with its boundary conditions for different wavenumbers k_y , we have to take the inverse **Fourier Transform**

$$\varphi(x, y, z) = \frac{2}{\pi} \int_0^\infty \tilde{\varphi}(x, k_y, z) \cos(k_y y) dk_y \quad (3.137)$$

to recover the potential $\varphi(x, z)$ in the spatial domain.

Theoretically our IP data are not only the apparent chargeability η_a but also the potentials φ_σ and φ_η , but in practical field work it is difficult to measure potentials at the instant the current is turned on or cut off. Instead the measured data are only the final potential φ_η and the remaining potentials ($\varphi_\eta - \varphi_\sigma$) after the current is switched

off. We usually approximate the potential $(\varphi_\eta - \varphi_\sigma)$ at the instant of cut-off with the remaining potential measured as soon as possible after the cut-off. Then the apparent chargeability η_a can be computed using formula (3.131).

3.2.4 The Inverse Problem

Linearization

The IP inverse problem is nonlinear and there are several ways to solve it [16]. One is to linearize the data equation (3.131) when the chargeability $\eta \ll 1$.

Let the earth model be partitioned into M cells and let η_i and σ_i denote the chargeability and electrical conductivity of the i^{th} cell. Linearizing the operator F_{dc} about the conductivity model σ yields

$$F_{dc}[\sigma - \eta\sigma] = F_{dc}[\sigma] + DF_{dc}[\sigma](-\eta\sigma) + o(\|\eta\sigma\|), \text{ i.e.,} \quad (3.138)$$

$$\varphi_\eta = \varphi(\sigma) - \sum_{j=1}^M \frac{\partial \varphi}{\partial \sigma_j} \eta_j \sigma_j + o(\|\eta\|) \quad (3.139)$$

Substituting into equation (3.131) yields

$$\eta_a = \frac{\varphi_\eta - \varphi_\sigma}{\varphi_\eta} \approx \frac{-\sum_j \frac{\partial \varphi}{\partial \sigma_j} \eta_j \sigma_j}{\varphi(\sigma) - \sum_j \frac{\partial \varphi}{\partial \sigma_j} \eta_j \sigma_j} \quad (3.140)$$

A further approximation yields

$$\eta_a \approx -\sum_j \frac{\sigma_j}{\varphi} \frac{\partial \varphi}{\partial \sigma_j} \eta_j. \quad (3.141)$$

Thus the i^{th} datum is

$$\eta_{ai} = \sum_{j=1}^M J_{ij} \eta_j, \quad i = 1, \dots, N, \quad (3.142)$$

where

$$J_{ij} = -\frac{\sigma_j}{\varphi_i} \frac{\partial \varphi_i}{\partial \sigma_j} \quad (3.143)$$

is the ij^{th} element of the sensitivity matrix.

So, instead of solving the nonlinear inverse problem (3.132) we can solve the linearized inverse problem

$$J\eta = \eta_a. \quad (3.144)$$

Computing The Sensitivity Matrix

To linearize the nonlinear inverse problem (3.132), we have to compute the sensitivity matrix J for equation (3.144). From the measured DC potentials a DC inverse problem is first solved to compute a background σ and a sensitivity matrix $(\partial\varphi_i/\partial\sigma_j)$. The DC inverse problem is nonlinear and we refer [12] for details based on the subspace iteration method. Here we briefly review the methods of computing the sensitivity matrix since it is useful in general linearization.

The most straightforward way to calculate the differential sensitivities is to approximate them using a one-side finite difference formula, but this method is inefficient since each sensitivity requires solving a completely new boundary-value problem.

Efficient methods of computing the sensitivity matrix use either sensitivity equations or Adjoint Equations.

1. **Sensitivity Equations.** For the steady state diffusion problem

$$L\varphi = -\vec{\nabla} \cdot (\sigma \vec{\nabla} \varphi) + q(\vec{x})\varphi = f(\vec{x}) \text{ in } D \quad (3.145)$$

with boundary condition

$$\alpha(\vec{x})\varphi + \beta(\vec{x})\frac{\partial\varphi}{\partial n} = 0 \text{ on } \partial D \quad (3.146)$$

we can use the sensitivity equation approach of Rodi [17].

Taking $\sigma(\vec{x})$ to be the model, and assuming the parameterization

$$\sigma(\vec{x}) = \sum_{i=1}^M \sigma_i \psi_i(\vec{x}) \quad (3.147)$$

where

$$\psi_i = \begin{cases} 1 & \text{if } \vec{x} \text{ in cell } i \\ 0 & \text{otherwise} \end{cases}$$

Substituting the σ from (3.147) into the equation (3.145) we get

$$L\varphi = -\vec{\nabla} \cdot \left(\sum_{i=1}^M \sigma_i \psi_i \vec{\nabla} \varphi \right) + q(\vec{x})\varphi = f(\vec{x}) \text{ in } D \quad (3.148)$$

Differentiating with respect to σ_i , and writing

$$\varphi_i(\vec{x}) = \frac{\partial \varphi(\vec{x})}{\partial \sigma_i} \quad (3.149)$$

yields the sensitivity problem

$$L\varphi_i = -\vec{\nabla} \cdot (\sigma \vec{\nabla} \varphi_i) + q(\vec{x})\varphi_i = \vec{\nabla} \cdot (\psi_i(\vec{x}) \vec{\nabla} \phi) \text{ in } D \quad (3.150)$$

$$\alpha(\vec{x})\varphi_i + \beta(\vec{x})\frac{\partial \varphi_i}{\partial n} = 0 \text{ on } \partial D \quad (3.151)$$

To compute the sensitivities for model $\sigma(\vec{x})$, the forward problem (3.145), (3.146) is first solved to obtain $\varphi(\vec{x})$ at all points \vec{x} in D . For each parameter p_i , the corresponding sensitivity problem must then be solved to obtain $\varphi_i(\vec{x})$ which is then evaluated at each of the observation locations. Since the source terms $\vec{\nabla} \cdot (\psi_i(\vec{x}) \vec{\nabla} \phi)$ are different for each i , a total of $M + 1$ forward problems must be solved to obtain all of the sensitivities.

Since the M sensitivity problems and the original forward problem differ only in terms of their right hand sides, they can all be solved using the same numerical forward algorithm. Since the sensitivity problems and the forward problem are identical except for their respective source terms, their numerical solution can be done very efficiently if a direct solver is used in the forward code.

2. **Adjoint Equations.** We can also calculate the sensitivities based on the adjoint Green's function. This approach has been used by Smith and Vozoff [19].

For the sensitivity problem (3.150)–(3.151) considered in the sensitivity equation approach, we consider the problem

$$L^* G^* = -\nabla \cdot (\sigma(\vec{x}) \nabla G^*) + q(\vec{x}) G^* = \delta(\vec{x} - \vec{x}_j) \text{ in } D \quad (3.152)$$

$$\alpha(\vec{x}) G^* + \beta(\vec{x}) \frac{\partial G^*}{\partial n} = 0 \text{ on } \partial D \quad (3.153)$$

where L^* is an adjoint operator and $G^*(\vec{x}, \vec{x}_j)$ is an adjoint Green's function.

Then we can get the sensitivity for an observation location from

$$\varphi_k(\vec{x}_j) = \int_D G^* \nabla \cdot (\psi_k \nabla \phi) dV. \quad (3.154)$$

For a $2D$ problem with M cells and N observation data we must solve $M + 1$ forward problems in the sensitivity equation approach and $N + 1$ forward problems in the adjoint equation approach. So we can use the sensitivity equation approach for overdetermined problems ($M \ll N$) and the adjoint equation approach for underdetermined problems ($M \gg N$). The IP problem here is underdetermined so we use the adjoint equation approach.

Solving The IP Inverse Problem

After we have linearized the discretized IP inverse problem and obtained

$$J\eta = d, \quad (3.155)$$

we can use standard regularization methods to solve it. Here we summarize the basic steps in solving IP inverse Problems.

1. Solve DC inverse problems with a subspace iteration method to get that background conductivity σ .
2. Use linearization to produce linear equation system (3.155).
3. Introduce the regularization operator $\| L\eta \|$ as in equation (3.111) with constant weighting functions.
4. Use *SLIM* to solve the regularized normal equation.

3.2.5 A Numerical Example

In this example, the forward problem is to produce the IP data and the inverse problem is to recover the IP model from the produced data. We work on the same domain introduced in the magnetic computing example since we will later be solving both problems together.

The chargeability distribution of the true model is illustrated in Figure 3.15: it consists of high chargeability part ($\eta=0.15$) and background chargeability part ($\eta=0.001$). The background conductivity is chosen as constant ($\sigma=0.6$ every cell).

Here we use a Pole-Pole configuration (similar to the four-electrode DC configuration in figure 3.12, but with one current electrode and one potential electrode at “infinity”) to introduce the current I and to measure the potential φ . We introduce electric current at 10 points sequentially. For each current source point, we measure the potential at 11 other points. So, we have 110 measurements all together.

A little different from the true experimental IP inverse problem, in this example we use linearization to produce the sensitivity matrix J first, then produce the synthetic data d with the formula

$$d = J\eta, \quad (3.156)$$

where η is the true model to be reconstructed. To simulate the measured data with

errors, we add normally distributed noises to the data, i.e.,

$$d_i^{obs} = d_i + \epsilon_i, \quad i = 1 \dots 110 \quad (3.157)$$

where ϵ_i has $N(0, \sigma_i)$ distribution. Here $\sigma_i = \sigma$ is 0.5% of the maximum of the absolute values of the exact 110 data.

After we get the data, we solve the discrete inverse problem according to the above steps and use the introduced regularization method to invert the data.

Here the regularization operator is

$$\begin{aligned} \| L(\eta - \eta_0) \|^2 = & \alpha_s \int_{area} w_s (\eta - \eta_0)^2 dx dz \\ & + \alpha_x \int_{area} w_x \left(\frac{\partial(\eta - \eta_0)}{\partial x} \right)^2 dx dz + \alpha_z \int_{area} w_z \left(\frac{\partial(\eta - \eta_0)}{\partial z} \right)^2 dx dz, \end{aligned}$$

where $\eta_0 = 0$, $w_s = w_x = w_z = 1$, $\alpha_s = 1$, $\alpha_x = 5$ and $\alpha_z = 10$, and its discretization produces the corresponding matrix

$$\| L\eta \| \approx \| W_\eta \eta \|^2. \quad (3.158)$$

According to regularization method, we solve optimization problem

$$\text{Minimize } \phi = \| W_\eta \eta \|^2 + \mu \| W_d(J\eta - d^{obs}) \|^2, \quad (3.159)$$

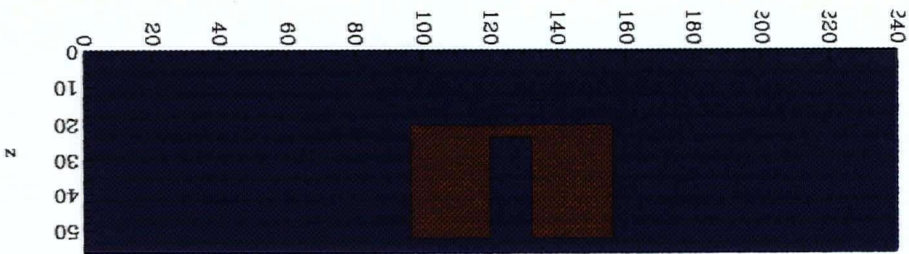
where W_d is a diagonal matrix and its elements are $1/\sigma_i$.

This is a standard optimization problem from linear inverse theory and we can solve it using *SLIM*.

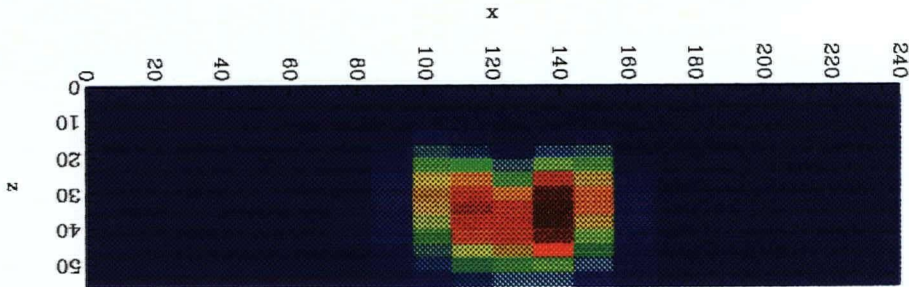
We plot the true model and the reconstructed model in figure (3.15)

From the calculated output, we can note that the reconstructed model is at the correct position corresponding to the true model and the chargeability value of the calculated model is consistent with the true model (figure 3.15). However, the inverse method can not reveal the fact that there is no chargeability in the interior of the true model.

This is due to the smoothing effect of the regularization operator which we use to solve the ill-posed IP inverse problem. However, it is possible to form IP structures with no chargeability in the interior, so it is important to develop inversion methods which can reveal this kind of IP model structure. This is what we will do in the next chapter.



True IP Chargeability



Reconstructed IP Chargeability



Figure 3.15: IP Model.

Chapter 4

Combined Inversion

To get reasonable approximate solutions to ill-posed inverse problems, we should use all the available information. This leads us to consider the combined inversion of geophysical data based on the correlation between the different physical parameters.

4.1 Known Approaches

To overcome the nonuniqueness inherent in geophysical inverse problems, we can use the regularization method. However, this mathematical technique does not provide the additional information needed to uniquely determine the true model. Instead, it simply chooses the approximate solution with the minimum (semi-)norm which is determined by its regularization operator. In this section we summarize the possible remedies for this problem.

First of all, to get a better approximation to the true model one natural approach is to collect more experimental data. Actually we use this approach for IP inverse problems. In IP field work the current electrode is moved to different points sequentially and the corresponding potentials are measured after the current electrode has been moved to each point. One obvious disadvantage of this approach is that we can not get as many data as we need, especially for practical problems.

The second approach is to reduce the set of possible reconstructed models. In this way one tries to overcome the nonuniqueness of ill-posed problems by restricting possible solutions to a smaller set based on reliable information regarding the true models. We use

this approach for both of our practical inverse problems. In solving both the magnetic problem and the IP problem, we require the inverted models to be nonnegative.

The third approach is to choose a proper regularization operator based on available information about the true models. For example, in solving magnetic inverse problems we choose weighting functions to compensate for the decay effects of the magnetic field. For general geophysical inverse problems, we should choose a regularization operator such that the reconstructed model is consistent with possible geological structures, mining recordings and so on.

Although these approaches are useful in solving inverse problems, they are based on one kind of observed data alone. If we try to combine the observed data of different kinds, we come to the correlation approach.

4.2 Correlation Between Parameters

To the above approaches for solving inverse problems, we would like to add a new approach—using the correlation between different model parameters.

In daily life, if we like to study a new material, it is natural for us to investigate its different characteristics. For example, to distinguish a kind of metal, we consider its color, density, conductivity and other characteristics at the same time. In the mineral industry and practical geophysical prospecting, it is common to use several kinds of prospecting methods in the same area. Therefore, observations of several different types are usually available.

Although different prospecting methods usually investigate different parameters of the ores, these parameters are often correlated. One good example regarding this kind of correlation is the relation between seismic wave velocity and density in an underground area. We can use gravity measurements to investigate the density of an underground

area. At the same time, we can use seismic experiments to study the velocity of waves propagating through the region. It is known that the density and the wave velocity in a region are positively correlated.

Here we would like to study the correlation between the chargeability η and the susceptibility κ of an underground region. From geophysical measurements [20], we know the magnetic susceptibility of an ore depends only upon the amount of ferrimagnetic minerals present, mainly magnetite, sometimes ilmenite or pyrrhotite. Chalcopyrite and pyrite are typical of many sulphide minerals which are basically nonmagnetic. At the same time, the chargeabilities of minerals increase with the increase of the sulphide concentration. Chalcopyrite and pyrite are typical sulphide minerals with high chargeabilities.

Here let us list the chargeability and susceptibility of some minerals from [20] in Table 4.2. The unit of chargeability η here is *msec* since it is computed using formula (3.121).

<i>name</i>	<i>chargeability η</i>	<i>susceptibility $\kappa \times 1000$</i>
Pyrite	13.4	1.5
Graphite	11.2	0.1
Chalcopyrite	9.4	0.4
Magnetite	2.2	6000
Hematite	0.0	6.5

Table 4.1: Chargeability and susceptibility of some minerals.

The chargeability and susceptibility data for various minerals and rocks in the tables of [20] suggest that these two quantities are negatively correlated.

This correlation can also be shown by a mineral example. McMillan [13] shows in the forming process of porphyry deposits that high chargeability material like pyrite gives way outward to high susceptibility metal like magnetite. This characteristic is simulated in our numerical example where there is high chargeability material is at the outside and material with high susceptibility inside.

We may not know the exact correlation between the susceptibility and chargeability

since it may depend on the sampling site and other unknown factors, but even an approximate effort to account for this correlation helps considerably to invert the magnetic and IP data.

4.3 Combined Inversion

Although we can invert the IP data for chargeability and the magnetic data for susceptibility independently, we propose to invert these two groups of data simultaneously. This will allow us to account for the correlation between the chargeability and the susceptibility. Since this amounts to adding more physical knowledge to our solution process, we expect it to give more reasonable solutions than the independent inversions do separately.

4.3.1 Optimization Problem

Now let us set up an optimization problem to invert the magnetic and IP data simultaneously. Under the assumptions for the magnetic problem stated in Chapter 3, the discrete magnetic inverse problem is

$$G_1 \kappa = d_1^{obs}, \quad (4.160)$$

where $\kappa \in R^M$ represents the model to be reconstructed. Assuming the discrete form of the regularization operator $\|L\kappa\|^2$ is

$$\phi_\kappa = \|W_\kappa(\kappa - \kappa_0)\|^2, \quad (4.161)$$

we arrive at the following optimization problem:

$$\begin{aligned} &\text{Minimize} \quad \phi_\kappa = \|W_\kappa(\kappa - \kappa_0)\|^2 \\ &\text{Subject to} \quad \phi_{d_1} = \|W_{d_1}(G_1 \kappa - d_1^{obs})\|^2 = \phi_{d_1}^* \end{aligned} \quad (4.162)$$

where ϕ_{d_1} is our misfit criterion, W_{d_1} is an $N_1 \times N_1$ data weighting matrix and $\phi_{d_1}^*$ is the target misfit.

Similarly for the IP inverse problem: if its discretized linear inverse problem is

$$G_2\eta = d_2^{obs}, \quad (4.163)$$

and its regularization operator has the form

$$\phi_\eta = \| W_\eta(\eta - \eta_0) \|^2, \quad (4.164)$$

then we arrive at the following optimization problem:

$$\begin{aligned} &\text{Minimize} \quad \phi_\eta = \| W_\eta(\eta - \eta_0) \|^2 \\ &\text{Subject to} \quad \phi_{d_2} = \| W_{d_2}(G_2\eta - d_2^{obs}) \|^2 = \phi_{d_2}^* \end{aligned} \quad (4.165)$$

where ϕ_{d_2} is our misfit criterion, W_{d_2} is an $N_2 \times N_2$ data weighting matrix and $\phi_{d_2}^*$ is the target misfit.

If the two parameters κ and η in each cell were independent, we could solve these two inverse problems independently. However, when there is a correlation between the two parameters, we should keep the solutions in the corresponding **Confidence Region**. Therefore, it is natural to solve the two inverse problems simultaneously.

If $(\kappa_i, \eta_i) \in R^2$ in cell i is a random vector and $\hat{\phi}_c(\kappa, \eta)$ represents its probability density function, then its confidence region is determined as:

$$\{(\kappa, \eta) : \hat{\phi}_c(\kappa, \eta) \leq \phi_c^*\}. \quad (4.166)$$

To keep the reconstructed model pair (κ_i, η_i) for each cell in its confidence region, we introduce a new optimization problem:

$$\begin{aligned} &\text{Minimize} \quad \phi_m := \| W_m(m - m_0) \|^2, \\ &\text{Subject to} \quad \phi_d := \| W_d(Gm - d^{obs}) \|^2 = \phi_d^*, \\ &\text{and} \quad \{(\kappa_i, \eta_i) : \phi_c(\kappa_i, \eta_i) \leq \phi_c^*\}, \end{aligned} \quad (4.167)$$

where

$$\begin{aligned}
 \kappa &= (\kappa_1 \dots \kappa_M)^T, \\
 \eta &= (\eta_1 \dots \eta_M)^T, \\
 m &= (\kappa, \eta)^T = (\kappa_1 \dots \kappa_M, \eta_1 \dots \eta_M)^T, \\
 d^{obs} &= (d_1^{obs}, d_2^{obs})^T, \\
 \phi_c &= \sum_{i=1}^M \hat{\phi}_c,
 \end{aligned}$$

and

$$\begin{aligned}
 G &= \begin{pmatrix} G_1 & \\ & G_2 \end{pmatrix}, \\
 W_d &= \begin{pmatrix} W_{d_1} & \\ & W_{d_2} \end{pmatrix}, \\
 W_m &= \begin{pmatrix} W_{m_1} & \\ & W_{m_2} \end{pmatrix}.
 \end{aligned}$$

The constrained optimization problem (4.167) can be transformed into the following unconstrained optimization problem

$$\begin{aligned}
 \text{Minimize } \phi &:= \phi_m + \mu \phi_d + \alpha_c \phi_c \\
 &= \|W_m(m - m_0)\|^2 + \mu \|W_d(Gm - d^{obs})\|^2 + \alpha_c \phi_c \quad (4.168)
 \end{aligned}$$

From (4.168) we see that choosing $\phi_c=0$ corresponds to solving for the two parameters $\eta \in R^M$ and $\kappa \in R^M$ independently; if we enforce a correlation between η and κ by setting $\phi_c \neq 0$, we may keep the parameters η and κ in a particular confidence region by choosing a suitable correlation weighting factor α_c .

Consider the case when ϕ_c is a convex quadratic function of $m = (\kappa, \eta)^T \in R^{2M}$, i.e.,

$$\phi_c = \frac{1}{2} m^T H m + g^T m + c, \quad (4.169)$$

where H is an $2M \times 2M$ positive definite matrix, g is a vector of $2M$ elements and c is a constant. Then solving optimization problem (4.168) amounts to solving its normal equation

$$Am = b, \quad (4.170)$$

where

$$A = W_m^T W_m + \mu G^T W_d^T W_d G + \alpha_c H, \quad (4.171)$$

and

$$b = W_m^T W_m m_0 + \mu G^T W_d^T W_d G d - \alpha_c g. \quad (4.172)$$

To solve linear equation system (4.170) we can use standard linear algebra techniques except choosing proper parameters μ and α_c . In fact both the direct methods and iterative methods reviewed in chapter 2 may be used to solve this equation system.

If ϕ_c is not a quadratic function of $m = (\kappa, \eta)^T$, we approximate ϕ_c using Taylor's expansion and solve problem (4.168) with iterative methods.

In fact, in a neighborhood of $m^{(n)}$, we have

$$\phi_c(m) \approx \phi_c(m^{(n)}) + g^T(m - m^{(n)}) + \frac{1}{2}(m - m^{(n)})^T H(m - m^{(n)}) \quad (4.173)$$

where g and H are the gradient and the Hessian matrix of ϕ_c at $m^{(n)}$ respectively.

Combining (4.168) and (4.173), we obtain the approximate optimization problem for iterative solution $m^{(n+1)}$:

$$\begin{aligned} \text{Minimize } \phi &:= \phi_m + \mu \phi_d + \alpha_c \phi_c \\ &\approx \|W_m(m - m_0)\|^2 + \mu \|W_d(Gm - d^{obs})\|^2 \\ &\quad + \alpha_c [\phi_c(m^{(n)}) + g^T(m - m^{(n)}) + \frac{1}{2}(m - m^{(n)})^T H(m - m^{(n)})]. \end{aligned} \quad (4.174)$$

To solve this optimization problem iteratively is an important part in the combined inversion, we will return to it later.

4.3.2 Statistical Description

To solve the optimization problem (4.174), we should know the correlation function $\phi_c(m)$.

Although it is impossible to determine an exact distribution of $(\kappa, \eta) \in R^2$ in each part, from the lists of magnetic susceptibilities and IP chargeabilities of various minerals and rocks in [20] we can find that *any mineral or rock does not likely have both high chargeability and high susceptibility*. We should approximate this correlation in our inversion for magnetic susceptibility and IP chargeability. Here are some possible choices.

1. Lognormal format. This is appropriate when the random vector $(\kappa, \eta) \in R^2$ has a lognormal distribution.

Let $x = \ln \kappa_i$, $y = \ln \eta_i$ in each cell, $i = 1 \dots M$. Under the assumption above, x and y have normal distributions with joint density function

$$f(x, y) = e^{-(x, y) \Sigma (x, y)^T}$$

where Σ is the covariance matrix.

Suppose (x_i, y_i) , $i = 1 \dots M$ are the observed values of (x, y) , then from statistical theory we know $\sum_{i=1}^M (x_i, y_i) \Sigma (x_i, y_i)^T$ is of $\chi^2(M)$ distribution. Therefore, the confidence region is

$$\{(x, y) : \sum_{i=1}^M (x_i - \bar{x}, y_i - \bar{y}) \Sigma (x_i - \bar{x}, y_i - \bar{y})^T = \chi^2(M) \leq C\}. \quad (4.175)$$

If we write in the function of κ and η , then the confidence region is determined by

$$\hat{\phi}_c(\kappa, \eta) = (\ln \kappa - \ln \kappa_0, \ln \eta - \ln \eta_0) \Sigma (\ln \kappa - \ln \kappa_0, \ln \eta - \ln \eta_0)^T. \quad (4.176)$$

Remark: Although the lognormal distribution is convenient theoretically, when we assume (κ, η) are lognormal we mean (κ_i, η_i) are always positive. However, (κ_i, η_i) may have

values being almost zero and this may result in computing difficulty when we compute the gradient and Hessian matrix of the function $\phi_c(\kappa, \eta)$.

2. Product format. In practical computing, a function like (4.176) is impractical, so we seek simpler functions that approximate the correlation $\hat{\phi}_c(\kappa, \eta)$.

One possible choice is the product of two probability density functions; one for group with high chargeability but low susceptibility and the other one for the group with low chargeability but high susceptibility. For example,

$$\begin{aligned} \hat{\phi}_c(\kappa, \eta) = & \{a_{11}(\kappa - \kappa_{01})^2 + 2a_{12}(\kappa - \kappa_{01})(\eta - \eta_{01}) + a_{22}(\eta - \eta_{01})^2\} \\ & \times \{b_{11}(\kappa - \kappa_{02})^2 + 2b_{12}(\kappa - \kappa_{02})(\eta - \eta_{02}) + b_{22}(\eta - \eta_{02})^2\}, \end{aligned} \quad (4.177)$$

where κ and η represent the susceptibility and chargeability of each cell, and κ_{01} and η_{01} are the average values of susceptibility and chargeability of groups with high susceptibility and low susceptibility. Similarly κ_{02} and η_{02} are the average values of susceptibility and chargeability of groups with low susceptibility and high susceptibility. Usually we choose $a_{12} = b_{12} = 0$.

In this choice, we have a lot of freedom to adjust the shape of the contours determining the boundaries of the confidence regions, but the computing is still somewhat complicated.

An even simpler example is

$$\hat{\phi}_c(\kappa, \eta) = \kappa^2 \eta^2. \quad (4.178)$$

Obviously, if the value of function $\phi_c(\kappa, \eta)$ is small, it is impossible for the both values of κ and η to be large. So, this function really reflects the correlation between κ and η in each cell. If the relation for κ and η is symmetric, (4.178) should be a good choice. Although $\hat{\phi}_c$ may remain small when one of κ and η becomes huge, this possibility is avoided by the presence of the regularization operator.

Furthermore, if we choose $a_{11} = b_{22} = 1$, $a_{12} = a_{22} = 0$, $b_{12} = b_{11} = 0$, $\eta_{01} = \kappa_{01} = 0$ and $\eta_{02} = \kappa_{02} = 0$ in function (4.177), what we get is just function (4.178). So function (4.178) is a special case of function (4.177).

4.4 Computing Techniques

After we approximate the correlation with the function $\phi_c(\kappa, \eta)$, we can solve the optimization problem of combined inversion. This is not trivial here since the objective function is neither quadratic nor convex.

4.4.1 Optimization Problem

For the optimization problem (4.168)

$$\begin{aligned} \text{Minimize } \phi &:= \phi_m + \mu\phi_d + \alpha_c\phi_c \\ &= \|W_m(m - m_0)\|^2 + \mu \|W_d(Gm - d^{obs})\|^2 + \alpha_c\phi_c, \end{aligned}$$

if we introduce the notation

$$\begin{aligned} \phi_d &\equiv \|Am - d^{obs}\|^2 \\ &= \|G_1\kappa - d_1^{obs}\|^2 + \|G_2\eta - d_2^{obs}\|^2 \\ &= \phi_{d_1} + \phi_{d_2} \end{aligned}$$

and

$$\begin{aligned} \phi_m &\equiv \|W_m(m - m_0)\|^2 \\ &= \|W_\kappa(\kappa - \kappa_0)\|^2 + \|W_\eta(\eta - \eta_0)\|^2 \\ &= \phi_\kappa + \phi_\eta. \end{aligned}$$

it will become the following unconstrained optimization problem

$$\begin{aligned} \text{Minimize } \phi &\equiv \mu\phi_d + \phi_m + \alpha_c\phi_c \\ &= \mu\phi_{d_1} + \mu\phi_{d_2} + \phi_\kappa + \phi_\eta + \alpha_c\phi_c. \end{aligned} \quad (4.179)$$

Here we choose a common μ for the misfit ϕ_d . Theoretically the misfit term should be chosen as $\mu_1\phi_{d_1} + \mu_2\phi_{d_2}$, but this choice will bring more parameters to be decided. In the numerical examples, the computing output shows $\phi_{d_1} \approx \phi_{d_1}^*$ and $\phi_{d_2} \approx \phi_{d_2}^*$ even though we just choose $\mu_1 = \mu_2 = \mu$.

4.4.2 Approximating the Nonlinear Correlation

If we let κ_i and η_i denote the susceptibility and chargeability of each cell i and function

$$\begin{aligned} \hat{\phi}_c(\kappa_i, \eta_i) &= \{a_{11}(\kappa_i - \kappa_{01})^2 + 2a_{12}(\kappa_i - \kappa_{01})(\eta_i - \eta_{01}) + a_{22}(\eta_i - \eta_{01})^2\} \\ &\quad \times \{b_{11}(\kappa_i - \kappa_{02})^2 + 2b_{12}(\kappa_i - \kappa_{02})(\eta_i - \eta_{02}) + b_{22}(\eta_i - \eta_{02})^2\} \end{aligned} \quad (4.180)$$

denote their correlation in the cell, then the total correlation ϕ_c in optimization problem (4.179) is

$$\begin{aligned} \phi_c &= \sum_{i=1}^M \hat{\phi}_c(\kappa_i, \eta_i) \\ &= \sum_{i=1}^M \{a_{11}(\kappa_i - \kappa_{01})^2 + 2a_{12}(\kappa_i - \kappa_{01})(\eta_i - \eta_{01}) + a_{22}(\eta_i - \eta_{01})^2\} \\ &\quad \times \{b_{11}(\kappa_i - \kappa_{02})^2 + 2b_{12}(\kappa_i - \kappa_{02})(\eta_i - \eta_{02}) + b_{22}(\eta_i - \eta_{02})^2\}. \end{aligned} \quad (4.181)$$

Similarly, if we use the correlation function

$$\hat{\phi}_c(\kappa_i, \eta_i) = \kappa_i^2 \eta_i^2 \quad (4.182)$$

instead of function (4.180), we can get corresponding total correlation

$$\phi_c = \sum_{i=1}^M \hat{\phi}_c(\kappa_i, \eta_i) = \sum_{i=1}^M \kappa_i^2 \eta_i^2. \quad (4.183)$$

For both cases (4.181) and (4.183), ϕ_c is not a quadratical function of

$$m = (\kappa_1 \dots \kappa_M, \eta_1 \dots \eta_M)^T, \quad (4.184)$$

so we have to use an iterative method to solve problem (4.179).

At each iteration we solve a problem of the form (4.174). This is Newton's method. In each step, we have to compute the gradient vector, $g = (g_1, \dots, g_{2M})^T$, and Hessian matrix, $H = (h_{ij})_{i,j=1}^{2M}$ of ϕ_c , that is,

$$g_i = \frac{\partial \phi_c(m)}{\partial m_i} \Big|_{m=m(n)} \quad (4.185)$$

and

$$h_{ij} = \frac{\partial^2 \phi_c(m)}{\partial m_i \partial m_j} \Big|_{m=m(n)}, \quad (4.186)$$

where $m = (\kappa_1 \dots \kappa_M, \eta_1 \dots \eta_M)^T$.

For our specific correlation function ϕ_c (4.183), we have

$$g_i = \begin{cases} 2\kappa_i \eta_i^2 & i \leq M \\ 2\kappa_i^2 \eta_i & i > M \end{cases}$$

and

$$h_{ij} = \begin{cases} 2\eta_i^2 & i = j \leq M, \\ 2\kappa_i^2 & i = j > M, \\ 4\kappa_i \eta_i & i \leq M \text{ and } j = M + i, \\ 4\kappa_i \eta_i & i > M \text{ and } j = i - M, \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, we can find the analytical expressions of g and H for the specific correlation function ϕ_c of (4.181).

4.4.3 Producing the Initial Model

Neither of correlation functions ϕ_c in (4.181 or (4.183)) is convex.

In the first case,

$$\hat{\phi}_c = \kappa^2 \eta^2, \quad (4.187)$$

has Hessian matrix

$$\begin{bmatrix} 2\eta^2 & 4\kappa\eta \\ 4\kappa\eta & 2\kappa^2 \end{bmatrix},$$

which has negative determinant, so it is nonconvex.

In the second case,

$$\begin{aligned} \hat{\phi}_c = & (\kappa_i - \kappa_{01})^2 + 2a_{12}(\kappa_i - \kappa_{01})(\eta_i - \eta_{01}) + a_{22}(\eta_i - \eta_{01})^2 \} \\ & \times \{b_{11}(\kappa_i - \kappa_{02})^2 + 2b_{12}(\kappa_i - \kappa_{02})(\eta_i - \eta_{02}) + b_{22}(\eta_i - \eta_{02})^2\}, \end{aligned}$$

with $a_{12} = b_{12} = 0$. To show it is a nonconvex function we can choose

$$m^1 = (\kappa_{01}, \eta_{01})^T \in R^2,$$

$$m^2 = (\kappa_{02}, \eta_{02})^T \in R^2,$$

then $\hat{\phi}_c(m^1) = 0$ and $\hat{\phi}_c(m^2) = 0$. However, when we let

$$m^3 = (m^1 + m^2)/2, \quad (4.188)$$

then $\phi_c(m^3) > 0$ unless $a_{11} = a_{22} = b_{11} = b_{22} = 0$ which is a trivial case.

Actually we are using Newton's method when we expand ϕ_c as in equation (4.174) and use iterative methods to solve it. To get a reasonable solution to this kind of optimization problem, we must start with a good initial model $m^{(0)}$.

Our approach is to generate an initial model by setting $\phi_c = 0$ in the optimization problem (4.179). In other words, we minimize

$$\begin{aligned} \phi(m) &= \mu \phi_d + \phi_m \\ &= \mu \| Gm - d^{obs} \|^2 + \| W_m(m - m_0) \|^2. \end{aligned} \quad (4.189)$$

This is a standard optimization problem for linear inverse theory and we solve it as in Chapter 3 to produce the initial model $m^{(0)}$.

4.4.4 Choosing the Correlation Weighting Parameter

After we introduce $\phi_c(m)$ to the optimization problem (4.179), we note the effects of correlation term ϕ_c and regularization term ϕ_m are in opposition. The regularization term ϕ_m increases as the correlation term ϕ_c decreases. The reason is that our regularization operator tends to smooth the reconstructed model while the correlation tends to reveal the nonsmooth characteristic of the true model. Therefore, we should choose a proper α_c .

If we choose α_c too large, the initial model gets a huge perturbation and may not converge to the ideal model anymore. On the other hand, if we choose α_c too small, the smoothing effect of ϕ_m may swamp the effect of correlation. To balance the effects of the correlation term and the regularization term, we should choose α_c such that these two terms are compatible.

However, the correlation ϕ_c decreases as the iteration proceeds, so α_c should be increased continually to maintain the compatibility between ϕ_c and $\|Lm\|^2$.

Actually, in each big step, we fix the parameter α_c and use *SLIM* to solve the linear system

$$Am = b, \text{ where} \quad (4.190)$$

$$A = W_m^T W_m + \mu G^T W_d^T W_d^T G + \alpha_c H, \quad (4.191)$$

$$b = W_m^T W_m m_0 + \mu G^T W_d^T W_d G d - \alpha_c g + \alpha_c H m^{(n)}. \quad (4.192)$$

Recall that (4.190) is the normal equation of the optimization problem (4.174).

After *SLIM* solves problem (4.195), we update the parameter α_c and run *SLIM* again. For our numerical examples, we usually need several steps to get the final results. How to modify *SLIM* such that it can choose α_c automatically will need further work in the future.

Here we update α_c by replacing it with $c\alpha_c$, where c is an increasing factor ($c > 1$). The initial α_c is chosen such that $\alpha_c\phi_c$ is compatible with ϕ_m . Similarly, in each following step, we choose α_c such that $\alpha_c\phi_c$ is compatible with the new value of ϕ_m .

4.4.5 Balancing the Regularization Operator

In problem (4.174), the model norm ϕ_m is the regularization operator. Here we have

$$\phi_m = \phi_\kappa + \phi_\eta, \quad (4.193)$$

where

$$\begin{aligned} \phi_\kappa &= \|L(\kappa - \kappa_0)\|^2 \\ &= \phi_{s1} + \phi_{x1} + \phi_{z1} \\ &= \alpha_{s1} \int_{area} w_s(\kappa - \kappa_0)^2 dx dz + \\ &\quad \alpha_{x1} \int_{area} w_x \left(\frac{\partial(\kappa - \kappa_0)}{\partial x} \right)^2 dx dz + \alpha_{z1} \int_{area} w_z \left(\frac{\partial(\kappa - \kappa_0)}{\partial z} \right)^2 dx dz. \end{aligned} \quad (4.194)$$

$$\begin{aligned} \phi_\eta &= \|L(\eta - \eta_0)\|^2 \\ &= \phi_{s2} + \phi_{x2} + \phi_{z2} \\ &= \alpha_{s2} \int_{area} w_s(\eta - \eta_0)^2 dx dz + \\ &\quad \alpha_{x2} \int_{area} w_x \left(\frac{\partial(\eta - \eta_0)}{\partial x} \right)^2 dx dz + \alpha_{z2} \int_{area} w_z \left(\frac{\partial(\eta - \eta_0)}{\partial z} \right)^2 dx dz. \end{aligned} \quad (4.195)$$

When we solve inverse problems using regularization method, our output is smoothed by the regularization operator. At the same time, the correlation term ϕ_c reduces the smoothing effect of the regularization operator. To keep the smoothing effects of ϕ_κ and ϕ_η to be compatible, we require the values of ϕ_κ and ϕ_η to be similar. So, we should choose proper α_{s1} , α_{x1} and α_{z1} , and α_{s2} , α_{x2} and α_{z2} such that ϕ_κ and ϕ_η are compatible. We will provide these values in our numerical examples.

4.4.6 Working with SLIM

SLIM [15] is designed mainly for linear inverse problems. For our combined inversion with correlation term ϕ_c which is neither quadratic nor convex we have to approximate it iteratively, so we should modify the original code.

After changing the code of *SLIM*, we can calculate H and g in each iteration. Furthermore, we can compute the gradient of ϕ_c in each step which is useful in producing the vectors of the subspace.

When we use the subspace method to solve problem (4.179), in each iteration we solve a linear system

$$B\delta m = b \quad (4.196)$$

for a perturbation δm in the subspace $V_s = \text{span}\{v_1, \dots, v_q\}$ instead of the whole solution space R^{2M} , that is, $\delta m = \sum_{k=1}^q \beta_k v_k$. Here

$$B = (W_m^T W_m + \mu G^T W_d^T W_d G + \alpha_c H),$$

and

$$b = -W_m^T W_m (m^{(n)} - m_0) - \mu G^T W_d^T W_d (Gm^{(n)} - d^{obs}) - \alpha_c g$$

Choosing the right subspace V_s is an important step in subspace methods. For our case, we produce the subspace vectors as follows:

$$\begin{aligned} v_1 &= \nabla \phi_{d_1}, & v_2 &= \nabla \phi_{d_2}, \\ v_3 &= \nabla \phi_{\kappa}, & v_4 &= \nabla \phi_{\eta}, \\ v_5 &= \nabla(\phi_{\kappa} + \phi_{\eta}), & v_6 &= \nabla(\phi_{\kappa} + \phi_{\eta}) \\ v_7 &= \nabla(\phi_{s1} + \phi_{s2}), & v_8 &= \nabla(\phi_{x1} + \phi_{s2}) \\ v_9 &= \nabla(\phi_{z1} + \phi_{z2}) & v_{10} &= \nabla \phi_c, \end{aligned}$$

so $q = 10$ here.

To produce the steepest descent directions, the Conjugate Gradient Method is used to solve the related linear system of equations. We refer to [15] for details.

4.5 A Numerical Example

We have already solved a magnetic inverse problem and an IP inverse problem independently in Chapter 3. Now we can solve them simultaneously with our combined inversion method.

According to the above steps, we first solve problem (4.179) with $\alpha_c = 0$. To solve this problem we choose 1.2 as the decreasing factor of ϕ_d in the *SLIM*. Here we list the parameters for the regularization operator $\|Lm\|^2$ in table (4.5).

<i>name</i>	α_s	α_x	α_z
ϕ_κ	160.	260.	2000.
ϕ_η	0.5	2.5	5.

Table 4.2: Parameters of regularization operator.

Then we use the correlation function ϕ_c defined by

$$\phi_c = \sum_{i=1}^M \hat{\phi}_c(\kappa_i, \eta_i) = \sum_{i=1}^M \kappa_i^2 \eta_i^2. \quad (4.197)$$

We choose the initial correlation coefficient $\alpha_c = 1.E4$ and increase α_c with the multiplying factor $c = 1.5$. The outputs are plotted here (see figure 4.16). In a small departure from the theoretical analysis, here I chose 320 as the target misfit for 310 data.

When we set $\alpha_c = 0$, we get the magnetic inversion model and IP inversion model plotted in figure (4.16). The output shows that the smoothing effect of the regularization operator makes the magnetic susceptibility κ low and the IP chargeability η smooth (compare the output in Table 4.5).

After we introduce $\alpha_c \neq 0$ at iteration 306, we get significantly improved reconstructions for both model components (figure 4.25) with $\phi_c(m)$ being defined as in (4.178)

<i>name</i>	$\phi_{d,\kappa}$	ϕ_κ	$\phi_{d,\eta}$	ϕ_η	Correlation ϕ_c
True Model	192	28	102	51	0.25×10^{-5}
Inverted Model ($\alpha_c = 0$)	228	20	92	58	0.62×10^{-2}
Inverted Model ($\alpha_c \neq 0$)	221	63	99	201	0.58×10^{-4}

Table 4.3: Output with $\alpha_c = 0$.

(see Table 4.5).

From the output we find the new calculated magnetic model is a better approximation to the true model and the new calculated IP model reveals the discontinuity of the true model.

The iteration process is shown in figures (4.17), (4.18), (4.19) and (4.20). In fact, although the total misfit does not change much after introducing correlation factor α_c , the correlation decreases slowly.

Similarly, if we choose a correlation function $\phi_c(m)$ defined as

$$\begin{aligned}
 \phi_c &= \sum_{i=1}^M \hat{\phi}_c(\kappa_i, \eta_i) \\
 &= \sum_{i=1}^M \{a_{11}(\kappa_i - \kappa_{01})^2 + 2a_{12}(\kappa_i - \kappa_{01})(\eta_i - \eta_{01}) + a_{22}(\eta_i - \eta_{01})^2\} \\
 &\quad \times \{b_{11}(\kappa_i - \kappa_{02})^2 + 2b_{12}(\kappa_i - \kappa_{02})(\eta_i - \eta_{02}) + b_{22}(\eta_i - \eta_{02})^2\},
 \end{aligned}$$

where

$$\begin{aligned}
 a_{11} &= 0.1, & a_{22} &= 64, \\
 \kappa_{01} &= 0.5, & \eta_{01} &= 0.01, \\
 a_{12} &= 0, & b_{12} &= 0, \\
 b_{11} &= 64, & b_{22} &= 0.1, \\
 \kappa_{02} &= 0.01, & \eta_{02} &= 0.15,
 \end{aligned}$$

then the correlation ϕ_c for true model is 0.89. If we solve the optimization problem (4.179) with $\alpha_c = 0$, then the correlation ϕ_c is 18. After we choose $\alpha_c > 0$ and use the

combined inversion, we also observe improved reconstruction relative to the independent inversion in Chapter 3. (figure 4.26) (see 4.5). Its correlation curve is plotted in Figure (4.23) and the total objective function ϕ curve is plotted in Figure (4.22). We note both the correlation ϕ_c and total objective function ϕ decrease with the iteration, after we let $\alpha_c > 0$.

<i>name</i>	ϕ_d	ϕ_m
Magnetic	232.6	55.9
IP	87.4	280.4
Correlation	(from) 18.0	(to) 0.84

Table 4.4: Output with $\alpha_c \neq 0$.

4.6 Conclusion

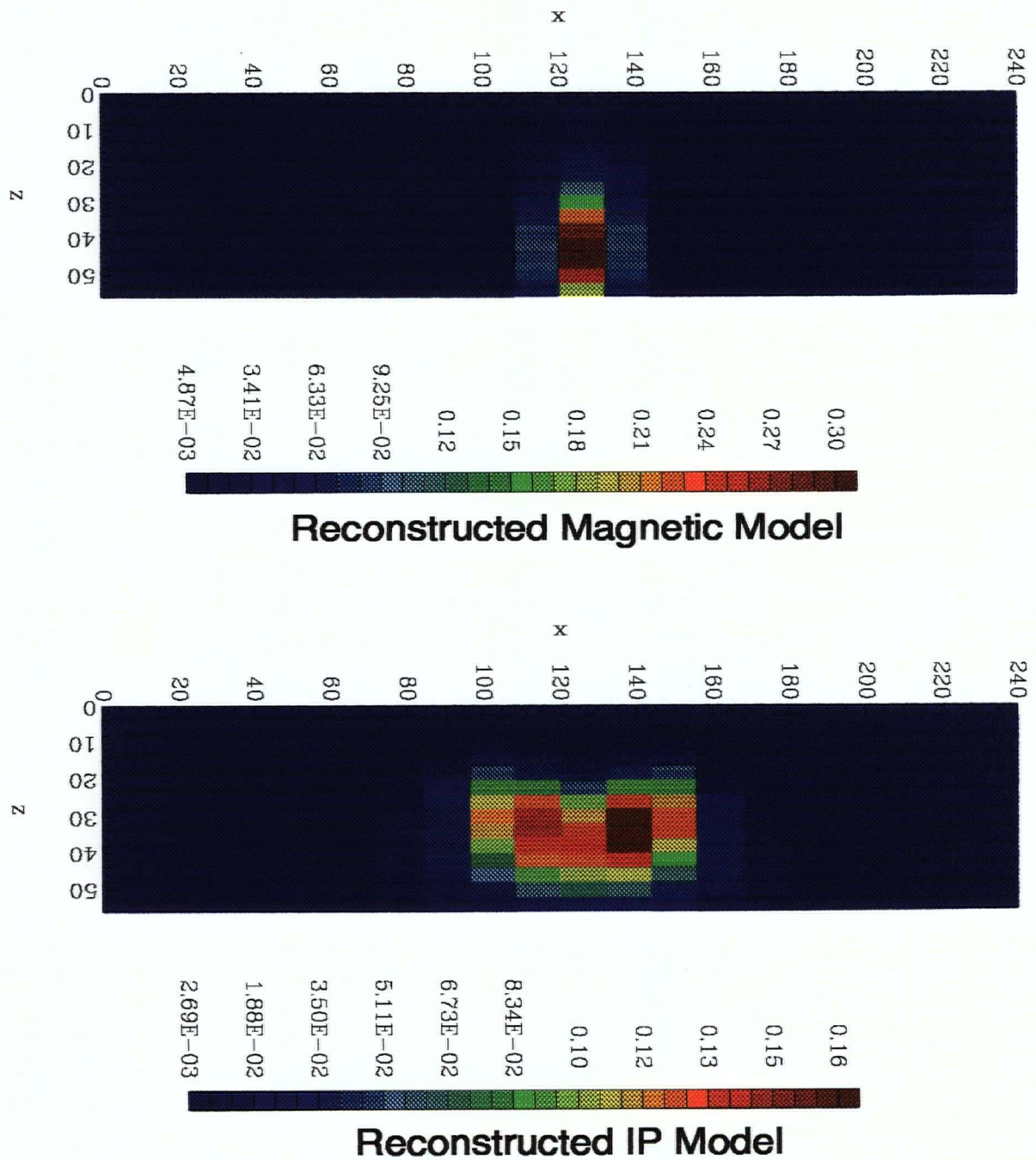
To solve ill-posed inverse problems, usually some kind of regularization is used. Regularization methods overcome the nonuniqueness of ill-posed problems by choosing a smooth model with minimum (semi-)norm. However, sometime the true model does not possess a smooth structure, so the regularization method based on smoothing does not work. To reveal this kind of nonsmooth structure, we consider different properties of the true model and find there is a correlation between its different physical parameters. With this additional information, it is natural to put it into the objective functional in solving the original inverse problem. In this thesis one more penalty is put into the objective function and corresponding combined inversion is developed.

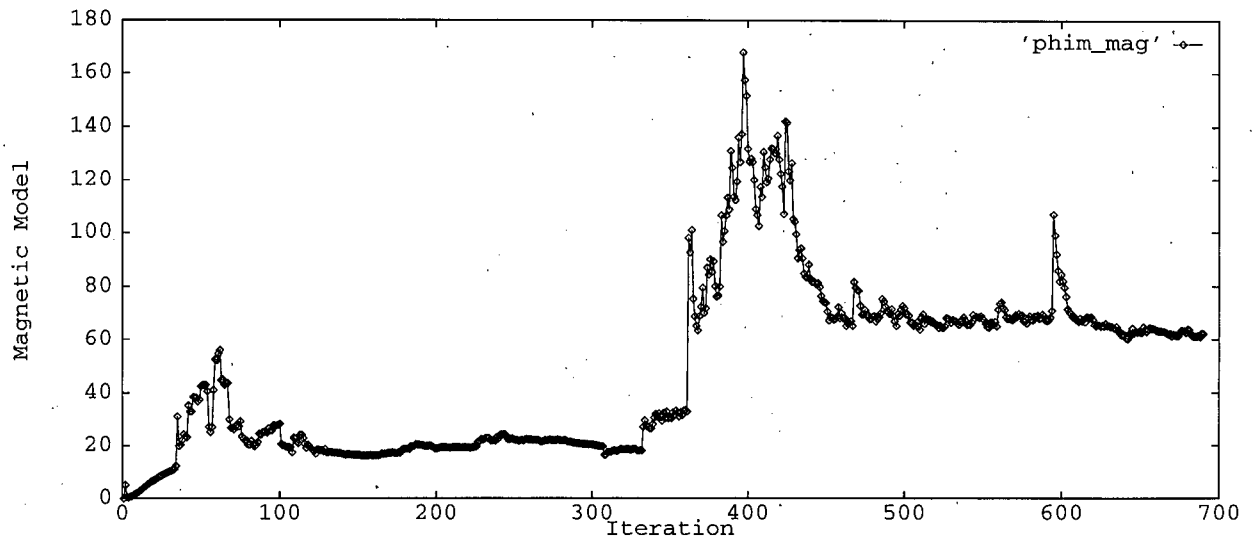
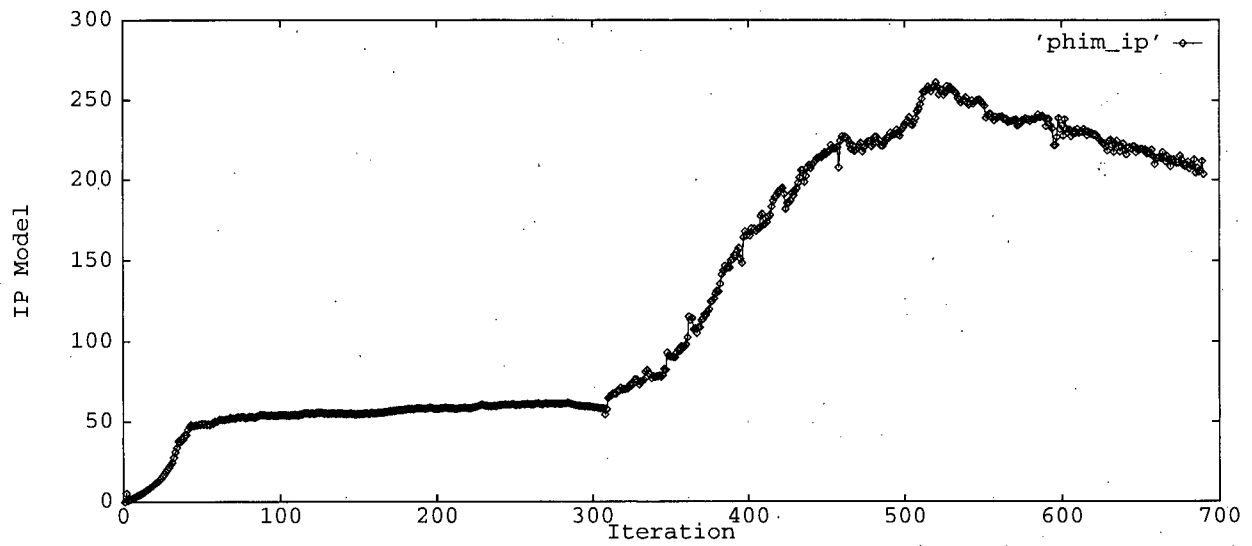
In doing this I found there is difficulty in solving the resulting optimization problem since the objective functional is neither convex nor quadratic. Even in the simple examples of this thesis we tried different computing techniques in solving the optimization problem. This combined inversion does reveal the nonsmooth characteristic of the true

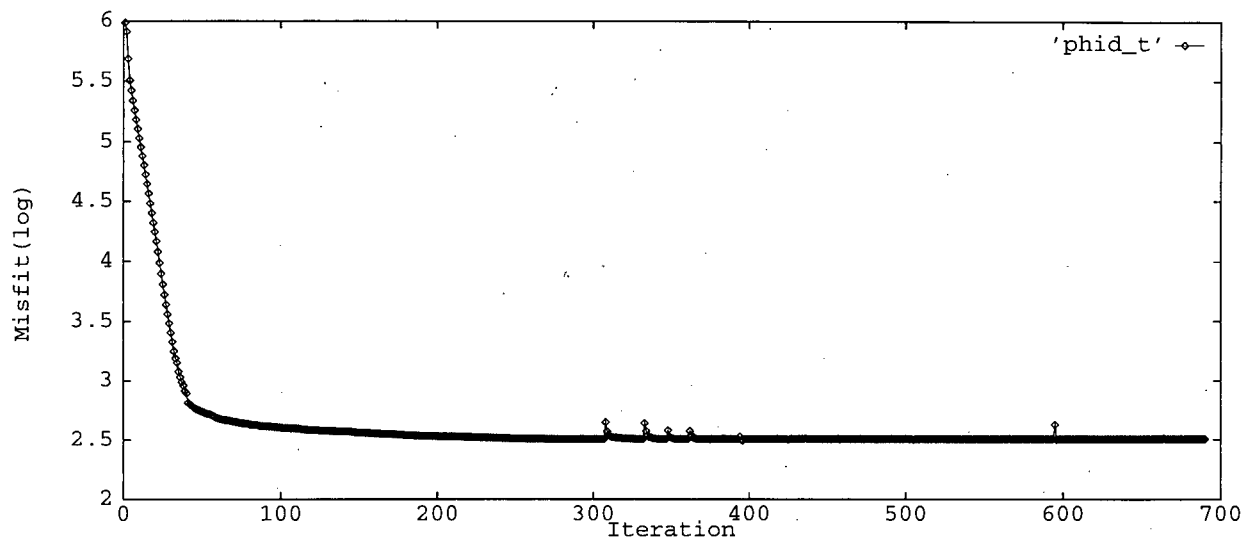
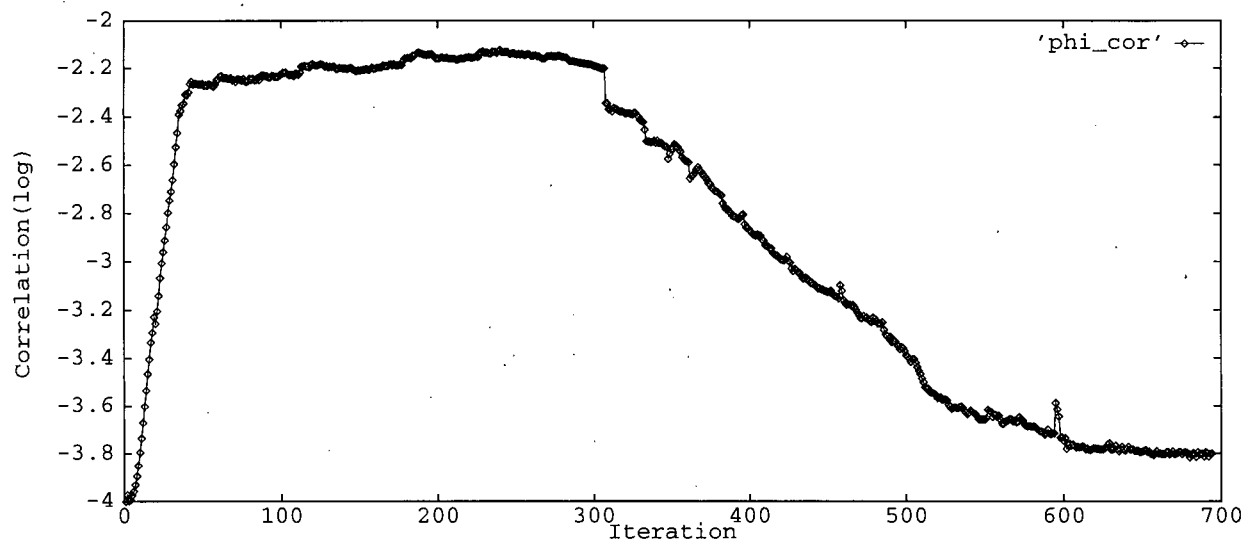
model and the reconstructed models are much better than those from the traditional regularization.

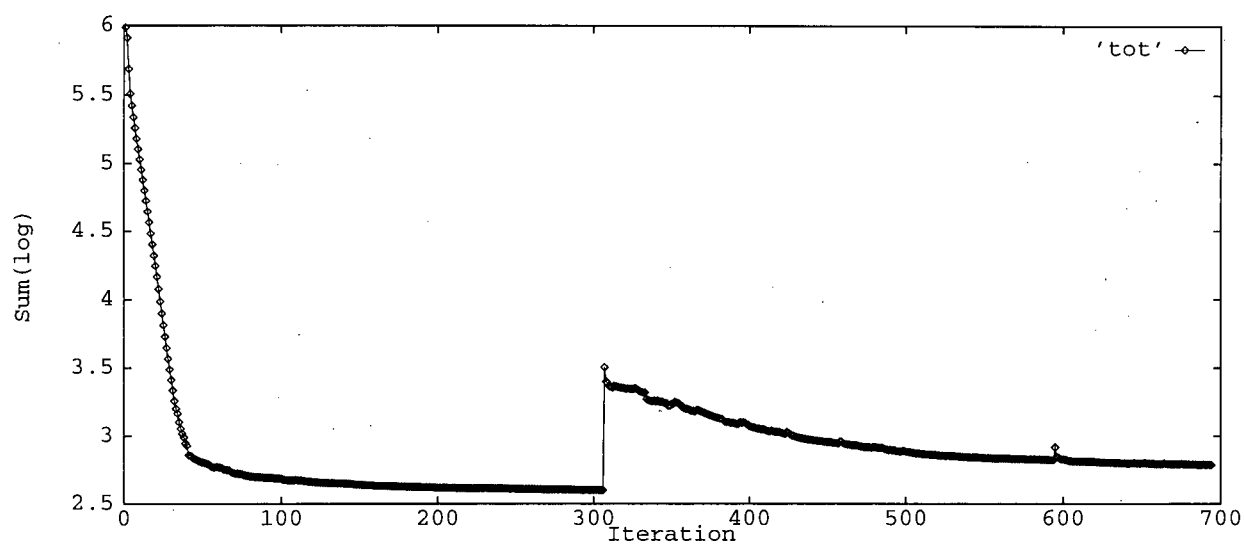
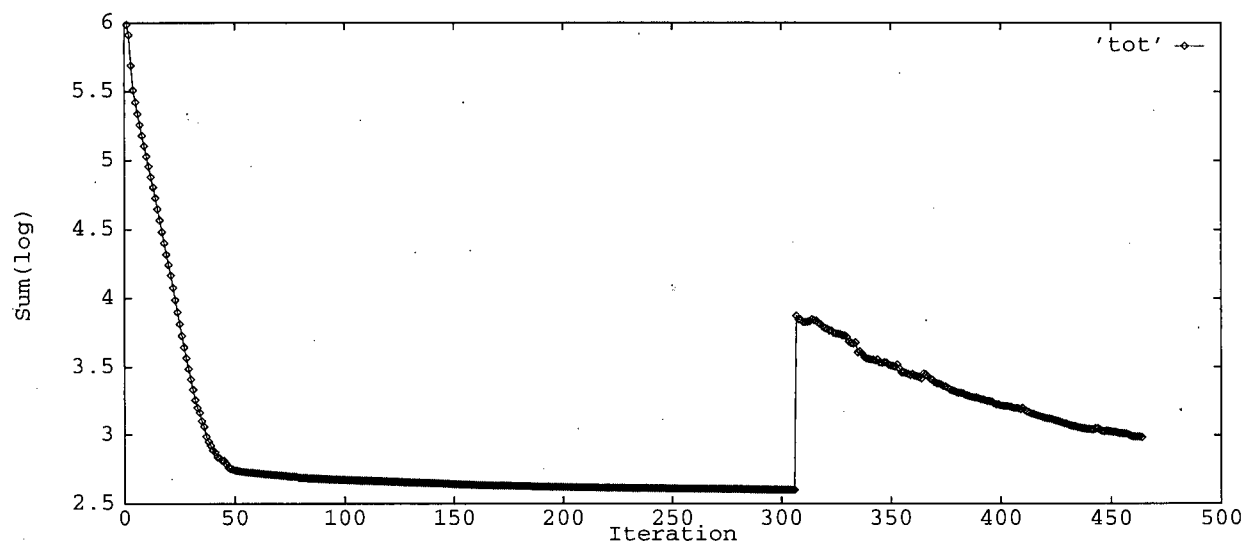
In contrast to traditional regularization methods which usually use one kind of data, this combined inversion method based on correlation uses observed data of different kinds. Although in this thesis we use the combined inversion to invert the susceptibility and the chargeability only, this method may be used to solve similar inverse problems arising from many areas. Generally speaking, if there is a correlation between two different physical parameters and each of these parameters can be inverted from corresponding data independently, we can use combined inversion to invert these two parameters simultaneously by keeping the reconstructed parameters in their confidence regions. The key part in using the combined inversion is to find the inherent correlation between the parameters and use proper functions to approximate them. Since combined inversion makes effective use of more information in solving ill-posed inverse problems, more reasonable outputs are obtained.

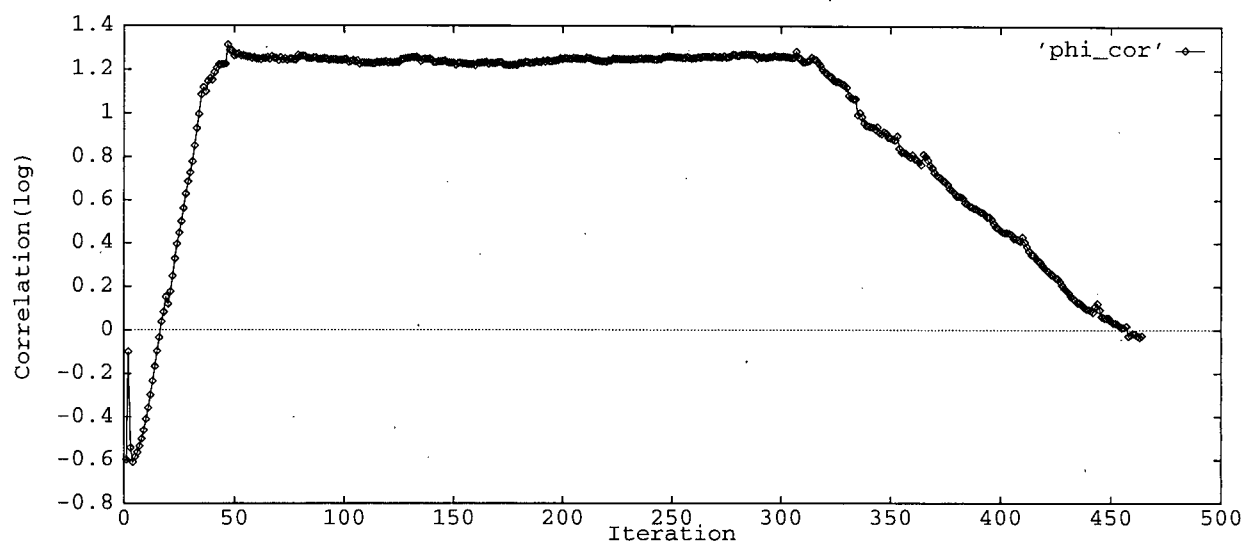
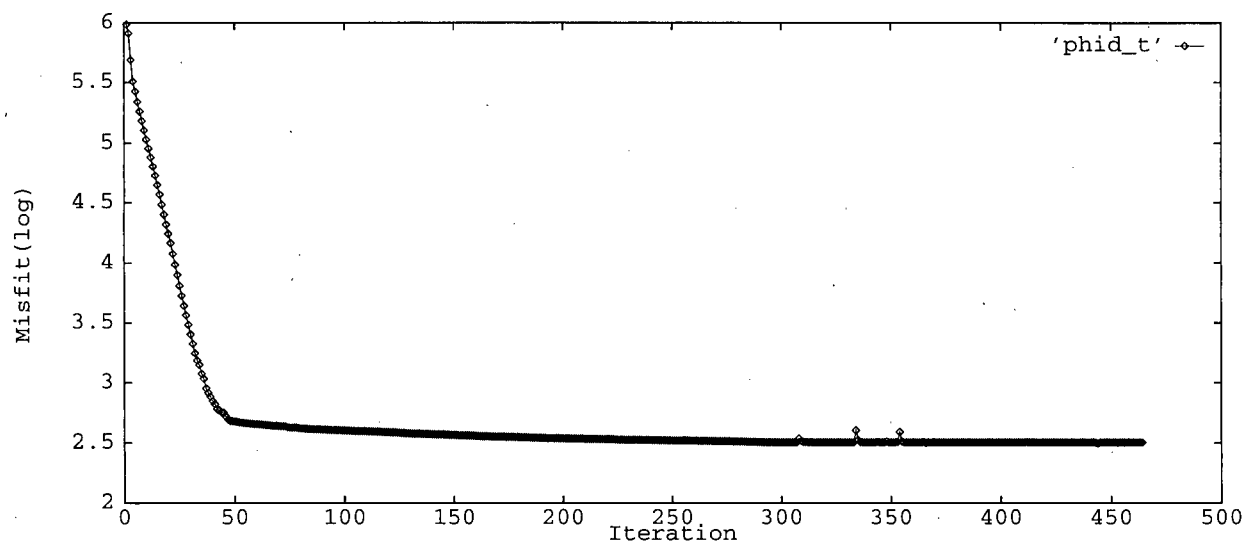
The combined inversion method here is developed to solve linear inverse problems, but it should be possible to develop similar combined inversion in solving nonlinear inverse problems. Based on this combined inversion, similar methods may be developed using the correlation among more than two physical parameters. In developing the combined inversion an efficient software package which can choose several parameters automatically in solving the resulting optimization problems should be developed.

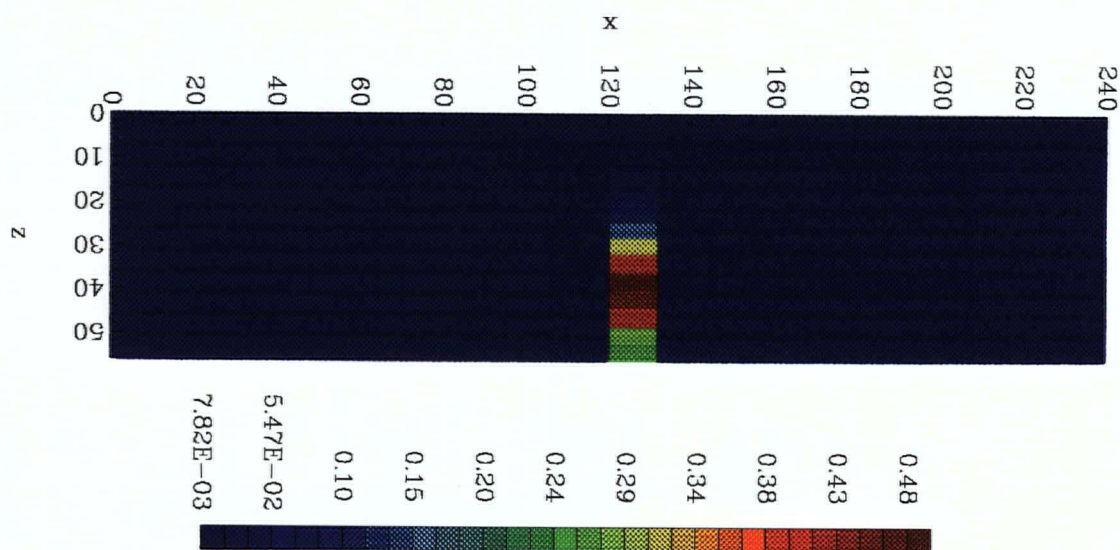
Figure 4.16: Reconstructed Models with $\phi_c = 0$.

Figure 4.17: Regularization term for magnetic part, ϕ_{κ} .Figure 4.18: Regularization term for IP part, ϕ_{η} .

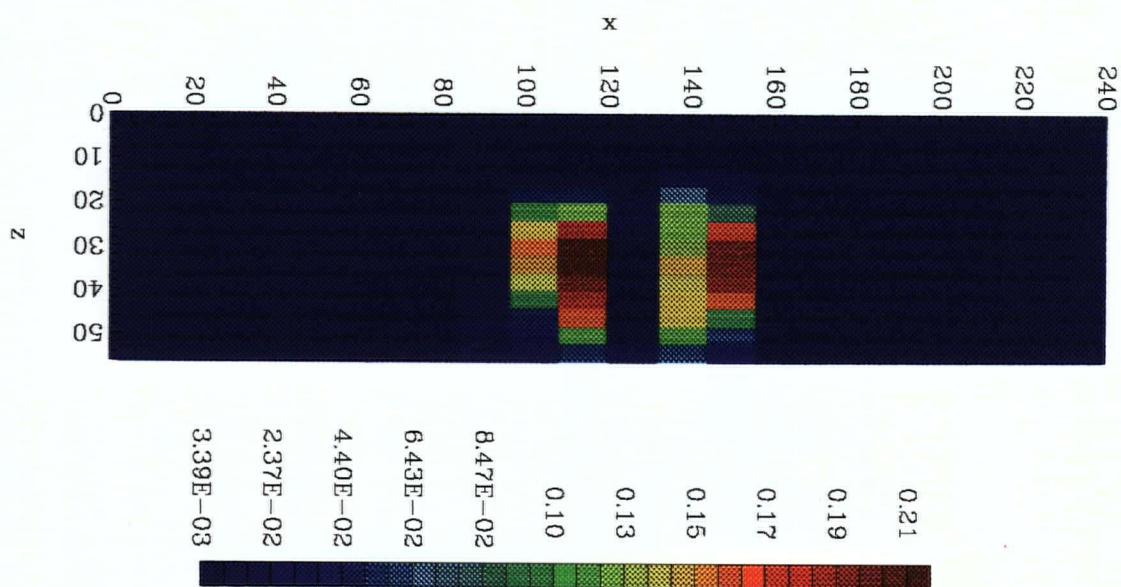
Figure 4.19: Total Misfit, ϕ_d .Figure 4.20: Correlation Function, $\phi_c(m)$.

Figure 4.21: Total sum of the objective function, ϕ .Figure 4.22: Total sum of the objective function, ϕ .

Figure 4.23: Correlation Function, $\phi_c(m)$.Figure 4.24: Total misfit, ϕ_d .



Reconstructed Magnetic Model



Reconstructed IP Model

Figure 4.25: Reconstructed Models with $\phi_c = \kappa^2 \eta^2$.

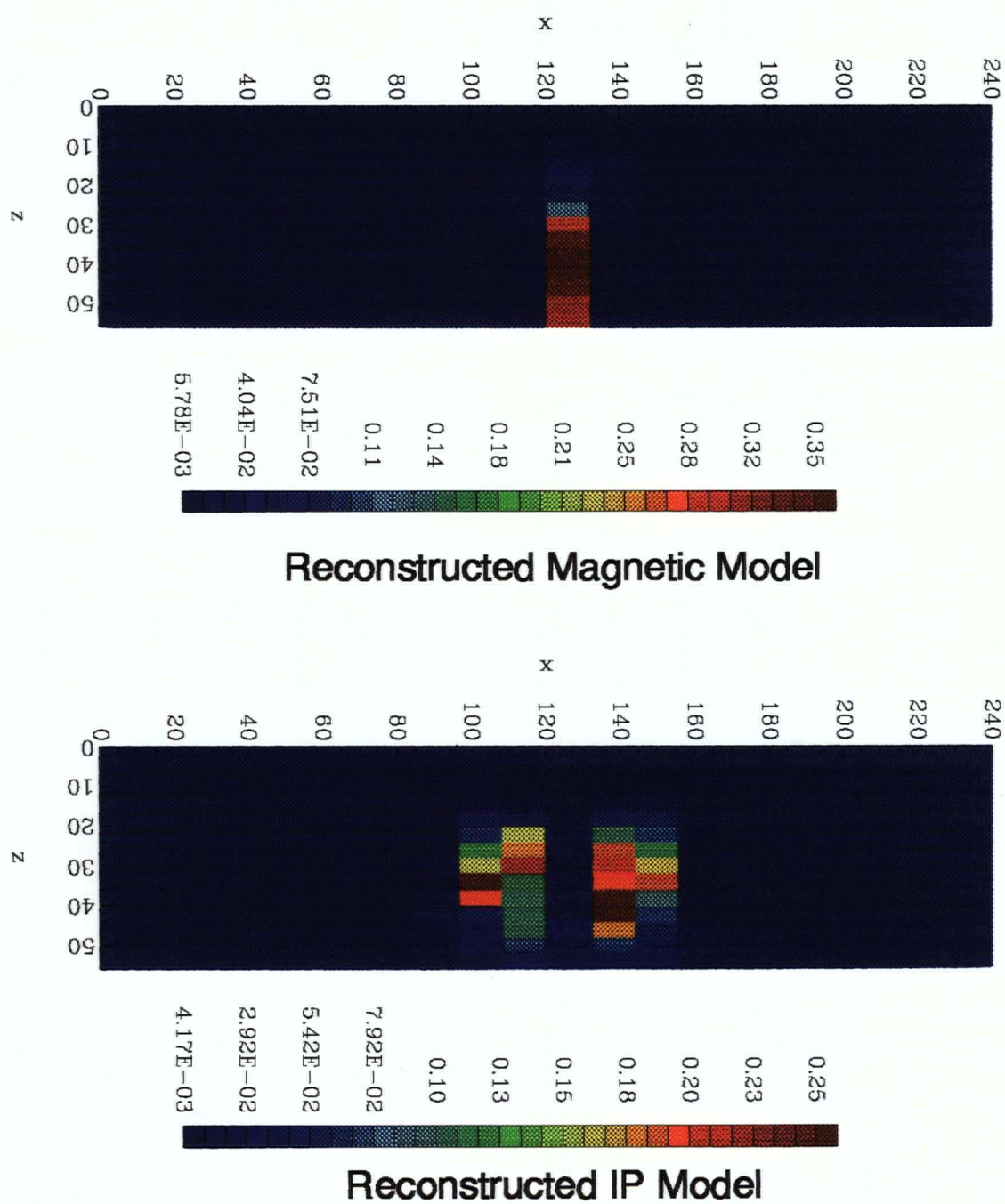


Figure 4.26: Reconstructed Models with ϕ_c =function product.

Bibliography

- [1] Anger, G., Gorenflo, R., Jochmann, H., Moritz, H. and Webers, W.: Inverse Problems: Principles and Applications in Geophysics, Technology and Medicine, Berlin: Akademie Verlag, (1993).
- [2] Courant, R. and Hilbert, D.: Methods of Mathematical Physics, Interscience, New York, (1953).
- [3] Dey, A., and Morrison, H. F. Resistivity modeling for arbitrarily shaped two-dimensional structures. Geophysical Prospecting 27, 106-136, (1979).
- [4] Gerkens, J. C.: Foundation of Exploration Geophysics. Elsevier Science Publishers B.V., The Netherlands, 1989.
- [5] Golub, G. H. and Van Loan, C. F.: Matrix computations. The Johns Hopkins University Press, Maryland (1983).
- [6] Groetsch, C. W.: Generalized inverses of linear operators: representation and approximation. Marcel Dekker, New York (1977).
- [7] Hadamard, J.: Lectures on Cauchy's problems in linear partial differential equations. Yale University Press, New Haven, 1923.
- [8] Heinz W. Engl, L.: Regularization methods for the stable solution of inverse problems. Surveys on Mathematics for Industry 3: 71-143 (1993).
- [9] Martin Hanke, K. and Per Christian Hansen, L.: Regularization methods for large-scale problems. Surveys on Mathematics for Industry 3: 253-315 (1993).
- [10] Lorrain, P., Corson, D. P. and Lorrain, F.: Electromagnetic fields and waves, W. H. Freeman and Company, New York, 1988.
- [11] McGillivray, P. R. and Oldenburg, D.W.: Methods for computing Frechet derivatives and sensitivities in the nonlinear inverse problem—a comparative study. Geophys. Prosp. 38: 499-524 (1990).
- [12] McGillivray, P. R.: Ph.D Thesis, University of British Columbia, (1992).
- [13] McMillan, W. J.: Porphyry Deposits in the Canadian cordillera; Ore deposits, tectonics and metallogeny in the Canadian Cordillera. Province of British Columbia; Ministry of Energy, Mines and Petroleum Resources, (1994).

- [14] Morozov, V. A.: Methods for solving incorrectly posed problems. Springer, New York (1984).
- [15] Oldenburg, D. W. and Li, Y.: Subspace Linear Inverse Method, *Inverse Problems* 10: 915-935 (1994).
- [16] Oldenburg, D. W. and Li, Y.: Inversion of Induced Polarization Data, to appear.
- [17] Rodi, W. L.: A technique for improving the accuracy of finited element solutions for MT data. *Geophysical Journal of the Royal Astronomical Society* 44, 483-506 (1976).
- [18] Sheriff, R. E.: Encyclopedic dictionary of exploration geophysics. Society of Exploration Geophysics, 1991.
- [19] Smith, N.C. and Vozoff, K.: Two-dimensional DC resistivity inversion for dipole-dipole data. *IEEE Transactions on Geoscience and Remote Sensing* GE-22, 21-28 (1984).
- [20] Telford, W. M., GelDart, L. P. and Keys, D. A.: *Applied Geophysics*. Cambridge; New York: Cambridge University Press, (1990).
- [21] Siegel, H. O.: A theory of induced polarization effects for step-function excitation. In *Overvoltage Research and Geophysical Applications*, J. R. Wait, ed. pp. 4-21. London: Pergamon, (1959).
- [22] Tikhonov, A. N., Arsenin, V.Y.: *Solutions of ill-posed problems*. Wiley, New York (1977).
- [23] Van Loan, C. F.: Generalizing the singular value decomposition. *SIAM J. Numerical Analysis* 13: 76-83 (1976).
- [24] Varah, J. M.: A practical examination of some numerical methods for linear discrete ill-posed problems. *SIAM Review* 21: 100-111(1979).
- [25] Varah, J. M.: Pitfalls in the numerical solution of linear ill-posed problems. *SIAM J. Sci. Statis. Comput.* 4: 164-176 (1983).
- [26] Wahba, G.: The practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.* 14: 651-667 (1977).